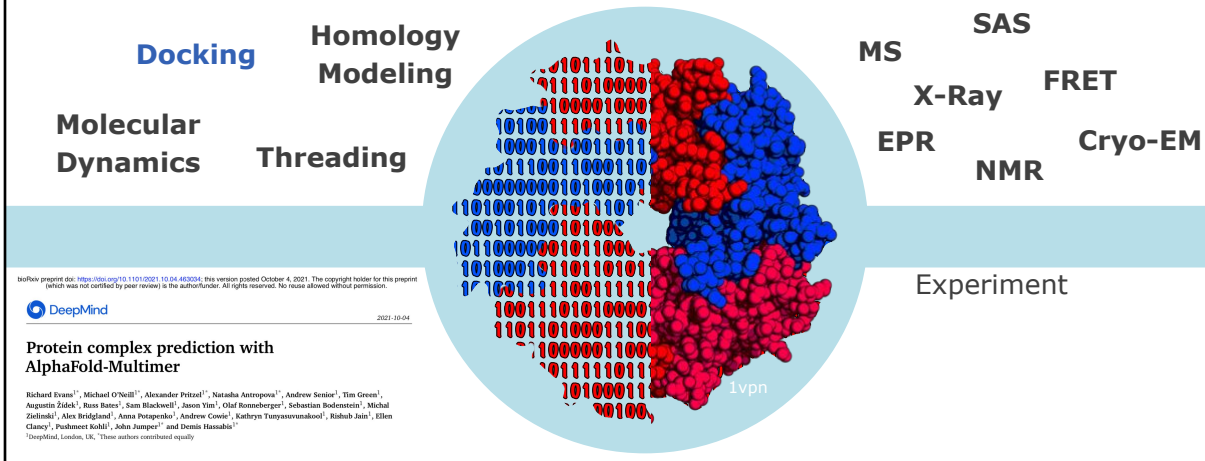
# Bioinformatyka

Andrzej Łyskowski, dr inż.
Katedra Biotechnologii i Bioinformatyki

andrzej.lyskowski@prz.edu.pl
H-237

## Biologia strukturalne | interakcje






DeepMind

2021-10-04

**Protein complex prediction with AlphaFold-Multimer**

Richard Evans[1*], Michael O'Neill[1*], Alexander Pritzel[1*], Natasha Antropova[1*], Andrew Senior[1], Tim Green[1], Augustin Žídek[1], Russ Bates[1], Sam Blackwell[1], Jason Yim[1], Olaf Ronneberger[1], Sebastian Bodenstein[1], Michal Zielinski[1], Alex Bridgland[1], Anna Potapenko[1], Andrew Cowie[1], Kathryn Tunyasuvunakool[1], Rishub Jain[1], Ellen Clancy[1], Pushmeet Kohli[1], John Jumper[1*] and Demis Hassabis[1*]

[1]DeepMind, London, UK, [*]These authors contributed equally

High-throughput computation vs. High-resolution experiments

computational models are often not trusted by the experimental community

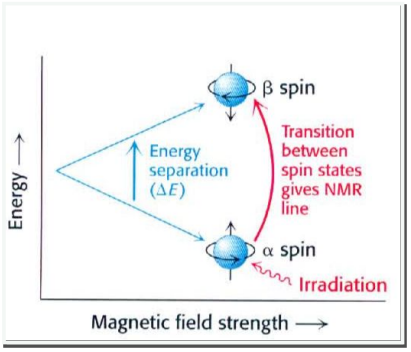
# Spektroskopia magnetycznego rezonansu jądrowego (NMR)

Podstawa:

niektóre jądra atomowe wykazują właściwości magnetyczne.

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

rozszczepienie energii jest zależne od środowiska.



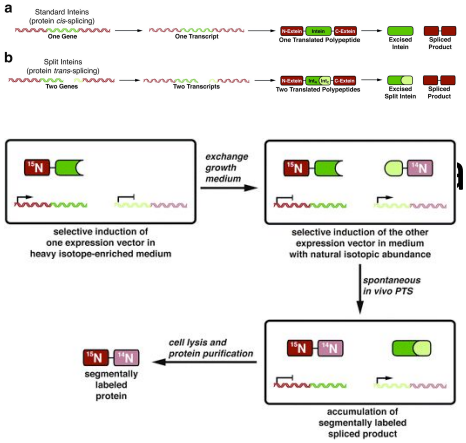

Naturalnie występujące izotopy pierwiastków o znaczeniu biologicznym.

| Nucleus | Natural abundance (% by weight of the element) |
|---|---|
| $^{1}H$ | 99.984 |
| $^{2}H$ | 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.

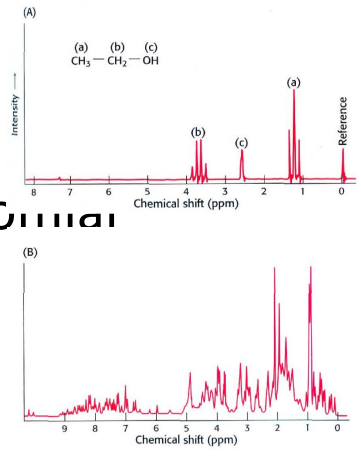
zastosowanie pożywek o zmodyfikowanym składzie izotopowym oraz odpowiednich szczepów ekspresyjnych.

znakowanie domen.




Stężona próbka białka:

1 mM lub 15 $mg/ml^{-1}$ dla białka 15 kDa



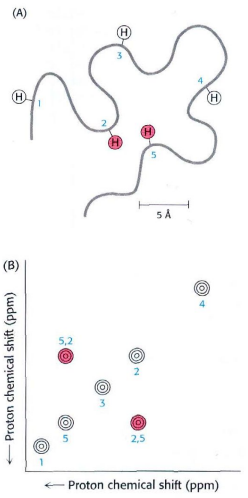
# Pomiar

Uproszczenie widma NMR osiąga się poprzez zastosowanie odpowiednich metod pomiarowych:

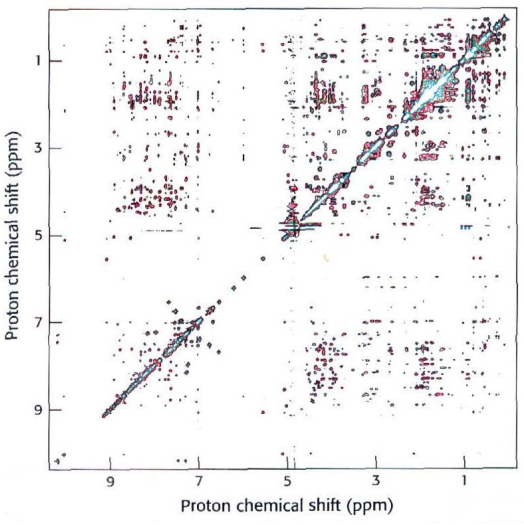
jądrowego efektu Overhausera (NOE) w technice spektroskopii jądrowego efektu Overhausera (NOESY);

spektroskopii korelacyjnej (COSY);

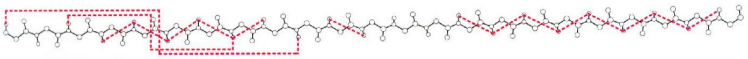
całkowitej spektroskopii korelacyjnej (TOCSY)




# Analiza



# Analiza

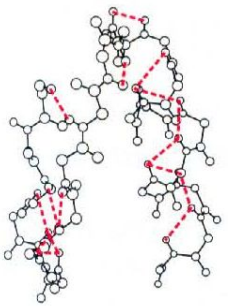


W trakcie interpretacji widm (ręcznej lub automatycznej) stosuje się następujące typy ograniczeń:

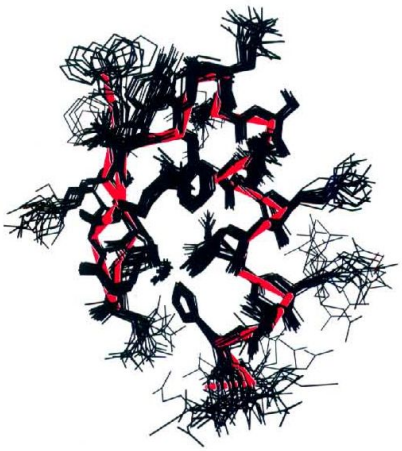
odległości;

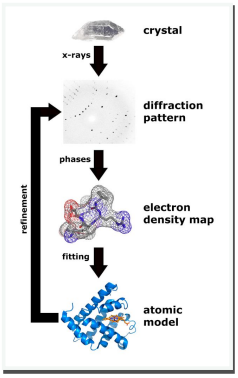
kątów;

orientacji przestrzennej.



