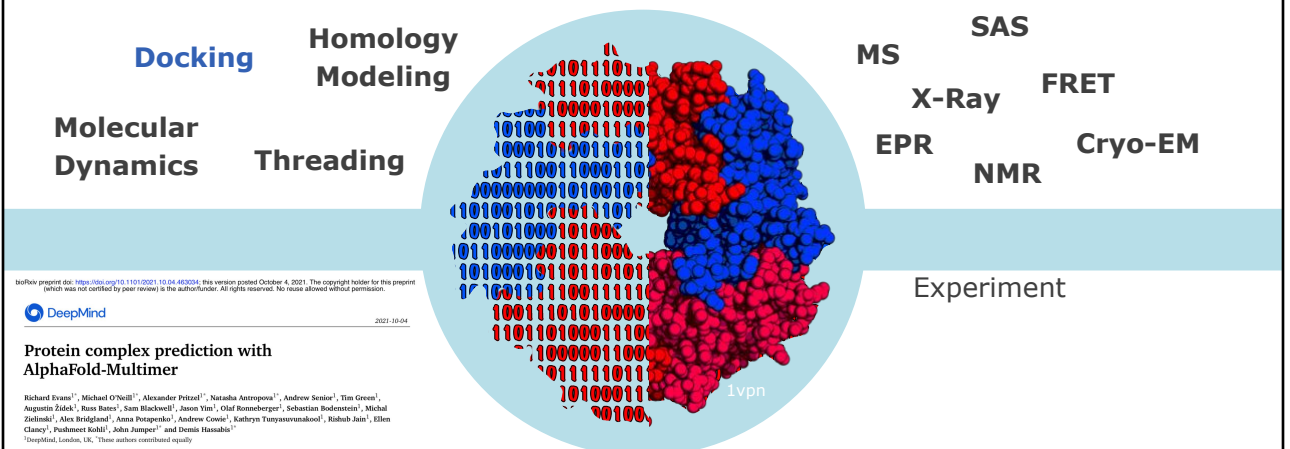


Bioinformatyka

Andrzej Łyskowski, dr inż.
Katedra Biotechnologii i Bioinformatyki
andrzej.lyskowski@prz.edu.pl
H-237

Biologia strukturalna | interakcje



High-throughput computation vs. High-resolution experiments
computational models are often not trusted by the experimental community

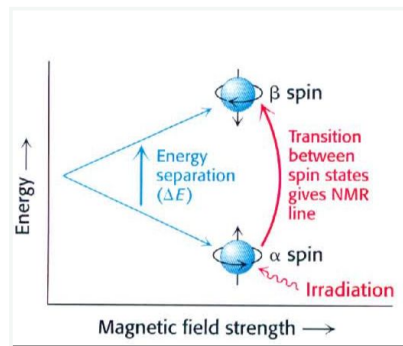
Spektroskopia magnetycznego rezonansu jądrowego (NMR)

Podstawa:

niektóre jądra atomowe wykazują właściwości magnetyczne.

różnica energii *spinów* jest proporcjonalna do siły przyłożonego pola magnetycznego.

rozszczipienie energii jest zależne od środowiska.

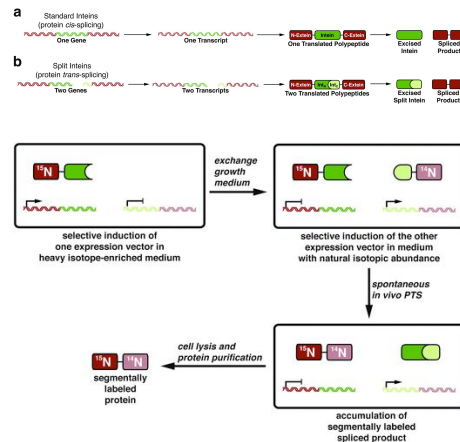


Naturalnie występujące izotopy pierwiastków o znaczeniu biologicznym.

Nucleus	Natural abundance (% by weight of the element)
^1H	99.984
^2H	0.016
^{13}C	1.108
^{14}N	99.635
^{15}N	0.365
^{17}O	0.037
^{23}Na	100.0
^{25}Mg	10.05
^{31}P	100.0
^{35}Cl	75.4
^{39}K	93.1

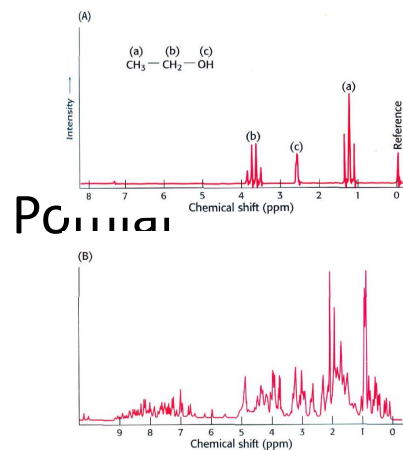
Skład izotopowy białek musi być wzbogacony w trakcie syntezy białek.

zastosowanie pożywek o zmodyfikowanym składzie izotopowym oraz odpowiednich szczepów ekspresyjnych.
znakowanie domen.



Stężona próbka białka:

1 mM lub 15 mg/ml⁻¹ dla białka 15 kDa



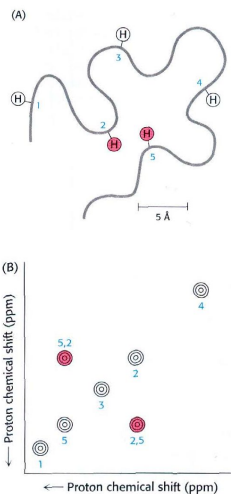
Pomiar

Uproszczenie widma NMR osiąga się poprzez zastosowanie odpowiednich metod pomiarowych:

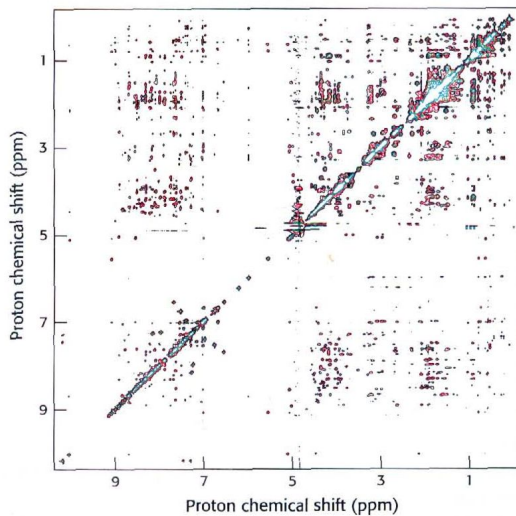
jądrowego efektu Overhausera (NOE) w technice spektroskopii jądrowego efektu Overhausera (NOESY);

spektroskopii korelacyjnej (COSY);

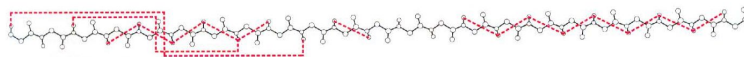
całkowitej spektroskopii korelacyjnej (TOCSY)



Analiza



Analiza

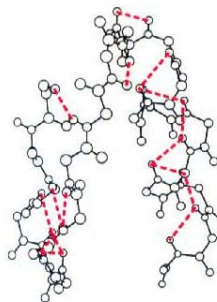


W trakcie interpretacji widm
(ręcznej lub automatycznej)
stosuje się następujące typy
ograniczeń:

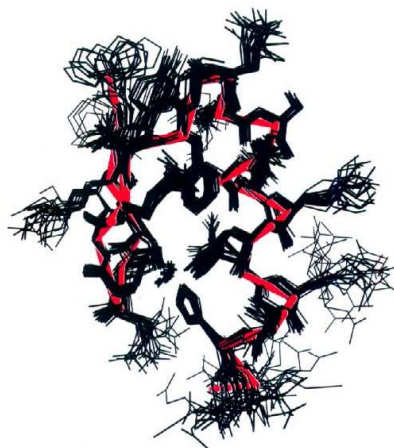
odległości;

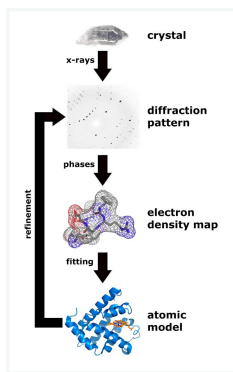
kątów;

orientacji przestrzennej.



Wynikiem
pomiarów jest
rodzina struktur.

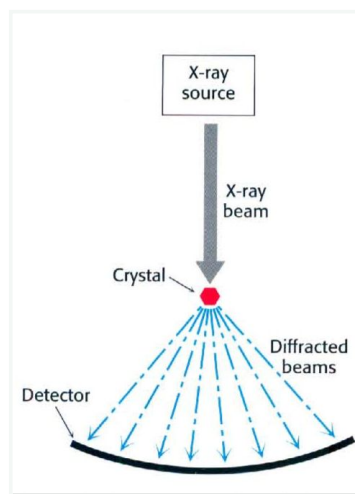




Etapy wyznaczania struktury

Krystalografia rentgenowska

To pierwsza i ciągle najczęściej wykorzystywana metoda oznaczania struktury przestrzennej białka na poziomie atomowym.



Kryształ

Kryształ

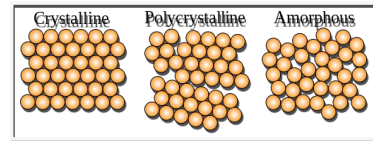
to materiał, którego materiał na poziomie atomów/ionów/cząstek jest rozmieszczony w regularny i powtarzalny sposób.

Krystalizacja białek

proces poszukiwania selektywnych, indywidualnych warunków krystalizacji białka poprzez:

wysalanie;

metody przesiewowego badania różnych warunków.

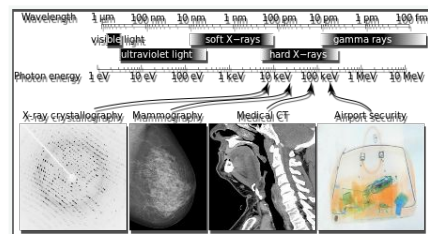


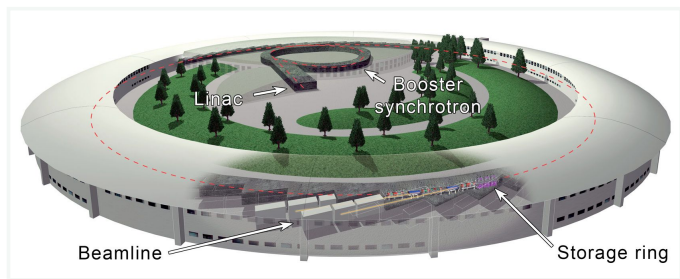
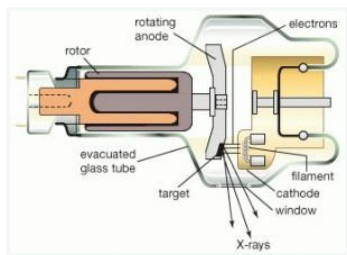
Promieniowanie rentgenowskie

Promieniowanie

w zakresie 0.1-10 nm (100 eV-100keV).

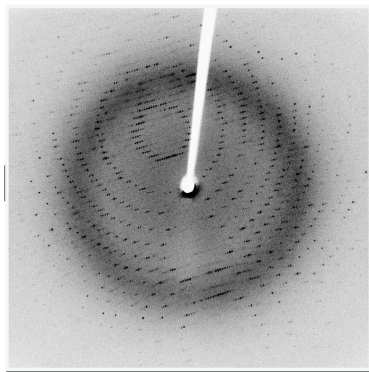
Dlaczego nie można stosować światła widzialnego?



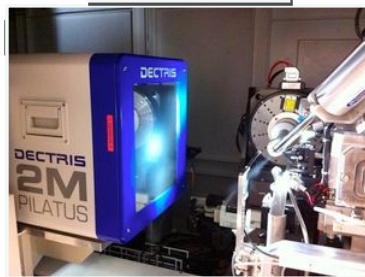


Promieniowanie
rentgenowskie ulega
rozszczepieniu na
elektronach
atomów.

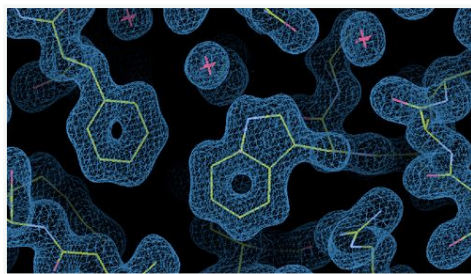
Rozproszone fale
nakładają się
tworząc obraz
interferencyjny, na
którym
wzmocnienie lub
wygaszenie fali jest
zależne od rodzaju i
położenia atomu w
kryształe.



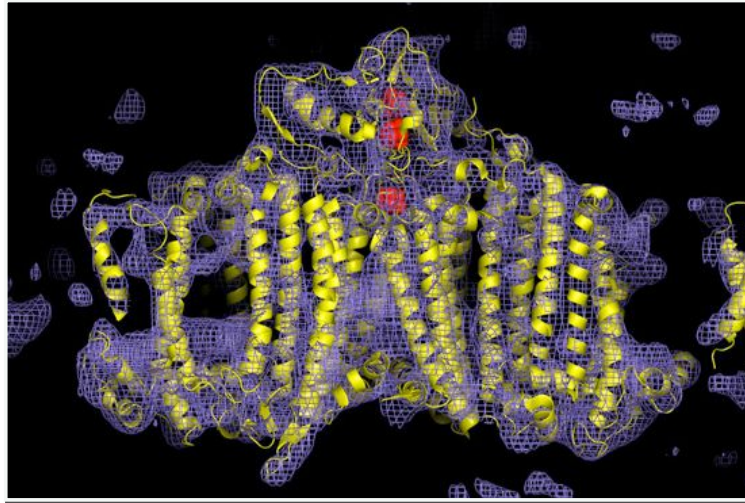
Historycznie rolę
detektora
spełniała klisza
fotograficzna.



Po wstępnej
analizie
(indeksacji) dane
poddawane są
przekształceniu
Fouriera. W jego
wyniku
otrzymuje się
mapę
elektronową,
którą musi
zostać
zinterpretowana.

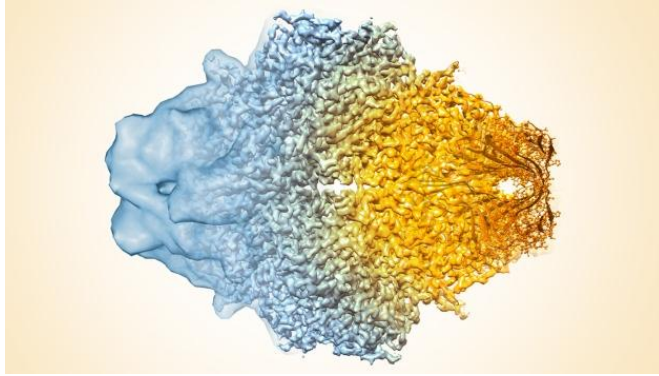


Struktura 3D



- › Kryształ vs. roztwór.
- › Pojedyncza struktura vs. rodzina.
- › Statyczny obraz vs. obraz dynamiczny.
- ›
- › Dowolny rozmiar cząsteczki vs. limit masy cząsteczkowej.

from the low-resolution density map to the atomic coordinates

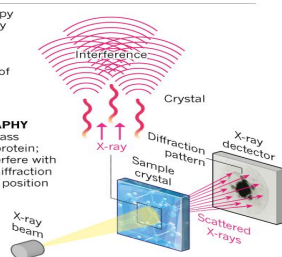


Metody eksperymentalne

Cryo-electron microscopy is taking over from X-ray crystallography as a method to deduce high-resolution protein structures, particularly of large molecules.

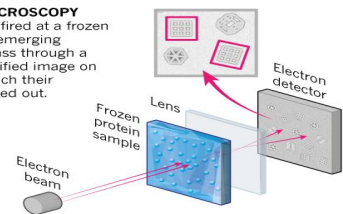
X-RAY CRYSTALLOGRAPHY

X-rays scatter as they pass through a crystallized protein; the resulting waves interfere with each other, creating a diffraction pattern from which the position of atoms is deduced.



CRYO-ELECTRON MICROSCOPY

A beam of electron is fired at a frozen protein solution. The emerging scattered electrons pass through a lens to create a magnified image on the detector, from which their structure can be worked out.

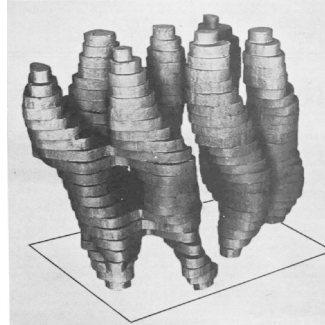


© nature

Transmission electron microscopy, as the technique is called, works more or less like ordinary microscopy, but a beam of electrons is sent through the sample instead of light. The electrons' wavelength is much shorter than that of light, so the electron microscope can make very small structures visible – even the position of individual atoms.

Problemy

- The intense electron beam necessary for obtaining high resolution images incinerates bio-logical material and, if the beam is weakened, the image loses its contrast and becomes fuzzy.
- Electron microscopy requires a vacuum, a condition in which biomolecules deteriorate because the surrounding water evaporates.

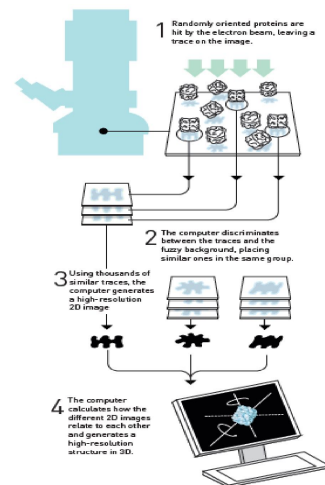


The first rough model of bacteriorhodopsin, published in 1975. Image from *Nature* 257: 28-32

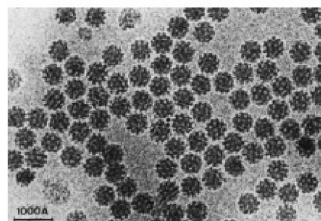
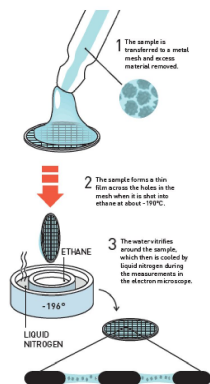
It was the best picture of a protein ever generated using an electron microscope. Many people were impressed by the resolution, which was 7 Ångström (0.0000007 millimetres), but this was not enough for Richard Henderson. His goal was to achieve the same resolution as that provided by X-ray crystallography, about 3 Ångström, and he was convinced that electron microscopy had more to give.

Procedura

1. Komputer rozpoznaje sygnał pochodzący od białek ułożonych w losowych orientacjach w przygotowanej próbce.
2. W oparciu o odpowiednie algorytmy zidentyfikowane zastają powtarzające się elementy.
3. Zebrane informacje są katalogowane i w ich oparciu powstaje model 2D.
4. Komputer generuje obraz 3D w oparciu o zebrane dane.



Witryfikacja



Dubochet generated the first images of viruses surrounded by vitrified water in 1984. Image from *Nature* 308: 32-36.

Initially, the research group attempted to vitrify tiny drops of water in liquid nitrogen at -196°C, but were successful only when they replaced the nitrogen with ethane that had, in turn, been cooled by liquid nitrogen. Under the microscope they saw a drop that was like nothing they had seen before. They first assumed it was ethane, but when the drop warmed slightly the molecules suddenly rearranged themselves and formed the familiar structure of an ice crystal. It was a triumph – particularly as some researchers had claimed it was impossible to vitrify water drops. We now believe that vitrified water is the most common form of water in the universe.

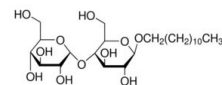
Wyznaczenie struktury GLP1-R

Stabilna biomolekuła

- For cryo-EM studies we used rabbit GLP-1R, which shares 92% identity with the human receptor, as it expressed at higher levels in insect cells than the human or mouse homologue.