Wynikiem pomiarów jest rodzina struktur.

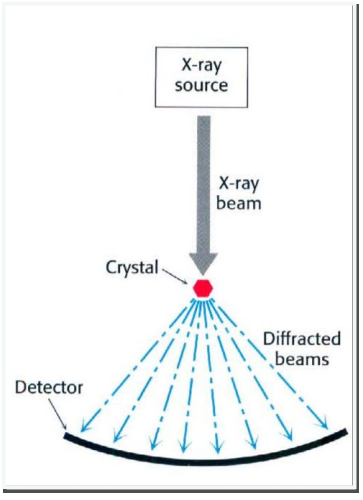





# Etapy wyznaczania struktury

## Krystalografia rentgenowska

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



# Kryształ

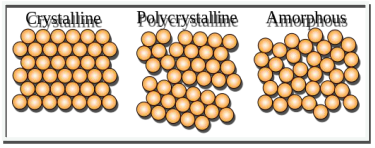
Kryształ

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

Krystalizacja białek

proces poszukiwanie selektywnych, indywidualnych warunków krystalizacji białka poprzez:

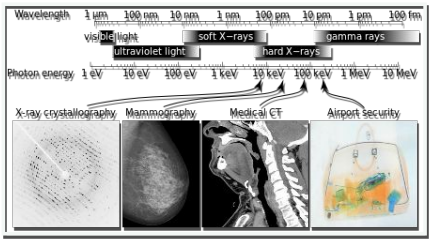
- wysalanie;
- metody przesiewowego badania różnych warunków.

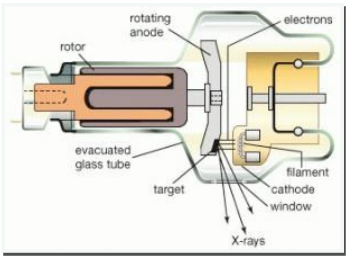


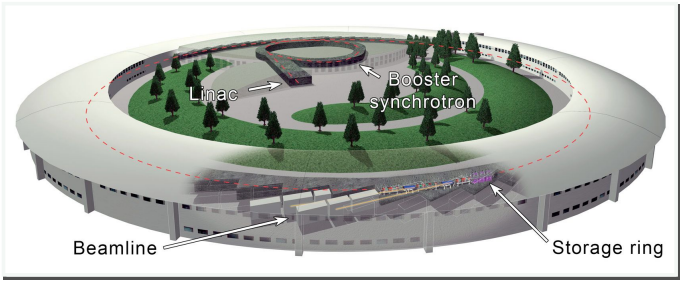

# Promieniowanie rentgenowskie

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

*Dlaczego nie można stosować światła widzialnego?*

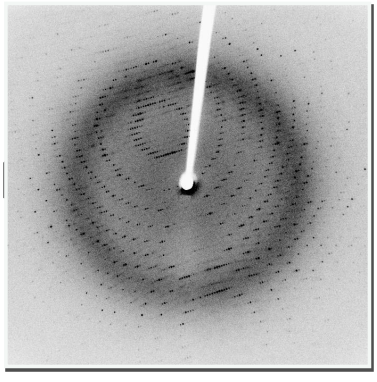




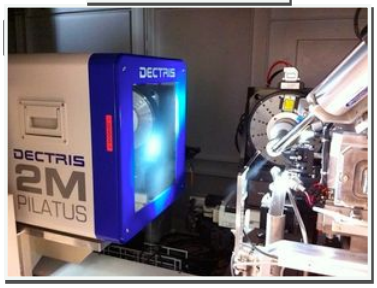




Promieniowanie rengenowskie ulega rozszczepieniu na elektronach atomów.

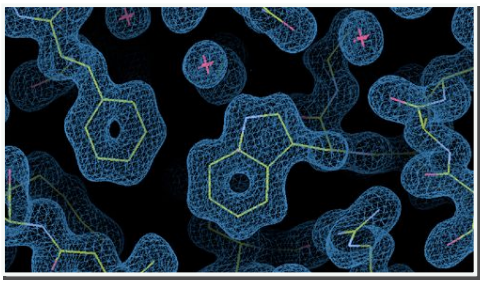
Rozproszone fale nakładają się tworząc obraz interferencyjny, na którym wzmocnienie lub wygaszenie fali jest zależne od rodzaju i położenia atomu w krysztale.



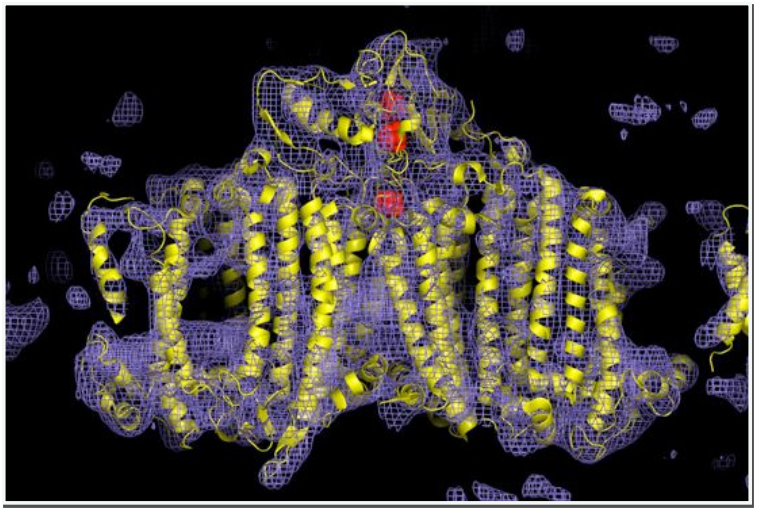
Historycznie rolę detektora spełniała klisza fotograficzne.