Adekwatna stabilizacja

- GLP-1R is very unstable to extraction from membranes using conventional detergents such as DDM (n-dodecyl β-D-maltoside) and MNG (maltose-neopentyl glycol); therefore, we formed the complex between receptor and purified Gs in insect cell membranes before extraction of the GLP-1R-Gs complex with MNG and purification by antibody affinity and size-exclusion chromatography.



n-Dodecyl β-D-maltoside

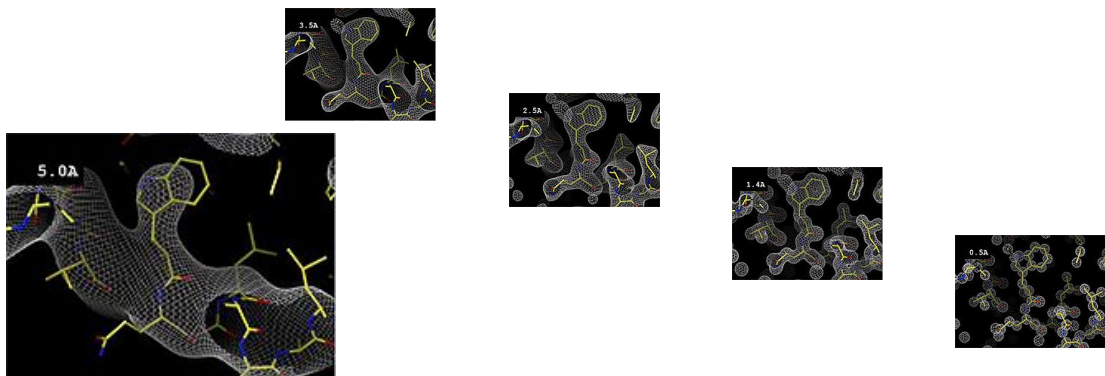
D4641-500MG	✓ Dostępny do wysyłki 17.10.17 - Z	184.00 830.25
D4641-1G	✓ Dostępny do wysyłki 17.10.17 - Z	268.00 1,206.00
D4641-5G	✓ Tylko 5 szt. w magazynie (więcej w drodze). - Z	887.00 3,991.50
D4641-25G	✓ Tylko 1 szt. w magazynie (więcej w drodze). - Z	3,640.00 15,930.00

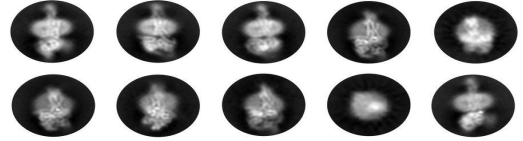
Struktura

• Sample evaluation by negative-stain EM and single-particle averaging confirmed a monodisperse particle population and stable complex formation. Initial cryo-EM experiments suggested that the protein complex avoided areas of thin vitreous ice, and we thus had to image the specimen in relatively thick ice at the expense of increased background noise in the micrographs.

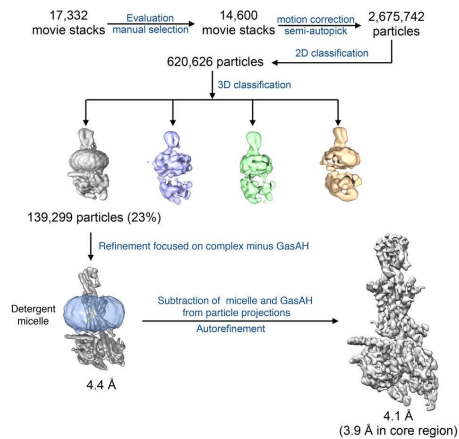
• Refinement and reconstruction of the selected particle projections after subtracting densities for the detergent micelle and the mobile α -helical domain enabled us to obtain a 3D reconstruction of the complex at 4.1 Å global resolution, with 3.9 Å nominal resolution in the core region that includes GLP-1, TMD and the $\alpha 5$ helix of the Gas Ras-like domain.

Rozdzielczość

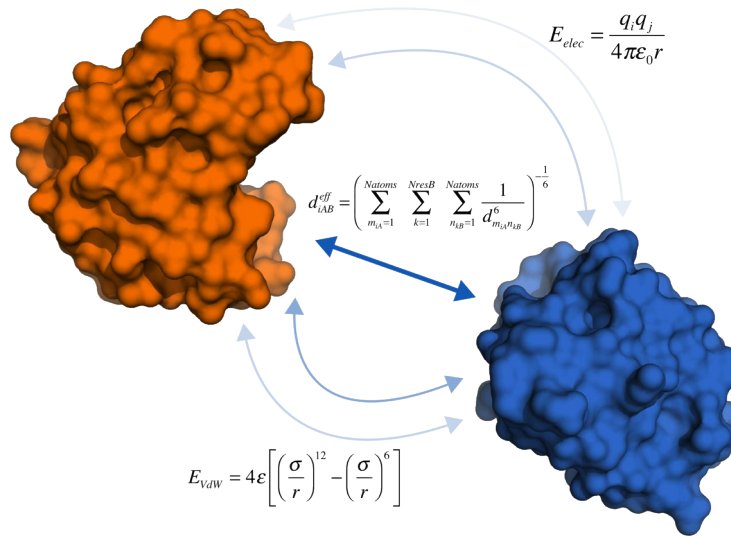




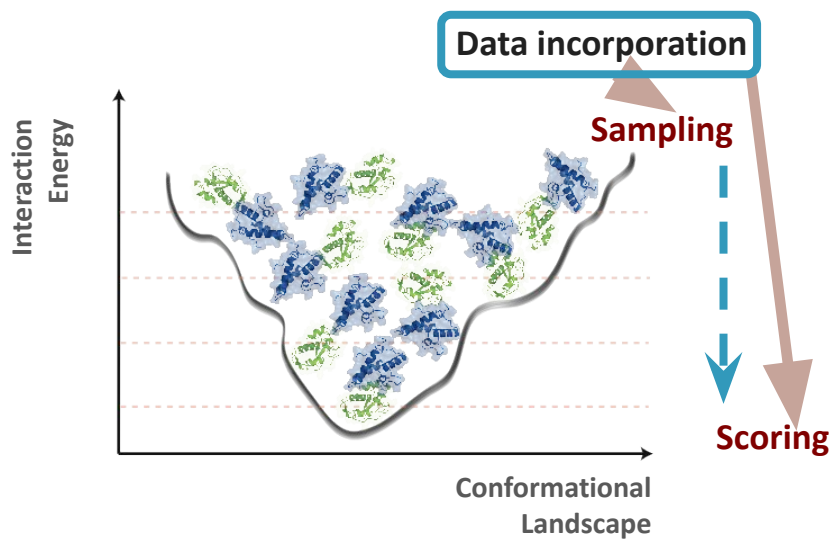
Rekonstrukcja EM



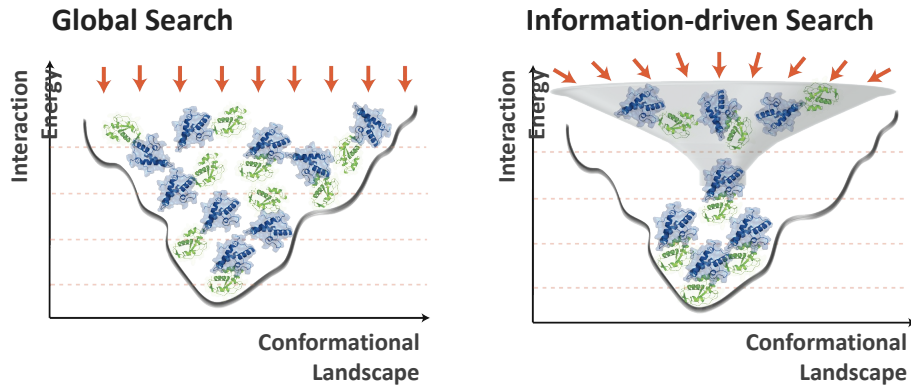
Molecular Docking



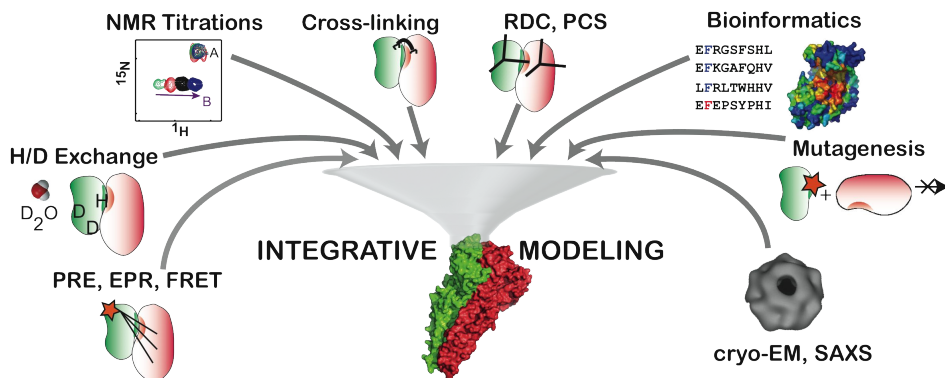
Metodologia



Integracja danych

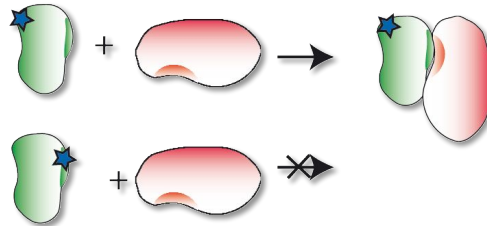


Kompleksowe modelowanie



Experimental sources:

mutagenesis



Advantages/disadvantages

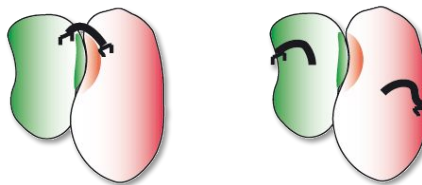
- + Residue level information
- Loss of native structure should be checked

Detection

- Binding assays
- Surface plasmon resonance
- Mass spectrometry
- Yeast two hybrid
- Phage display libraries, ...

Experimental sources:

cross-linking and other chemical modifications



Advantages/disadvantages

- + Distance information between linker residues
- Cross-linking reaction problematic
- Detection difficult

Detection

- Mass spectrometry

Experimental sources:

H/D exchange



Advantages/disadvantages

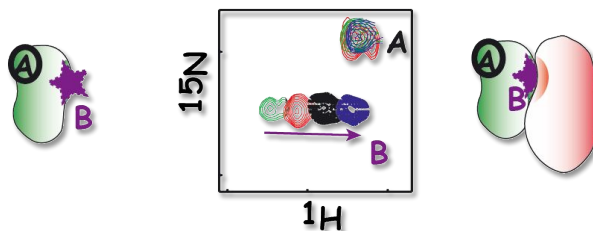
- + Residue information
- Direct vs indirect effects
- Labeling needed for NMR

Detection

- Mass spectrometry
- NMR ^{15}N HSQC

Experimental sources:

NMR chemical shift perturbations



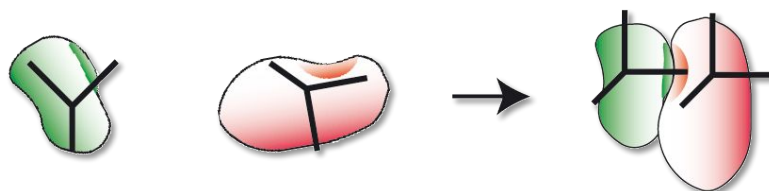
Advantages/disadvantages

- + Residue/atomic level
- + No need for assignment if combined with a.a. selective labeling
- Direct vs indirect effects
- Labeling needed

Detection

- NMR ^{15}N or ^{13}C HSQC

Experimental sources: NMR orientational data (RDCs, relaxation)