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

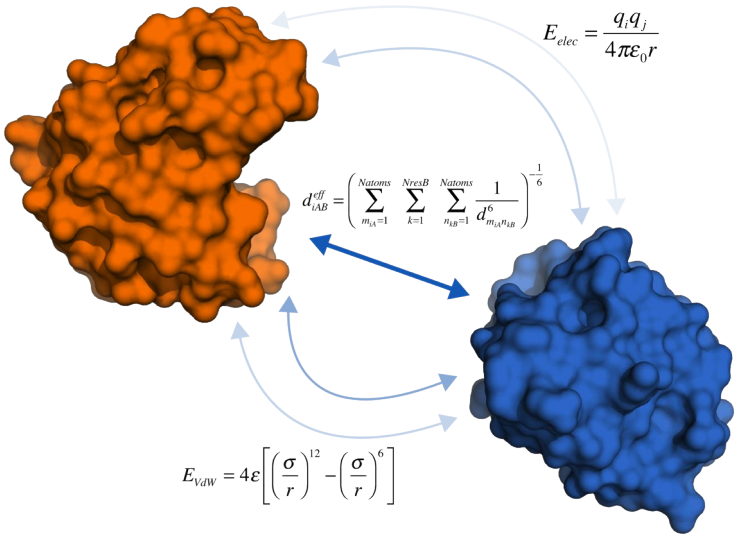


# Struktura 3D

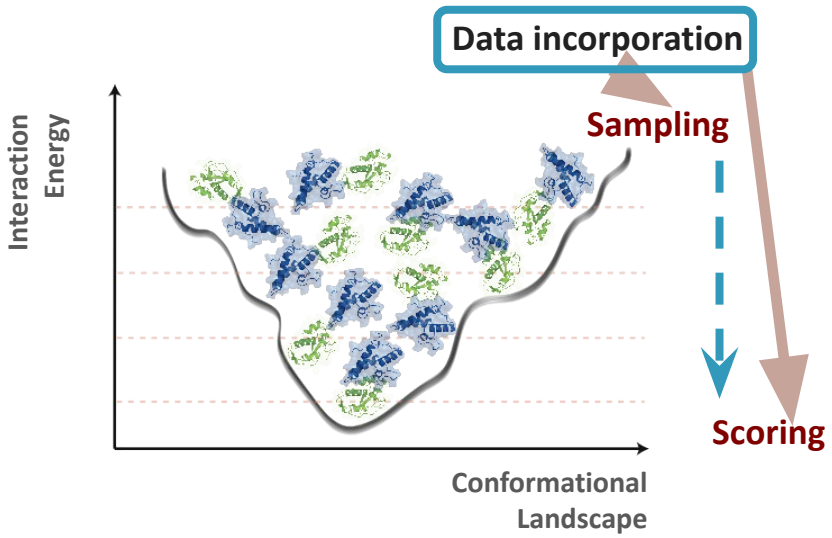


- Kryształ *vs.* roztwór.
- Pojedyncza stuktura *vs.* rodzina.
- Statyczny obraz *vs.* obraz dynamiczny.
-
- Dowolny rozmiar cząsteczki *vs.* limit masy cząsteczkowej.

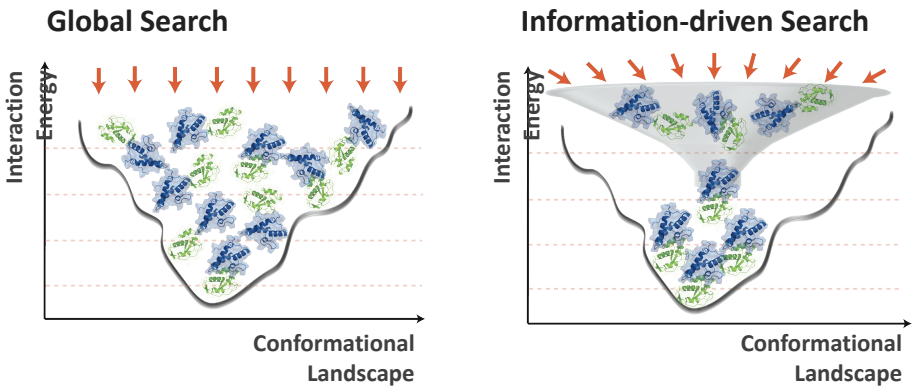
# Molecular Docking



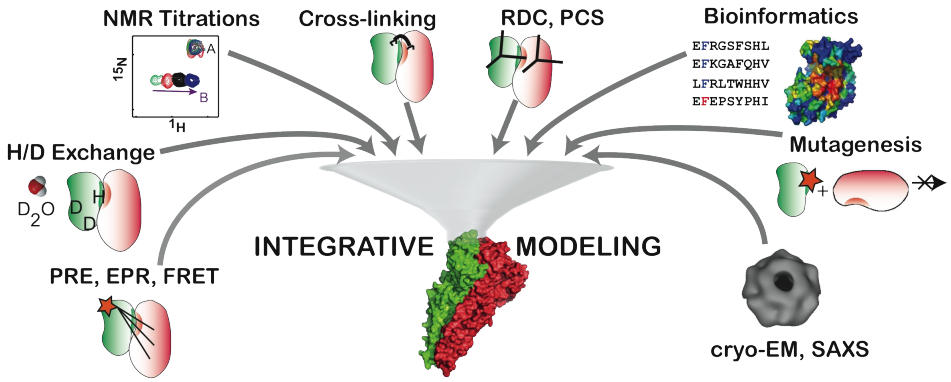

# Metodologia



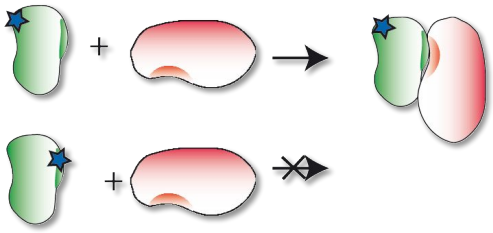
# Integracja danych




# Kompleksowe modelowanie



# Experimental sources:
## mutagenesis



**Advantages/disadvantages**

**+ Residue level information**

**- Loss of native structure should be checked**

**Detection**

**- Binding assays**

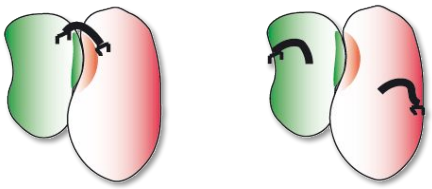
**- Surface plasmon resonance**

**- 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**

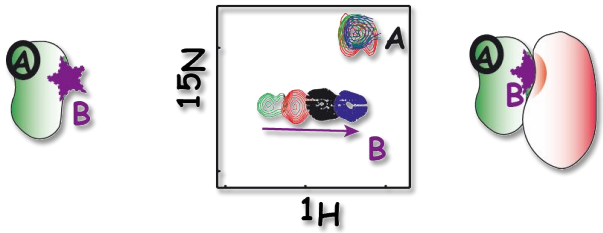
**- Labeling needed for NMR**

**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**

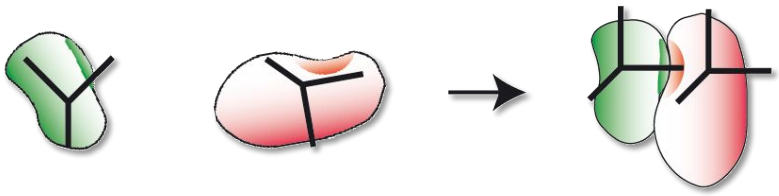
**- Direct vs indirect effects**

**- 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**

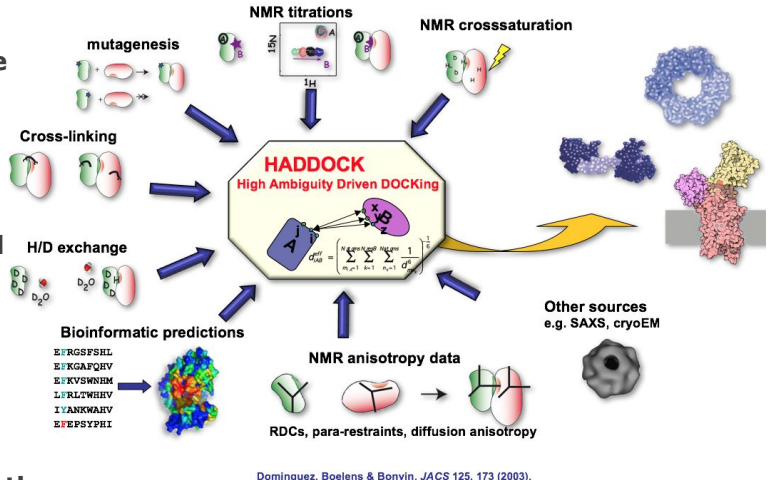
**Capable of docking up to 20 molecules (2.4 version)**

**Symmetries can be leveraged**

**Allows for flexibility at the interface**

**Final flexible refinement in explicit solvent**

**Consistent performance over the years in CAPRI**



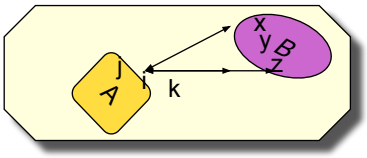

Dominguez, Boelens & Bonvin. *JACS* 125, 173 (2003).

http://www.bonvinlab.org/software

# Data-driven docking with HADDOCK

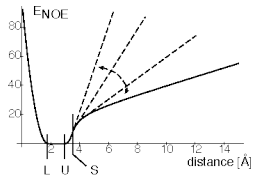
**List of interface residues for protein A**

**List of interface residues for protein B**




**Effective distance $d_{iAB}^{eff}$ calculated as**

$$d_{iAB}^{eff} = \left( \sum_{m_{iA}=1}^{N_{atoms}} \sum_{k=1}^{N_{resB}} \sum_{n_k=1}^{N_{atoms}} \frac{1}{d_{mn_k}^6} \right)^{-\frac{1}{6}}$$




**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**

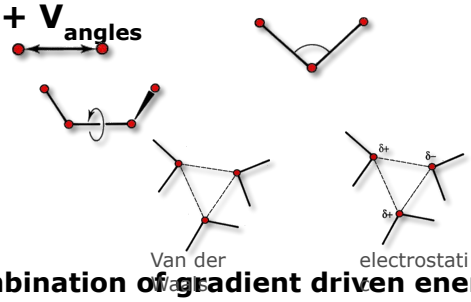
(Nilges & Brunger 1991)

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

# Searching the interaction space in HADDOCK

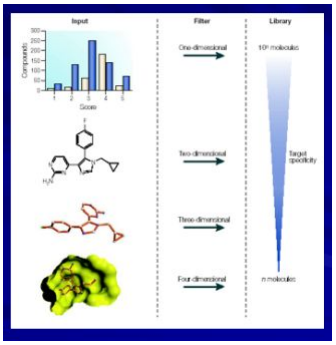
**Experimental and/or predicted information is combined with an empirical force field into an energy function whose minimum is searched for**

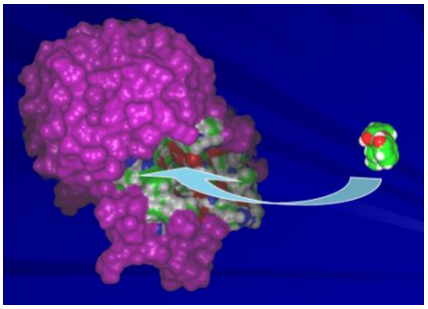
$$V_{potential} = V_{bonds} + V_{angles} + V_{torsion} + V_{non\text{-}bonded} + V_{exp}$$

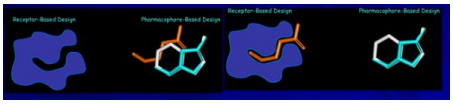



**Search is performed by a combination of gradient driven energy minimization and molecular dynamics simulations**

# Projektowanie leków oparte o receptor





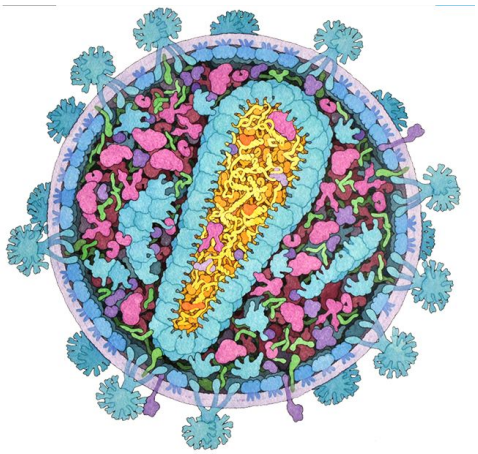

# >projektowanie leków oparte o receptor

## 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

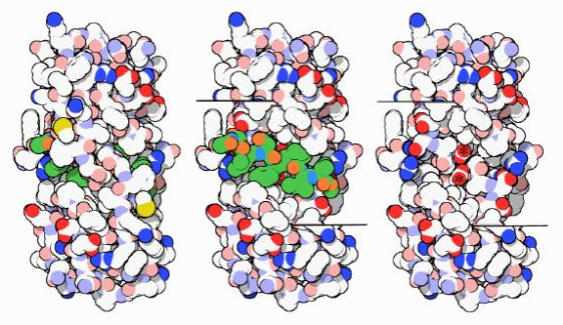
Zoptymalizuj interakcje

Wiązania wodorowe

Oddziaływania hydrofobowe

Zawadę steryczną

Oceń

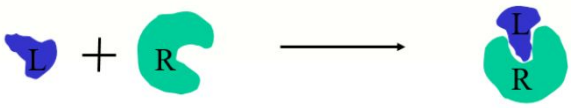


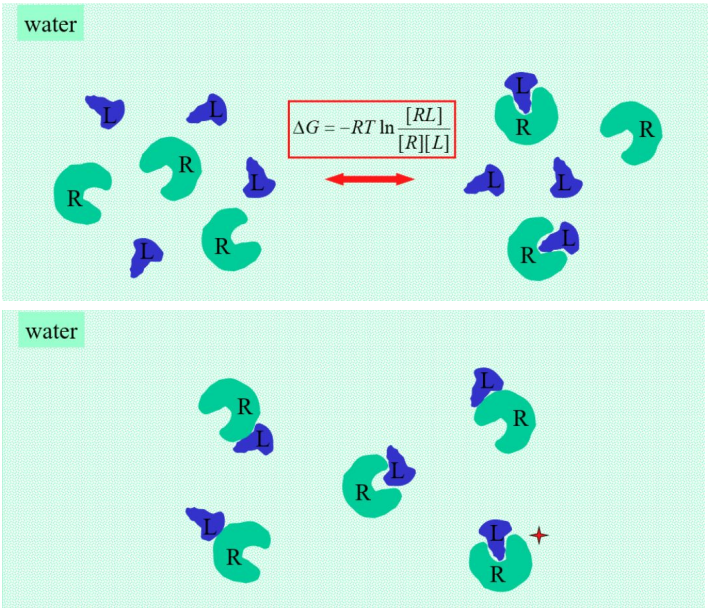
## >dokowanie

# Cel:

wyznaczenie za pomocą metod obliczeniowych kompleksu ligand-receptor.

# Aktywność biologiczna oparta jest o specyficzne odziaływania






## >dokowanie: przygotowanie modelu

# Punkt wyjścia:

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

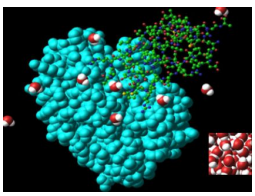
Ograniczona przestrzeń dokowania

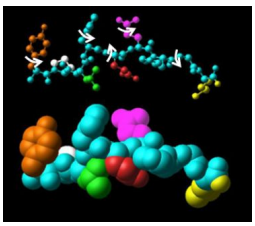
Znana ‚energia'

Brak ‚giętkości'

‚Opisana' natura oddziaływań

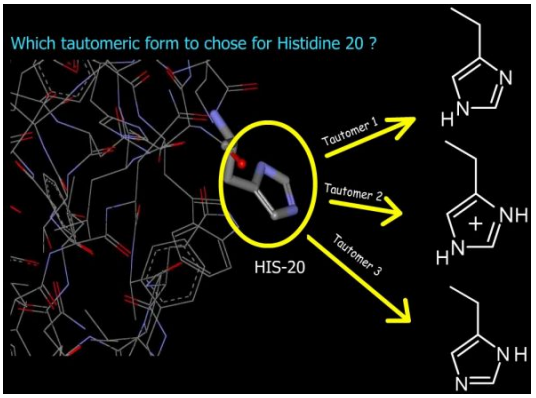
Struktura kompleksu nie jest znana (*ang. unbound docking)*





## >dokowanie: przygotowanie receptora

Każda symulacja wymaga precyzyjnego określenia parametrów startowych. Właściwości aminokwasów zmieniają się wraz z parametrami środowiska.