Advantages/disadvantages

- + Atomic level
- Labeling needed

Detection

- NMR

HADDOCK: An integrative modeling platform

Incorporates ambiguous and low-resolution data to aid the docking

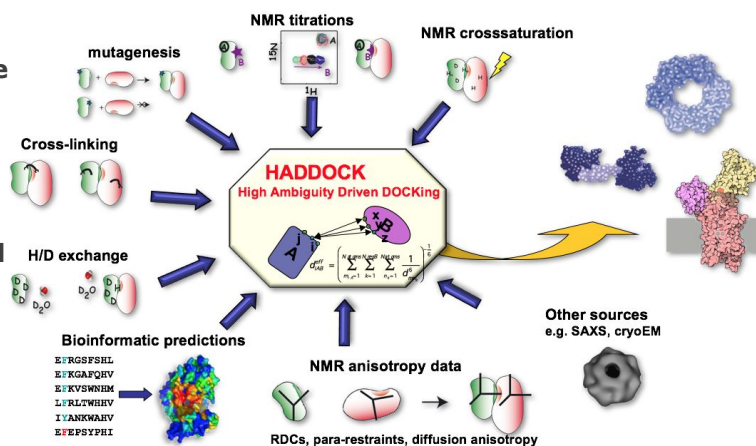
Capable of docking up to 20 molecules (2.4 version)

Symmetries can be leveraged

Allows for flexibility at the interface

Final flexible refinement in explicit solvent

Consistent performance over the years in CAPRI



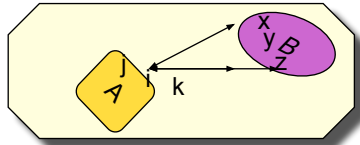
Dominguez, Boelens & Bonvin. JACS 125, 173 (2003).

<http://www.bonvinlab.org/software>

Data-driven docking with HADDOCK

List of interface residues
for protein A

List of interface residues
for protein B

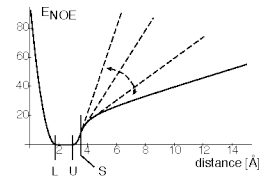


Effective distance

d_{iAB}^{eff}

calculated as

$$d_{iAB}^{eff} = \left(\sum_{m_i=1}^{N_{Lqms}} \sum_{k=1}^{N_{resB}} \sum_{n_k=1}^{N_{Lqms}} \frac{1}{d_{mn_k}^6} \right)^{\frac{1}{6}}$$



Ambiguous Interaction Restraint:

a residue must make contact with any residue from the other list

Different fraction of restraints (typically 50%) randomly deleted for each docking trial to deal with inaccuracies and errors in the information used

(Nilges & Brunger 1991)

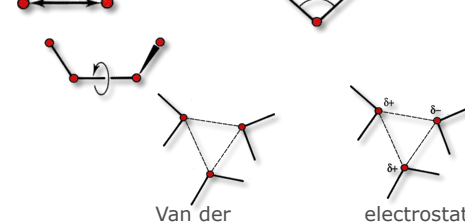
$$E_{NOE} = \begin{cases} (r-L)^2 & \text{if } r < L \\ 0 & \text{if } L < r < U \\ (U-r)^2 & \text{if } U < r < S \\ A(r-U)^{-1} + B(r-U) + C & \text{if } r > S \end{cases}$$

Searching the interaction space in HADDOCK

Experimental and/or predicted information is combined with an empirical force field into an energy function whose minimum is searched for

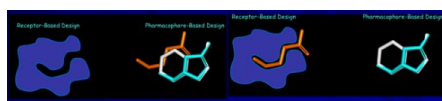
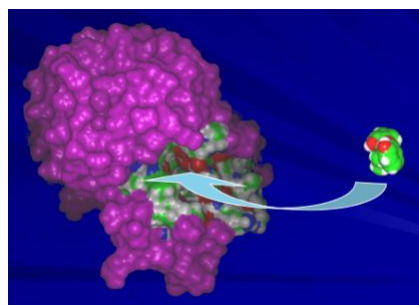
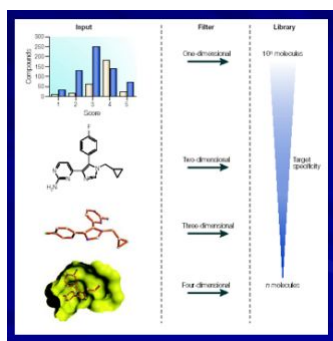
$$\begin{aligned} V_{\text{potential}} &= V_{\text{bonds}} \\ &+ V_{\text{torsion}} \\ &+ V_{\text{non-bonded}} \\ &+ V_{\text{exp}} \end{aligned}$$

+ V_{angles}



Search is performed by a combination of gradient driven energy minimization and molecular dynamics simulations

Projektowanie leków oparte o receptor



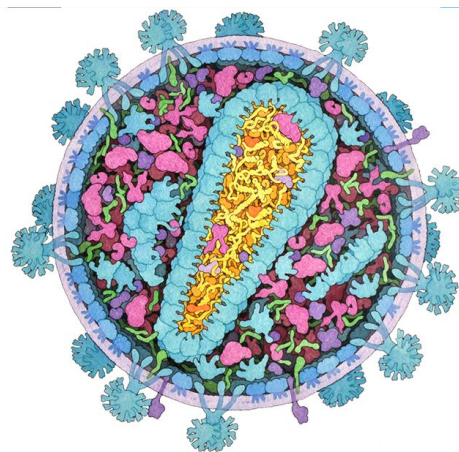
>projektowanie leków oparte o receptor

Analiza struktury 3D
Identyfikacja miejsca
aktywnego
Identyfikacja interakcji:

Orientacja liganda

Odziaływania ligand-receptor

Wprowadzenie
komplementarnych
modyfikacji



>dokowanie (*ang. docking*)

Procedura:

Umieść ligand w miejscu aktywnym

Zminimalizuj energię

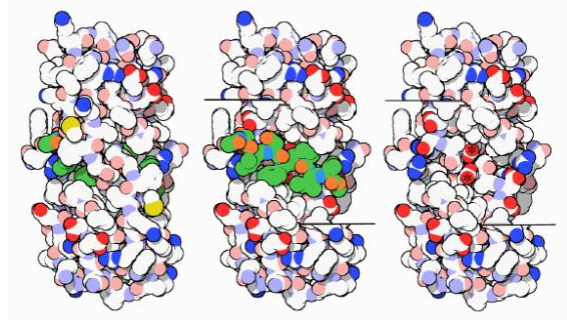
Zoptymalizuj interakcje

Wiązania wodorowe

Oddziaływania hydrofobowe

Zawadę steryczną

Oceń

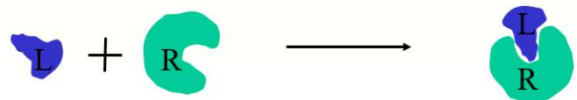


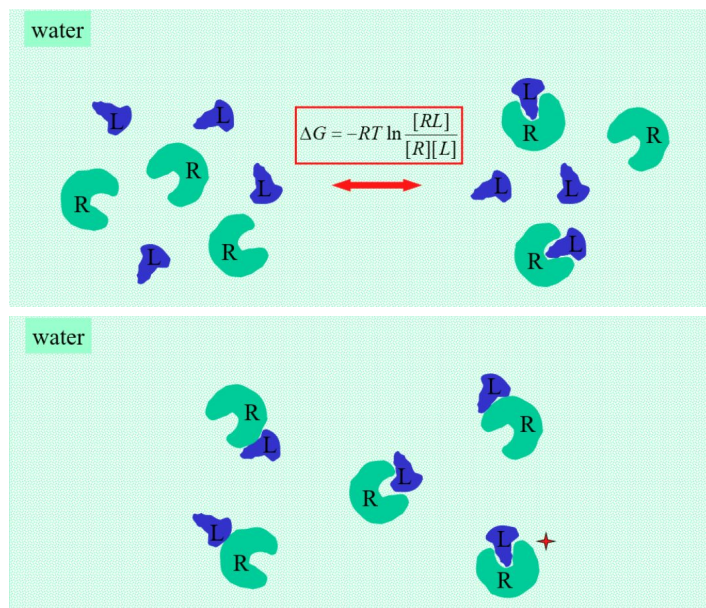
>dokowanie

Cel:

wyznaczenie za pomocą metod obliczeniowych kompleksu ligand-receptor.

Aktywność
biologiczna oparta
jest o specyficzne
oddziaływania





>dokowanie: przygotowanie modelu

Punkt wyjścia:

Struktura kompleksu jest znana (*ang. bound docking*)

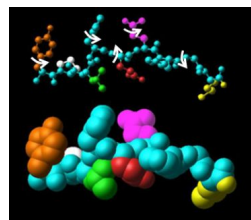
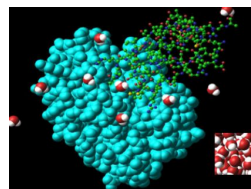
Ograniczona przestrzeń dokowania

Znana ,energia'

Brak ,giętkości'

,Opisana' natura oddziaływań

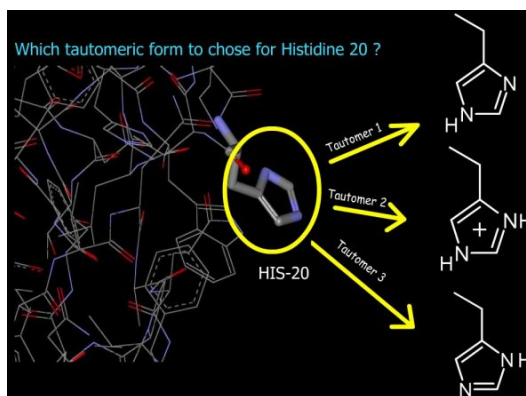
Struktura kompleksu nie jest znana (*ang. unbound docking*)



>dokowanie: przygotowanie receptora

Każda symulacja wymaga precyzyjnego określenia parametrów startowych.

Właściwości aminokwasów zmieniają się wraz z parametrami środowiska.

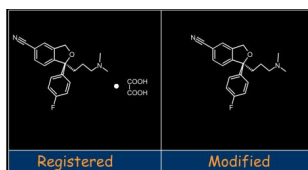


>dokowanie: przygotowanie ligandu

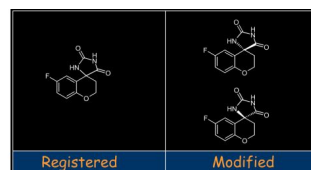
Struktura ligandu musi zawierać:

wyłącznie ligand

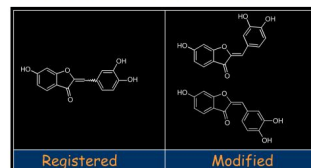
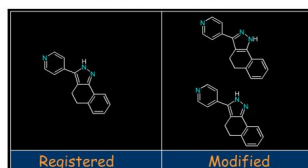
wszystkie wariacje strukturalne (enancjomery, izomery cis/trans, tautomery, warianty ,H')



Tautomers generation



Double bond cis/trans isomers generation



>dokowanie: optymalizacja

Dokowanie:

Statyczne (*ang. static*)

Dynamiczne (*ang. flexible*)

Uwzględnia zmiany w strukturze 3D ligandu i receptora

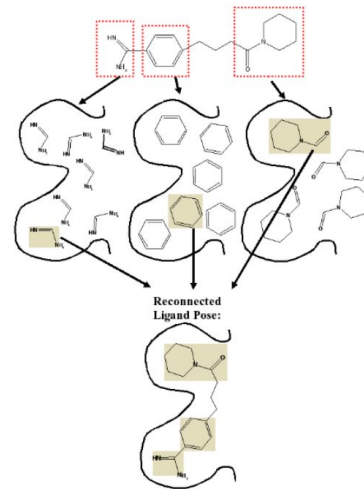
Zmiany strukturalne można oszacować za pomocą:

Dynamiki molekularnej i minimalizacji energii kompleksów

Metod Monte Carlo

Bibliotek (rotamerów, białek (NMR))

Lokalnych potencjałów



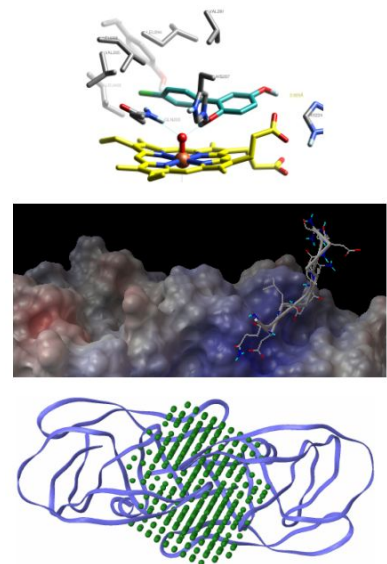
>dokowanie: wizualizacja

Oddziaływanie ligand-recetor można analizować na różnych poziomach:

Atomowym

Powierzchni initeracji

Siatki (*ang. grid*)

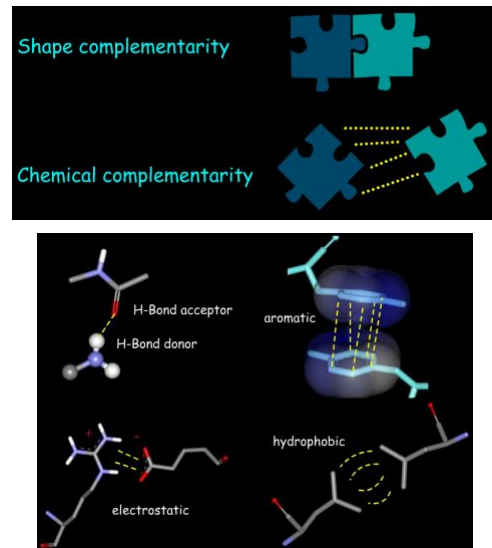


>dokowanie: interpretacja wizualizacji

Interpretacja komplementarności oddziaływań ligand-receptor:

Na poziomie kształtu

Na poziomie właściwości fizyko-chemicznych



>dokowanie: strategie

Elastyczny ligand | Swobodne wiązania

Zmiany w entropii | Swobodne wiązania

Solvatacja | Interakcje ligand-receptor

Interakcje związków

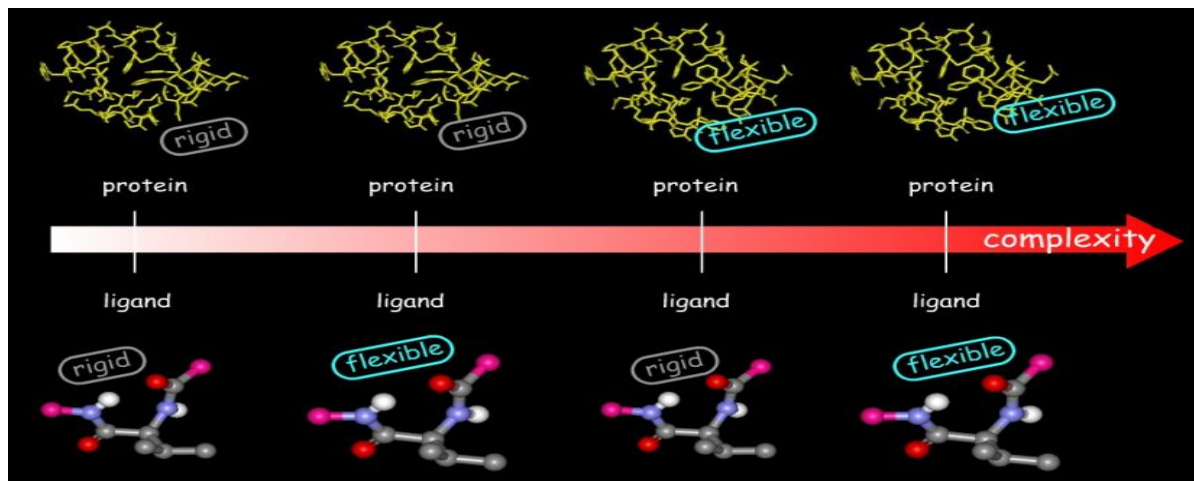
małocząsteczkowych (woda, jony, itp.)

Tautomery

Elastyczność białek | Wymuszone dopasowanie

Specyficzność wiązania | Identyfikacja kluczowych oddziaływań

Efekty farmakologiczne



>dokowanie: minimalizacja energii

$$\Delta G = \Delta E_{\text{vdW}} + \Delta G_{\text{desol}} + \Delta E_{\text{elec}} + \Delta G_{\text{const}}$$

ΔE_{vdW} : van der Waals energy; Shape complementarity

ΔG_{desol} : Desolvation energy; Hydrophobicity

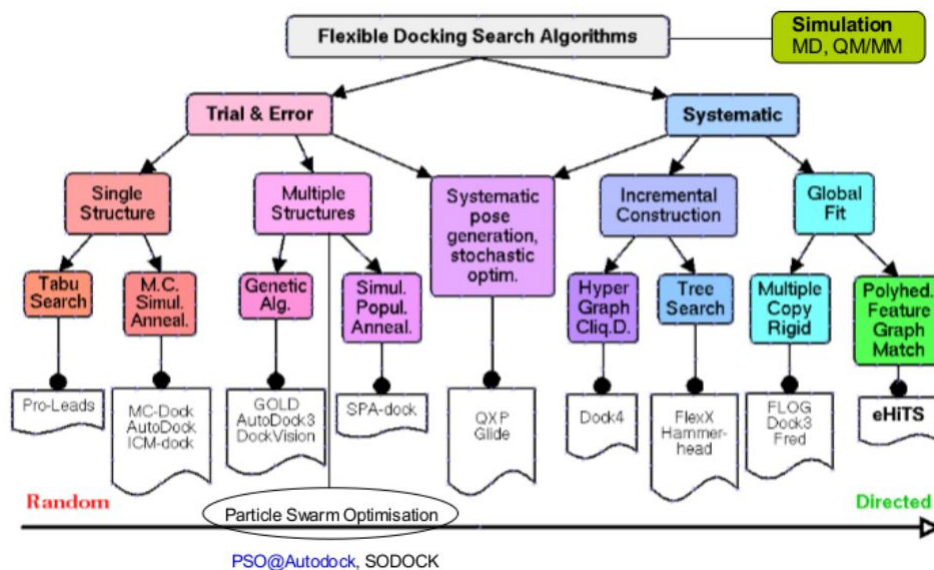
ΔE_{elec} : Electrostatic interaction energy

ΔG_{const} : Translational, rotational and vibrational free energy changes

$$\Delta G_{\text{desol}} = \sum_i N_i \Delta G_i$$

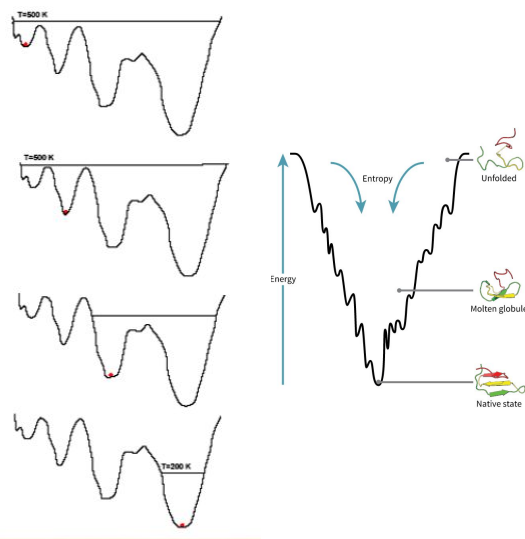
N_i : Number of atoms of type i

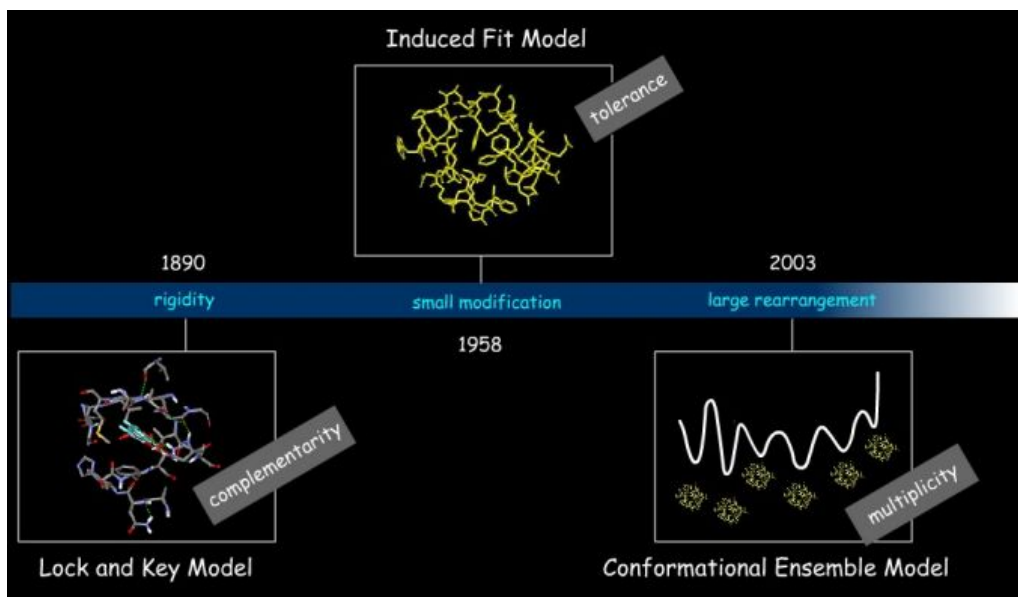
ΔG_i : Desolvation energy for an atom of type i