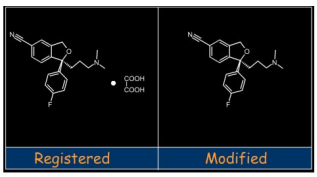



## >dokowanie: przygotowanie ligandu

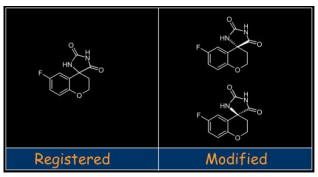
Struktura ligandu musi zawierać:

*wyłącznie ligand*

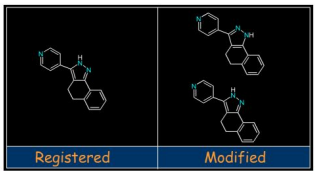
*wszystkie wariacje strukturalne (enancjomery, izomery cis/trans, tautomery, warianty ‚H')*

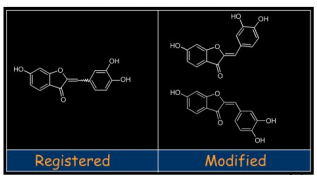



Tautomers generation




Double bond cis/trans isomers generation



## >dokowanie: optymalizacja

# Dokowanie:

Statyczne *(ang. static)*

Dynamiczne *(ang. flexible)*

Uwzględnia zmiany w strukturze 3D ligandu i receptora

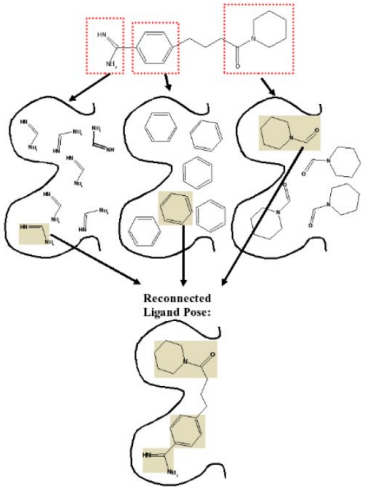
*Zmiany strukturalne można oszacować za pomocą:*

*Dynamiki molekularnej i minimalizacji energii kompleksów*

*Metod Monte Carlo*

*Bibliotek (rotamerów, białek (NMR))*

*Lokalnych potencjałów*



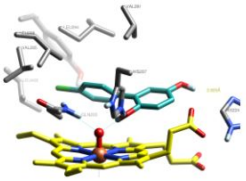

## >dokowanie: wizualizacja

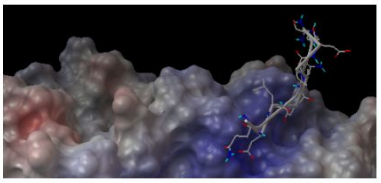
# Oddziaływanie ligand-recetor można analizować na różnych poziomach:

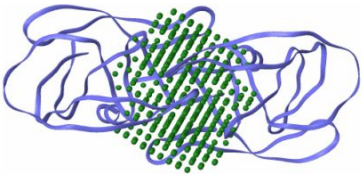
Atomowym

Powierzchni initerakcji

Siatki *(ang. grid)*





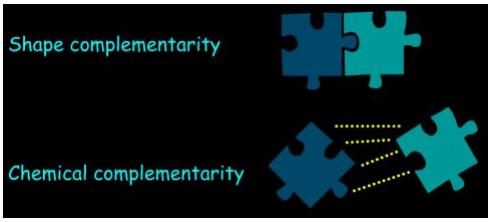


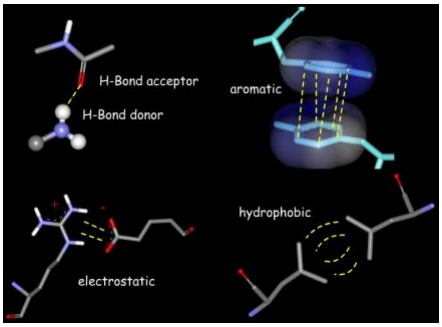
## >dokowanie: interpretacja wizualizacji

# Interpretacja komplementarności oddziaływań ligand-receptor:

Na poziomie kształtu

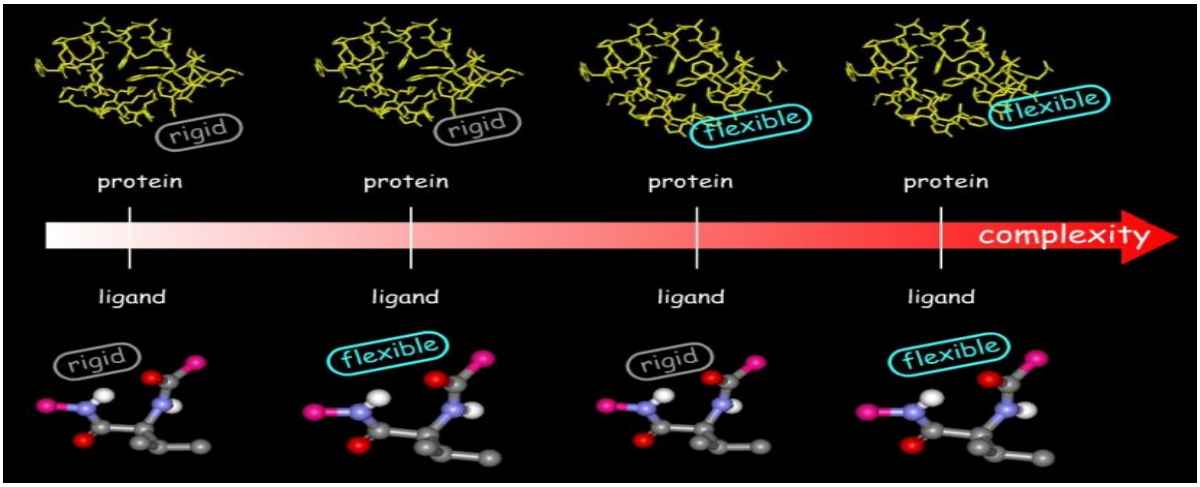
Na poziomie właściwości fizyko-chemicznych




## >dokowanie: strategie

Elastyczny ligand | Swobodne wiązania
Zmiany w entropii | Swobodne wiązania
Solwatacja | Interakcje ligand-receptor
Interakcje związków małocząsteczkowych (woda, jony, itp.)
Tautomery
Elastyczność białek | Wymuszone dopasowanie
Specyficzność wiązania | Identyfikacja kluczowych oddziaływań
Efekty farmakologiczne




# >dokowanie: minimalizacja energii

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

$\Delta E_{vdW}$: van der Waals energy; Shape complementarity

$\Delta G_{desol}$: Desolvation energy; Hydrophobicity

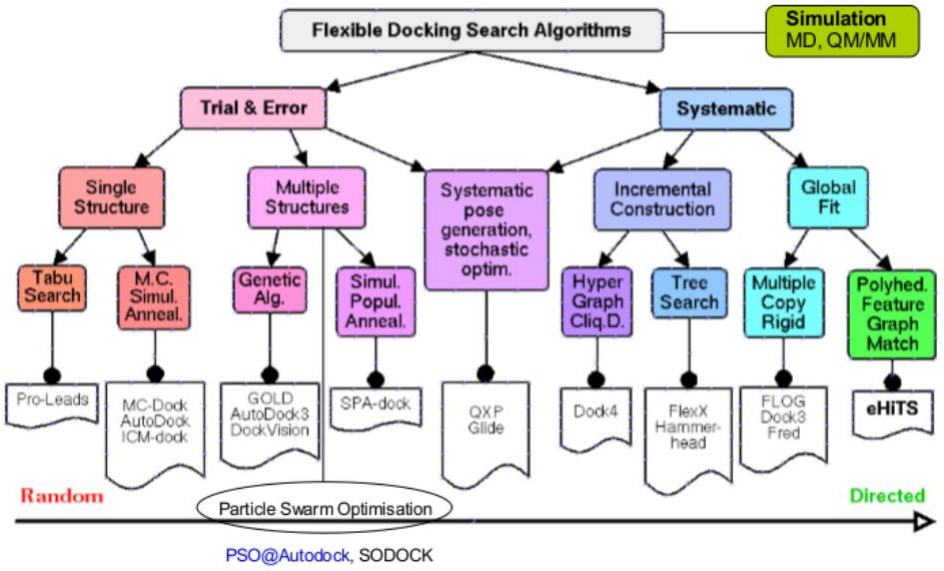
$\Delta E_{elec}$: Electrostatic interaction energy

$\Delta G_{const}$: Translational, rotational and vibrational free energy changes