>dokowanie:
 minimalizacja energii
 Znalezienie globalnego minimum za pomocą technik obliczeniowych

Translacja, rotacja ligandu
 Wykorzystanie istniejącej wiedzy
 Losowe zmiany parametrów
 Ocena parametryczna

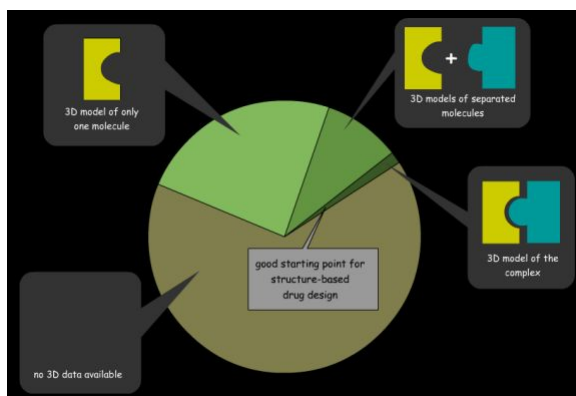




>dokowanie: podsumowanie

Ograniczenia:

Ograniczone informacje strukturalne



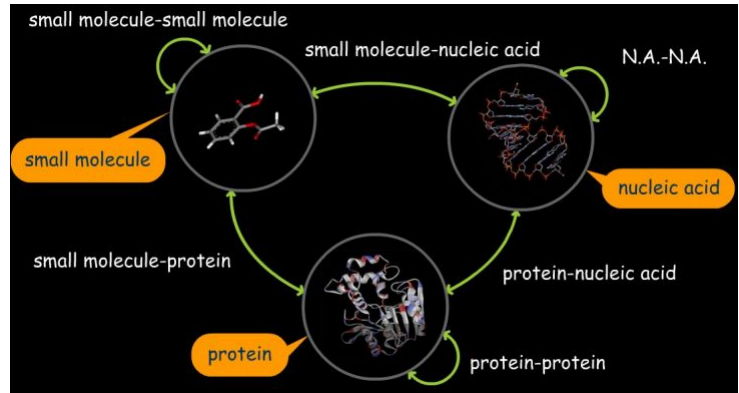
>dokowanie: klasyfikacja

Ze względu na obiekty:

Małe cząsteczki (ligandy)

Białka

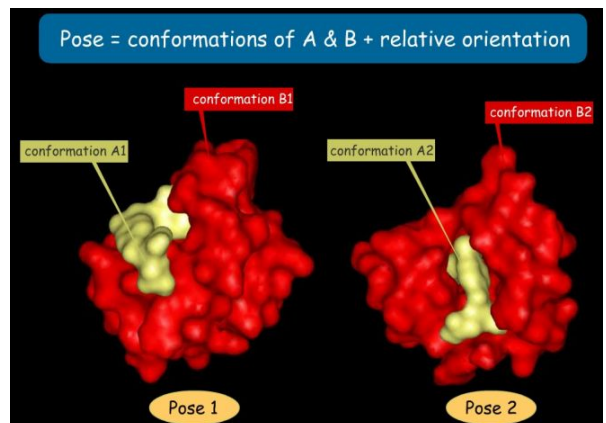
Kwasy nukleonowe



>dokowanie: efekty

Binding mode (pose)
wzajemny geometryczny
układ ligandu i receptora

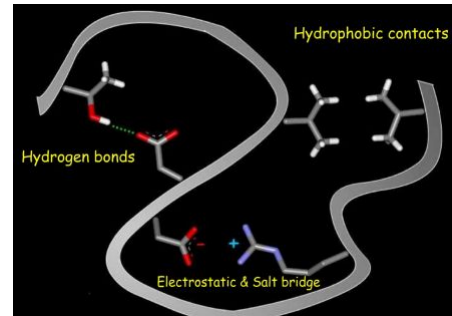
Układ opisuje relatywne
orientacje zadokowanej
cząsteczki względem
receptora uwzględniając
jego konkretną konformację



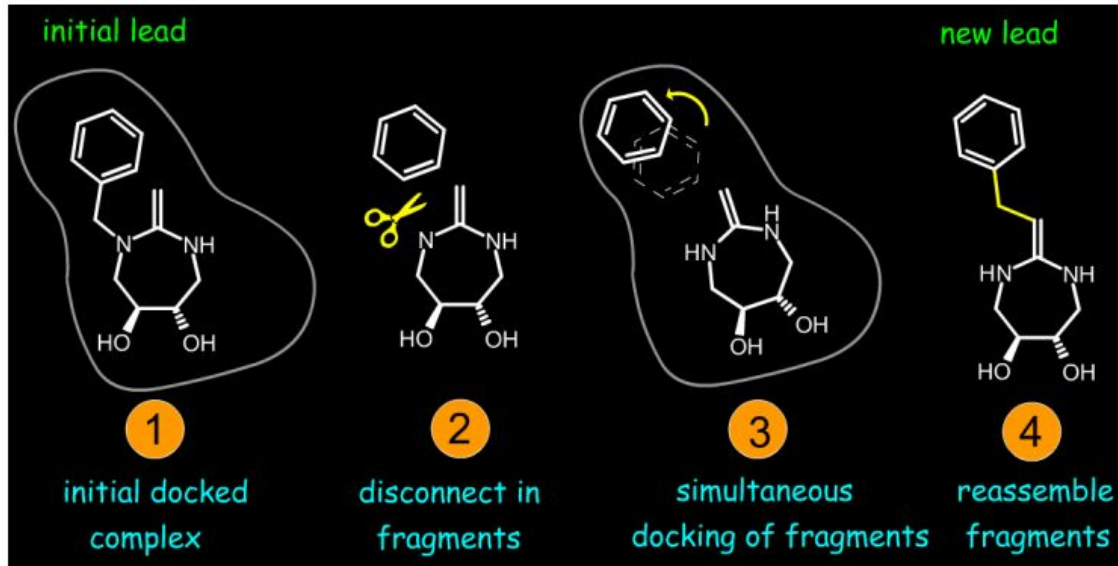
>dokowanie: ocena

OCENA EKSPERYMENTU DOKOWANIA MA CHARAKTER RELATYWNY

Zależy od zastosowanych
parametrów oraz funkcji
użytych przez wybrany
pakiet do ceny wyniku.



- WYNIKÓW DOKOWANIA OTRZYMANYCH Z DWÓCH RÓŻNYCH PAKIETÓW
NIE MOŻNA BEZPOŚREDNIO PORÓWNAĆ!

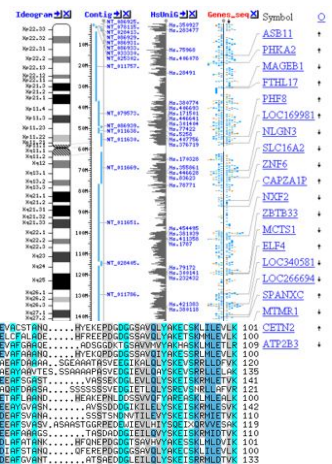
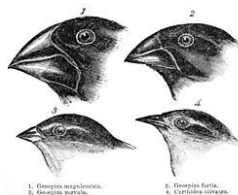


Bioinformatyka

Kombinacja informacji biologicznych i technik informacyjnych.
Analiza bioinformatyczna opiera się na analizie informacji zebranych w bazach danych w celu wyjaśnienia obserwowanych faktów doświadczalnych

Rezultatem analizy są:

- Określenie wzajemnych zależności,
- Pokrewieństwa sekwencji kodujących,
- Podobieństwa struktur,
- Profile ekspresji,
- Profile szlaków metabolicznych.



Rola komputerów

Źródłem danych są organizmy żywe.

Każdy gatunek stanowi niepowtarzalny i unikalny zbiór danych, którego wielkość jest różna w zależności.

Każda naturalna wariacja w ramach gatunku stanowi nowy zbiór danych.

Komputery są niezbędne do gromadzenia, zarządzania i przetwarzania danych w akceptowalnym zakresie czasowym.

GENBANK AND WGS STATISTICS

Release	Date	GenBank		WGS	
		Bases	Sequences	Bases	Sequences
3	Dec 1982	680338	606		
14	Nov 1983	2274029	2427		
20	May 1984	3002088	3665		
129	Apr 2002	19072679701	16769983	692266338	172768
130	Jun 2002	20648748345	17471130	3267608441	397502
131	Aug 2002	22616937182	18197119	3848375582	427771
237	Apr 2020	415770027949	216531829	7788133221338	1267547429
238	Jun 2020	427823258901	217122233	8114046262158	1302852615
239	Aug 2020	654057069549	218642238	8841649410652	1408122887

67

The collage displays several key biological databases and research portals:

- NCBI (National Center for Biotechnology Information):** Shows a COVID-19 alert and a search bar.
- EMBL-EBI (European Molecular Biology Laboratory - European Bioinformatics Institute):** Features a search bar and a 'Protein Data Bank in Europe' section.
- UniProt (Universal Protein Resource):** Shows a search bar and a 'Getting started' section.
- PDBe (Protein Data Bank in Europe):** Features a 'New PDBe-KB COVID-19 Data Portal' and a 'Featured structure' section.
- UniParc (Universal Protein Resource):** Shows a search bar and a 'Getting started' section.
- EMBL-EBI (European Molecular Biology Laboratory - European Bioinformatics Institute):** Shows a search bar and a 'Protein Data Bank in Europe' section.

68

Omica

Dane omiczne to dane ze specyficznych eksperymentów naukowych padające poszczególne aspekty aktywności biologicznej na różnych jej poziomach molekularnych:

- Genomika: DNA
- Proteomika: białka
- Metabolomika: małocząsteczkowe związki chemiczne (np. pochodzące ze szlaków metabolicznych)

69 /

Proteom

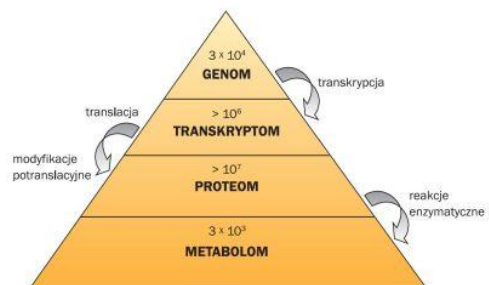
Proteom
component białkowy komórek kodowany przez genom.