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

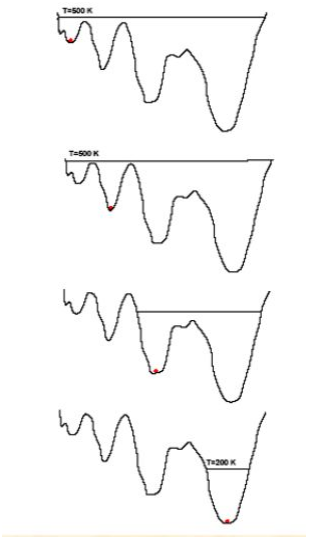
$N_i$: Number of atoms of type $i$

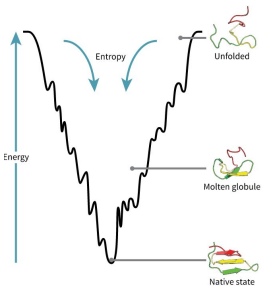
$\Delta G_i$: Desolvation energy for an atom of type $i$

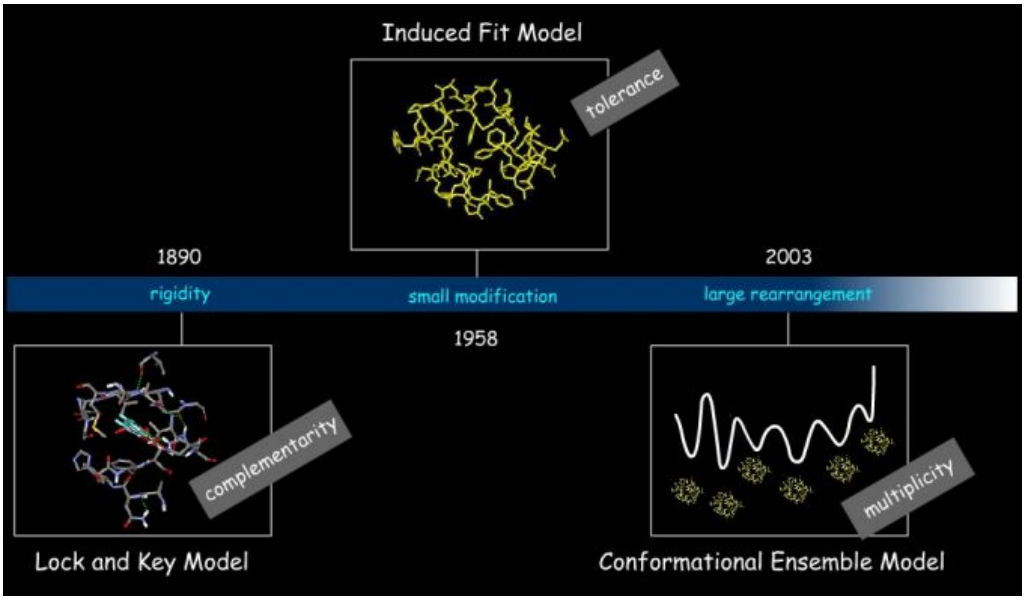



>dokowanie:
minimalizacja energii
Znalezienie globalnego minimum za pomocą technik obliczeniowych

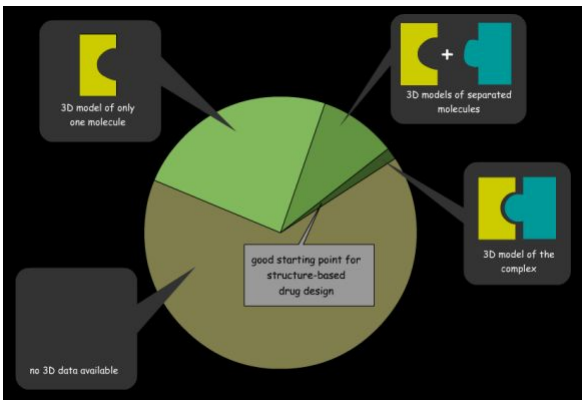
Translacja, rotacja ligandu
Wykorzystanie istniejącej wiedzy
Losowe zmiany parametrów
Ocena parametryczna






## >dokowanie: podsumowanie




# Ograniczenia:

Ograniczone informacje strukturalne

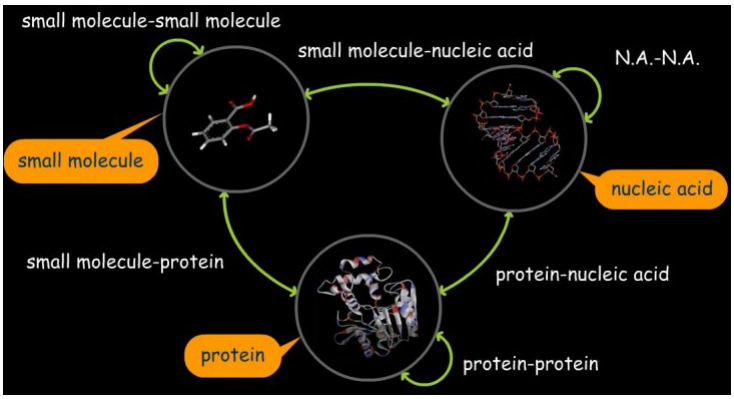
## >dokowanie: klasyfikacja

# Ze względu na obiekty:

Małe cząsteczki (ligandy)

Białka

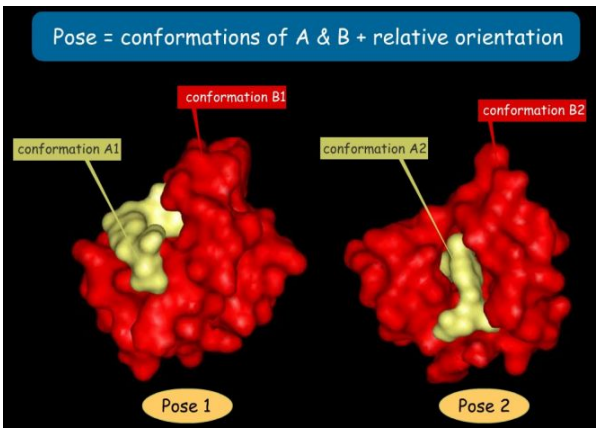
Kwasy nukleionowe




## >dokowanie: efekty

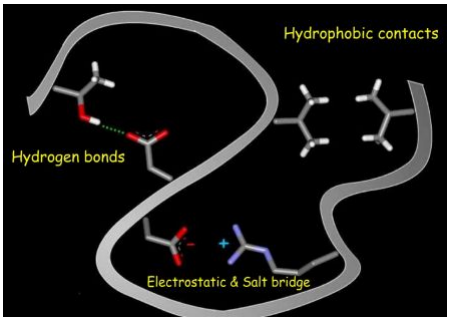
Binding mode (pose) wzajemny geometryczny układ ligandu i receptora

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

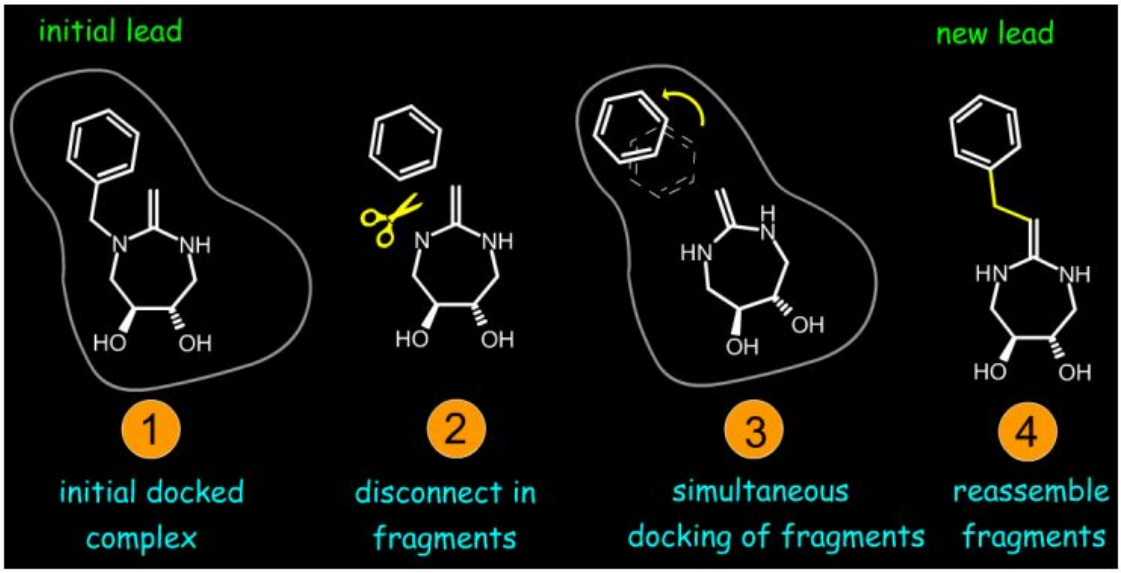


>dokowanie: ocena

OCENA EKSPERYMENTU DOKOWANIA MA CHARAKTER RELATYWNY
Zależy od zastosowanych parametrów oraz funkcji użytych przez wybrany pakiet do ceny wyniku.