Stężenie białek w komórce...

- nie jest prostą funkcją ekspresji genów
- podlega licznym modyfikacjom, które decydują o końcowych właściwościach.

W połączeniu z genomem pozwala:

- identyfikować szlaki metaboliczne zaangażowane w badane procesy
- monitorować procesy patologiczne na poziomie komórkowym

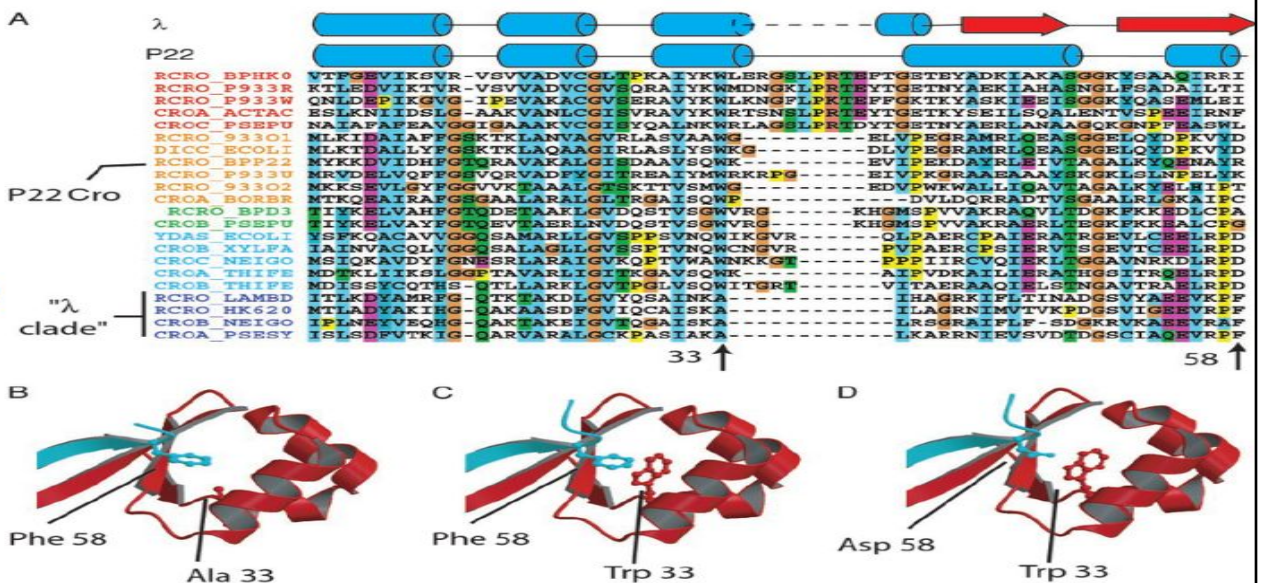
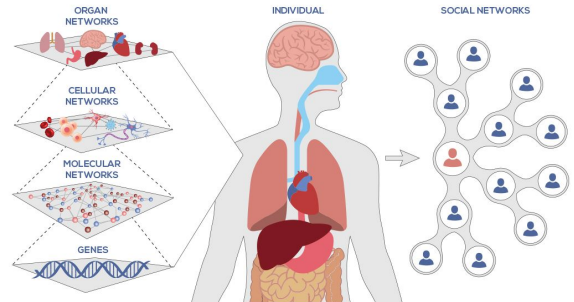


Genom	Proteom	Metabolom
statyczny	dynamiczny	dynamiczny
możliwość amplifikacji	brak	brak
jednorodny	niejednorodny	niejednorodny
stałe stężenie	zmiennie stężenie	zmiennie stężenie

Biologia systemów

...to całościowe, holistyczne, spojrzenie na funkcjonowanie komórek czy organizmów opierające się na analizie wycinkowych prac badawczych umożliwiające modelowanie układów biologicznych.

Biologia systemów wykorzystuje w równym stopniu informacje genetyczne, proteomiczne i metabolomiczne poprzez analizę funkcjonalną (analizę funkcji cząsteczek).



(A) Alignment of Cro sequences generated from sequence similarity searches of microbial genomes using previously annotated Cro sequences. Alignment corresponds to residues 1 – 58 of P22 Cro and 5 – 58 of λ Cro. Sequences are named as described in Materials and Methods. RCRO λ LAMBD, λ Cro; RCRO λ BPP22, P22 Cro. (B) Portion of λ Cro showing ball-and-socket dimer interface and Ala-33 and Phe-58 side chains. Different monomers are shown in blue and red. (C) Model of λ Cro-A33W showing Trp-33 occupying a portion of the hydrophobic socket and clashing sterically with the Phe-58 ball. (D) Model of λ Cro-A33W F58D.

Dopasowanie sekwencji

Warianty:

- identyczne sekwencje
- sekwencje z błędami
(ang. *match & mismatch*)
- insercje i delecje
- konserwacja

```
ATGGCATGCCTGATC
|||||
ATGGCATGCCTGATC
```

```
MSTPAGSDQERMILV
|||||
MSTPAGSDQERMILV
```

```
ACGGCTTACCTGGCC
|||||
ATGGCATGCCTGATC
```

```
MGTPAASFOERMSTV
|||||
MSTPAGSDQERMILV
```

```
ATGGC-TGCCGATC
|||||
ATGGCATGCC-GATC
```

```
MSTPA-SDWERMILV
|||||
MSTPAGSDQE-MILV
```

```
MSKMAGSNVERMILV
|||.|||.:.|||.
MSRVAGSDIERMIMV
```

Typy dopasowań

Needleman and Wunsch (1970)

Global alignment



Local alignment



Smith and Waterman (1981)

podobieństwo i homologia sekwencji

- Każda para sekwencji będzie wykazywać pewien poziom podobieństwa.
- Podobieństwo nie świadczy o wspólnym pochodzeniu lub właściwościach – może mieć ono charakter losowy.
- Sekwencje homologiczne to sekwencje pochodzące od wspólnego przodka.
Dwie lub więcej sekwencji nigdy nie będzie homologiczne w 50%.
Dwie lub sekwencji może posiadać 50 % podobieństwa co może świadczyć o homologii pomiędzy nimi.
- Ortologi
dwa homologiczne geny występujące w odrębnych gatunkach posiadające te same funkcje.
- Paralogi
dwa homologiczne geny u tego samego (lub odrębnego) gatunku posiadające różne funkcje.

Zmiany ewolucyjne

Mutacje (zmiana tożsamości, usunięcie lub dodanie aminokwasu/domeny) na poziomie genetycznym zachodzą tylko w określonych warunkach i gdy nie wpływają na strukturę lub funkcję kodowanego białka

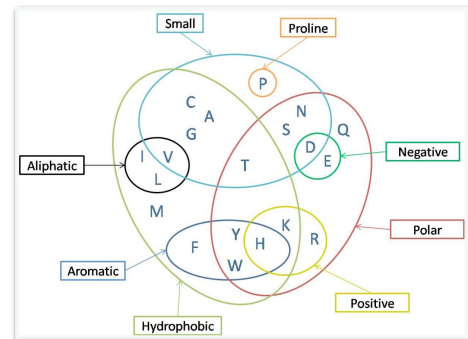
Problem: jak ocenić podobieństwo sekwencji uwzględniając zmiany ewolucyjne?

Ewolucja białek

Zmiany ewolucyjne mogą zachodzić w dwóch obszarach sekwencji:

- *kluczowym*, czyli obszarze odpowiedzialnym za konkretną funkcję biologiczną;
- *powłoce*, czyli obszarze odpowiedzialnym za uformowanie struktury białka ale nie biorącym bezpośredniego udziału w katalizie.

Zachodzące mutacje w obszarze kluczowym mają tendencję do zamiany aminokwasów na inne o podobnych właściwościach fizykochemicznych. Np. zamiana hydrofobowych aminokwasów w rdzeniu białka (Leu, Ile, Val).



blosum (blocks substitution matrix)

Macierze BLOSUM zostały opracowane w oparciu o analizę lokalnych dopasowań spokrewnionych fragmentów sekwencji.

W wyniku analizy spokrewnionych i zakonserwowanych sekwencji nie zawierających przerw w bazie danych BLOCKS ustalono częstotliwość występujących mutacji. Następnie dokonano obliczenia odpowiednich parametrów dla wszystkich potencjalnych 210 zmian dla wszystkich standardowych 20 aminokwasów.

Macierze BLOSUM zostały wprowadzone przez Steven Henikoff i Jorja Henikoff w 1992 r.

Ala	4																			
Arg	-1	5																		
Asn	-2	0	6																	
Asp	-2	-2	1	6																
Cys	0	-3	-3	-3	9															
Gln	-1	1	0	0	-3	5														
Glu	-1	0	0	2	-4	2	5													
Gly	0	-2	0	-1	-3	-2	-2	6												
His	-2	0	1	-1	-3	0	0	-2	8											
Ile	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
Leu	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
Lys	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
Met	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
Phe	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
Pro	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
Ser	1	-1	1	0	-1	0	0	0	-1	-2	0	-1	-2	-1	4					
Thr	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	1	5				
Trp	-3	-3	-4	-4	-2	-2	-3	-2	-3	-2	-3	-1	1	-4	-3	-2	11			
Tyr	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-2	1	3	-3	-2	2	7			
Val	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	0	-3	-1	4	

Query sequence: PQGEFG

Word 1: PQG

```
;LCBO - Prolactin precursor - Bovine
; a sample sequence in FASTA format
MDSKGSSQKGRSRLLLLVSNLLCQGVSTPVPNCPGNGQVSLRDLFRAVMVSHYIHDLS
EMFNEFKRYAQKGFTIMALLNSCHTSSLPTEDEKQAQQTHIEVLMSLILGLRWSNDPLVHL
VTEVRMGKAPDAILSRATIEENKRLLEGMEIFGQVPGAKETEPYVWSGLSPDKTDED
ARYSAFYNLHLCLRDSSKIDTYLLKLCRLCIINNIC*
```

- >MCHU - Calmodulin - Human, rabbit, bovine, rat, and chicken
ADQLTEEQIAEFKEAFSLFDKDGDTITTKELGTVMRSLQGNPTAEALQDMINEVDADNGNTID
FPEFLTMMARKMKDITDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEKLTDEEVDEMIREA
DIDGDGQVNYEEFVQMMTAK*

>gl|5524211|gb|AA444166.1| cytochrome b [Elephas maximus maximus]
 GLCYTHGIRNIYGSVYSLTWTNTGIMLLITMAFGVYVWPGNSFWGATVITNLFSAIPYITGNLV
 EIWJGSDVFNKATLNRFFAFHFTLPTVMAALAGHTLFLHETGSGNGLTDSDDIKPFHPYVITKDFLG
 LLTITLLLLALLSPDMLGDPDNIHMPADNPITPLHIKPEWYFLFAYAIRSVPNKLGGVLALFLSIVIL
 GLMPFLHTSKHSRSMMLRPSQLFWITLMDLLTITWIGSQPVEYPTYIIGQMASIYFLSITLAFPIAGX
 IENY

[illegible]

W przypadku dopasowania wielu sekwencji jest to nie możliwe!

Ocena jakości dopasowania wielu sekwencji jest w gestii badacza.

Czy kolejność wprowadzenia sekwencji w dopasowaniu będzie mieć znaczenia?

czy kolejność wprowadzenia sekwencji w dopasowaniu będzie mieć znaczenia?

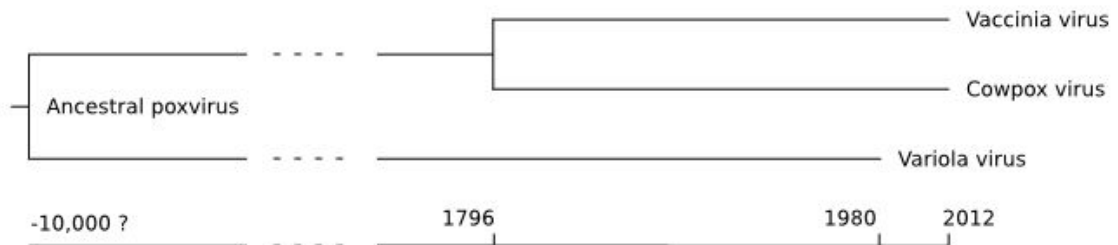
Im więcej sekwencji tym większy 'ciężar' poprawnie dopasowanych fragmentów sekwencji.

Bez interwencji badacza automatyczna korekta dopasowanie jest niezwykle trudna.



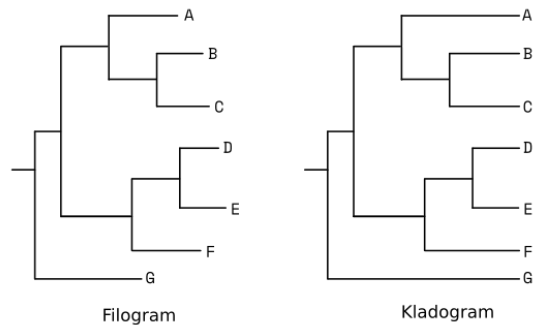
SeqA GARFIELD THE LAST FA-T CAT
SeqB GARFIELD THE FAST CA-T ---
SeqC GARFIELD THE VERY FAST CAT
SeqD ----- THE ---- FA-T CAT

graficzna reprezentacja zmian ewolucyjnych

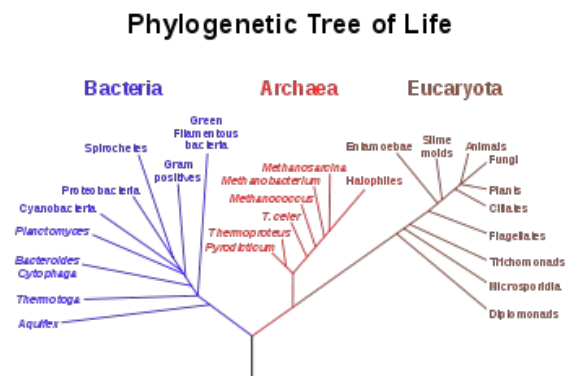
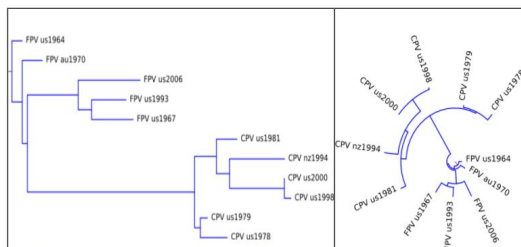


drzewo filogenetyczne | topologia

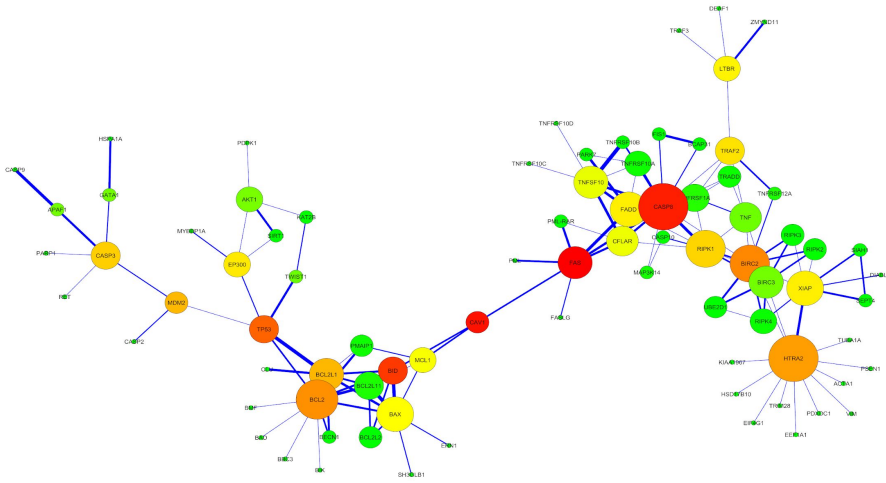
- Gdy długość gałęzi drzewa odzwierciedla odległość ewolucyjną badanych sekwencji, wtedy drzewo nazywamy **filogramem**.
- **Kladogram** natomiast pokazuje jedynie pokrewieństwa między badanymi taksonami.



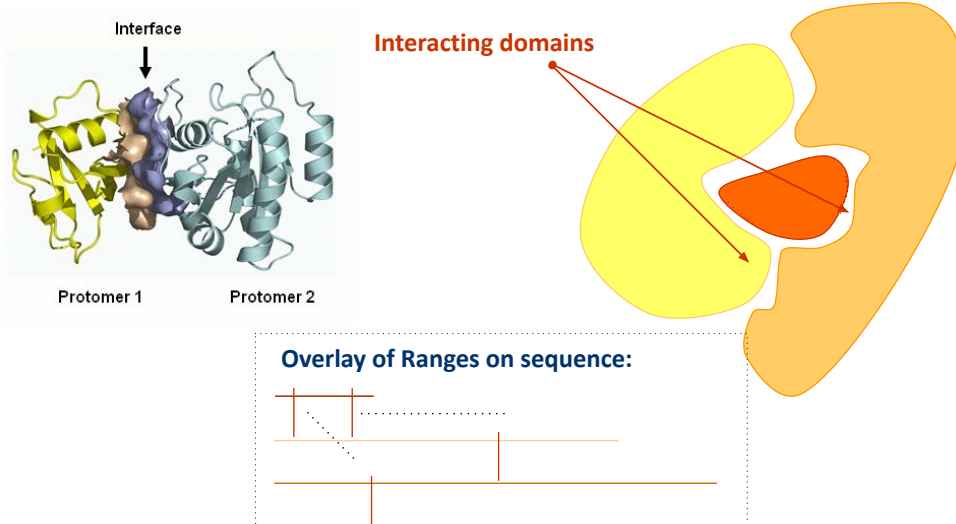
drzewa filogenetyczne | reprezentacja



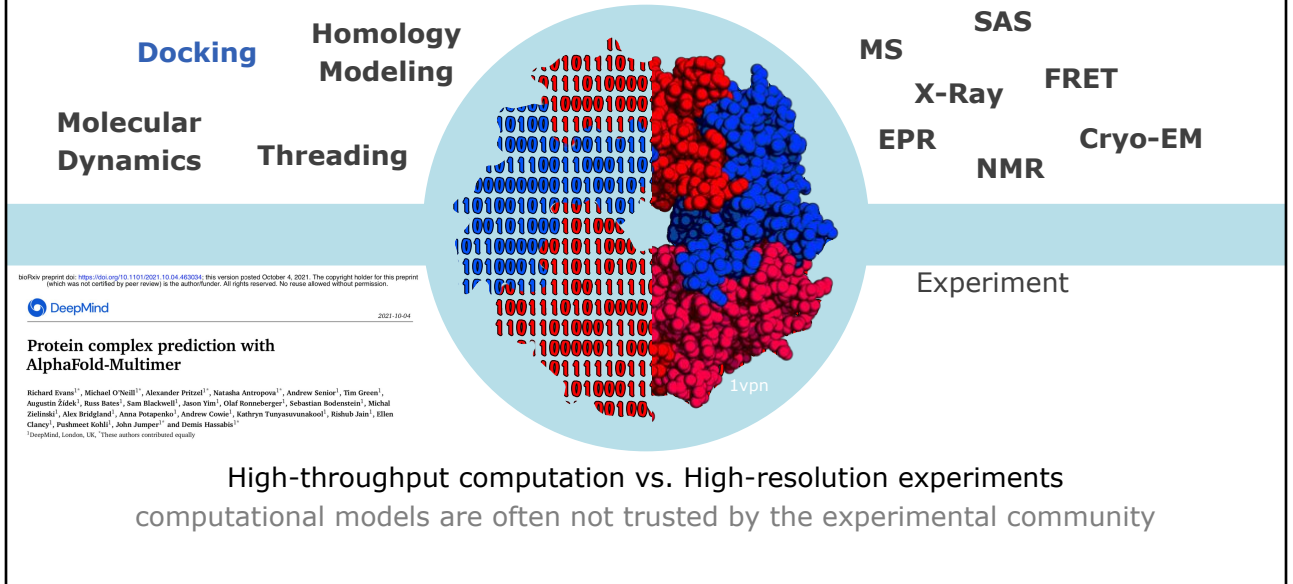
PPI: efekt centralizacji



Representing protein-protein interactions: interacting domains



Biologia strukturalne | interakcje



bioRxiv preprint doi: <https://doi.org/10.1101/2020.10.04.403014>; this version posted October 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



2021-10-04

Protein complex prediction with AlphaFold-Multimer

Richard Evans¹, Michael O'Neill¹, Alexander Pritzel¹, Natasha Antropova¹, Andrew Senior¹, Tim Green¹, Augustin Zidek¹, Russ Bates¹, Sam Blackwell¹, Jason Yim¹, Olaf Ronneberger¹, Sebastian Bodenstein¹, Michal Zielinski¹, Alex Bridgland¹, Anna Piatkowski¹, Andrew Cowie¹, Kathryn Tunyasuvunakool¹, Rishabh Jain¹, Ellen Clancy¹, Pushmeet Kohli¹, John Jumper¹ and Demis Hassabis¹

¹DeepMind, London, UK. These authors contributed equally