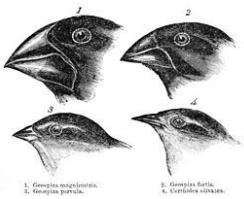

- WYNIKÓW DOKOWANIA OTRZYMANYCH Z DWÓCH RÓŻNYCH PAKIETÓW NIE MOŻNA BEZPOŚREDNIO PORÓWNAĆ!

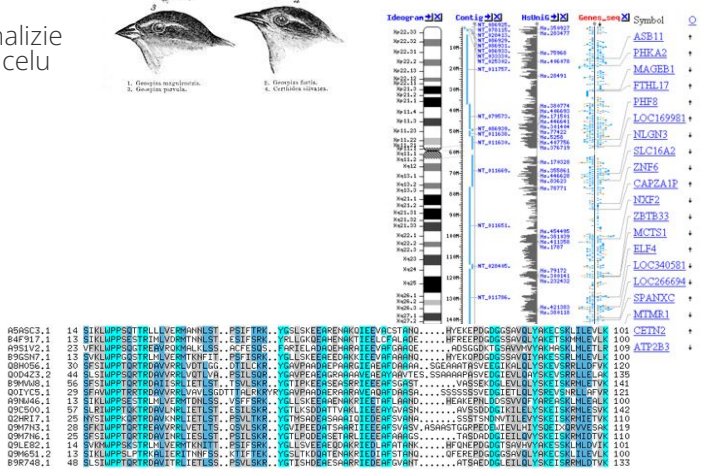



# Bioinformatyka

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.

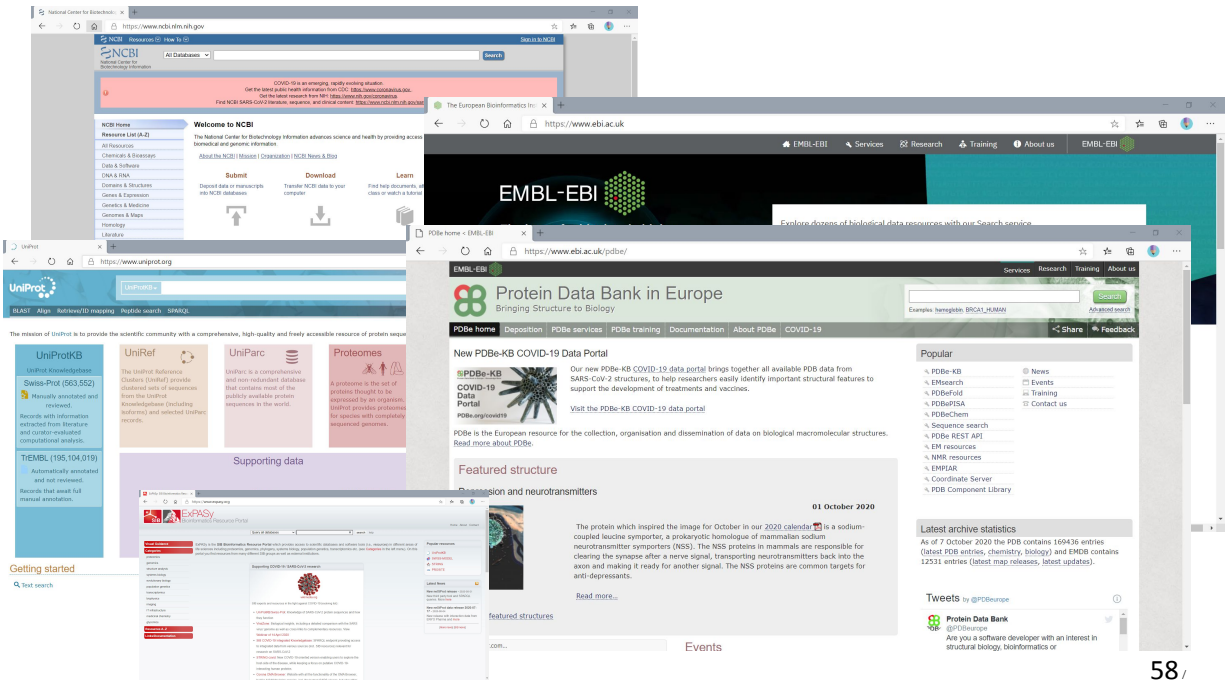
Każda naturalna wariacja w ramach gatunku stanowi nowy zbiór danych.

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

GENBANK AND WGS STATISTICS

| | | GenBank | | WGS | |
|---|---|---|---|---|---|
| Release | Date | 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 |





## 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)



## Proteom

Proteom
component białkowy komórek kodowany przez genom.

Stężenie białek w komórce...

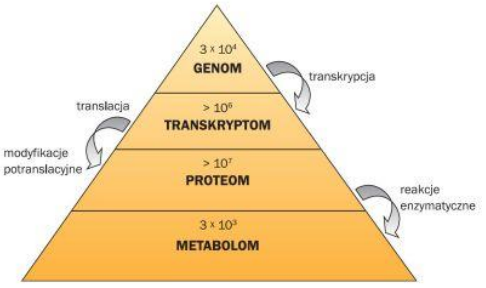
-nie jest prostą funkcją ekspresji genów

-podlega licznym modyfikacjom, które decydują o końcowych właściwościach.

W połączeniu z genomem pozwala:

-identyfikować szlaki metaboliczne zaangażowane w badane procesy

-monitorować procesy patologiczne na poziome komórkowym

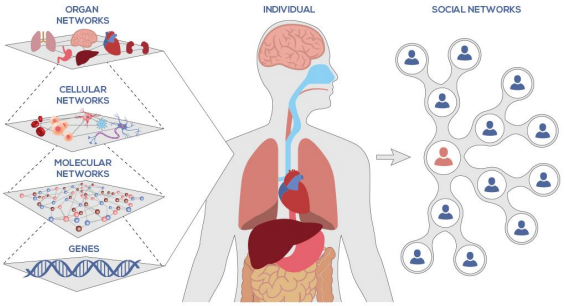


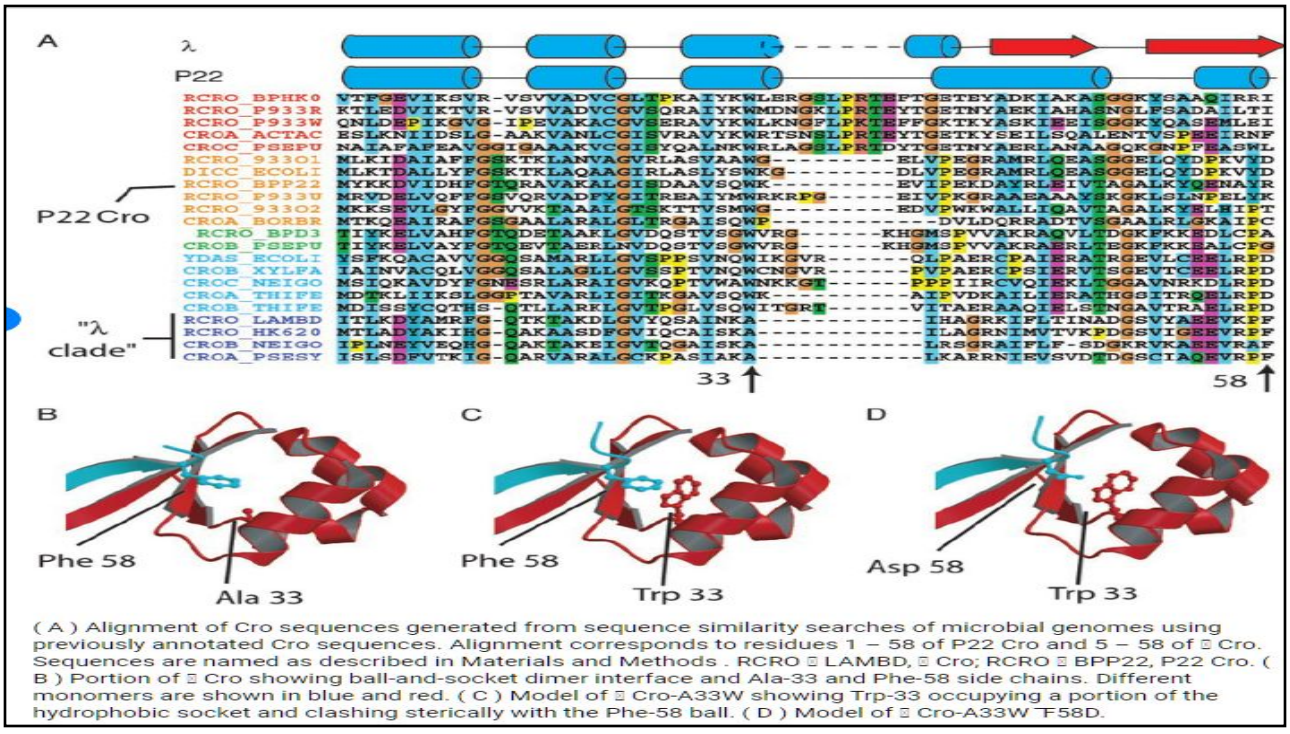

| Genom | Proteom | Metabolom |
|---|---|---|
| statyczny | dynamiczny | dynamiczny |
| możliwość amplifikacji | brak | brak |
| jednorodny | niejednorodny | niejednorodny |
| stałe stężenie | zmienne stężenie | zmienne stężenie |

# 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).



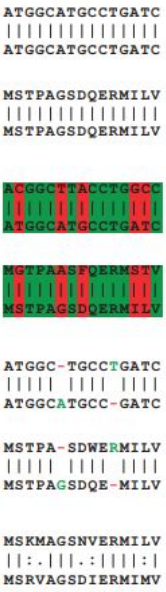




( A ) Alignment of Cro sequences generated from sequence similarity searches of microbial genomes using previously annotated Cro sequences. Alignment corresponds to residues 1 – 58 of P22 Cro and 5 – 58 of □ Cro. Sequences are named as described in Materials and Methods . RCRO □ LAMBD, □ Cro; RCRO □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

# Dopasowanie sekwencji

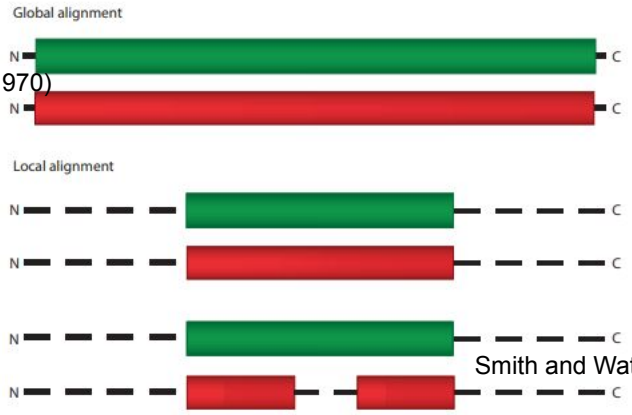
Warianty:

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




# Typy dopasowań

Needleman and Wunsch (1970)




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

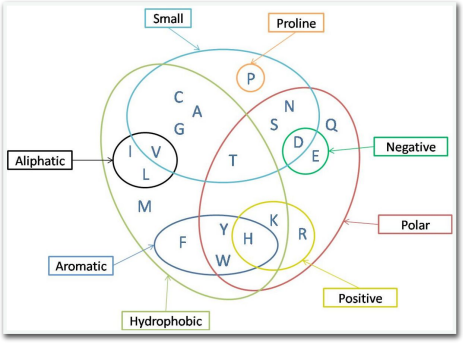
*Problem: jak ocenić podobieństwo sekwencji uwzględniając zmiany ewolucyjne?*

## 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 dopasowani 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.

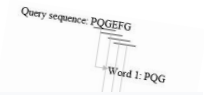
Macierze BLOSUM zostały wprowadzone przez Steven Henikoff i Jorja Henikoff w 1992 r.

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

## blast

BLAST (ang.: basic local alignmet search tool) pozwala na wyszukiwanie w parach:

- DNA:DNA (blastn)
- białko:białko (blastp)
- DNA tłumaczenie:białko (blastx)
- DNA tłumaczenie: DNA tłumaczenie (tblastx)
- białko: DNA tłumaczenie (tblastn)
- megablast
- PSI-BLAST

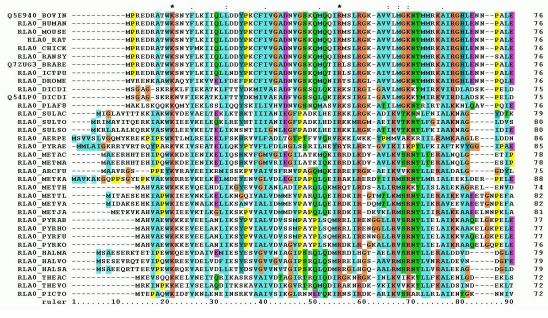



```
;LCBO - Prolactin precursor - Bovine
; a sample sequence in FASTA format
MDSKGSSQKGSRLLLLLVVSNLLLCQGVVSTPVCPNGPGNCQVSLRDLFDRAVMVSHYIHDLSS
EMFNEFDKRYAQGKGFITMALNSCHTSSLPTPEDKEQAQQTHHEVLMSLILGLLRSWNDPLYHL
VTEVRGMKGAPDAILSRAIEIEEENKRLLEGMEMIFGQVIPGAKETEPYPVWSGLPSLQTKDED
ARYSAFYNLLHCLRRDSSKIDTYLKLLNCRIIYNNNC*

>MCHU - Calmodulin - Human, rabbit, bovine, rat, and chicken
ADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTID
FPEFLTMMARKMKDTDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEKLTDEEVDEMIREA
DIDGDGQVNYEEFVQMMTAK*

>gi|5524211|gb|AAD44166.1| cytochrome b [Elephas maximus maximus]
LCLYTHIGRNIYYGSYLYSETWNTGIMLLLITMATAFMGYVLPWGQMSFWGATVITNLFSAIPYIGTNLV
EWIWGGFSVDKATLNRFFAFHFILPFTMVALAGVHLTFLHETGSNNPLGLTSDSDKIPFHPYYTIKDFLG
LLILILLLLLLALLSPDMLGDPDNHMPADPLNTPLHIKPEWYFLFAYAILRSVPNKLGGVLALFLSIVIL
GLMPFLHTSKHRSMMLRPLSQALFWTLTMDLLTLTWIGSQPVEYPYTIIGQMASILYFSIILAFLPIAGX
IENY
```

## ocena wyniku msa



Dla dopasowania dwóch sekwencji możliwe jest obliczenie parametru będącego oceną jakości przyrównania.

W przypadku dopasowania wielu sekwencji jest to nie możliwe!
Nie ma dopasowania idealnego | poprawnego.
Jest tylko dopasowanie **optymalne**.

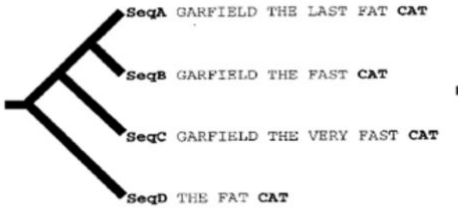
Ocena jakości dopasowania wielu sekwencji jest w gestii badacza.

*Czy kolejność wprowadzenia sekwencji w dopasowaniu będzie mieć znaczenia?*

## czy kolejność wprowadzenia sekwencji w dopasowaniu będzie mieć znaczenia?

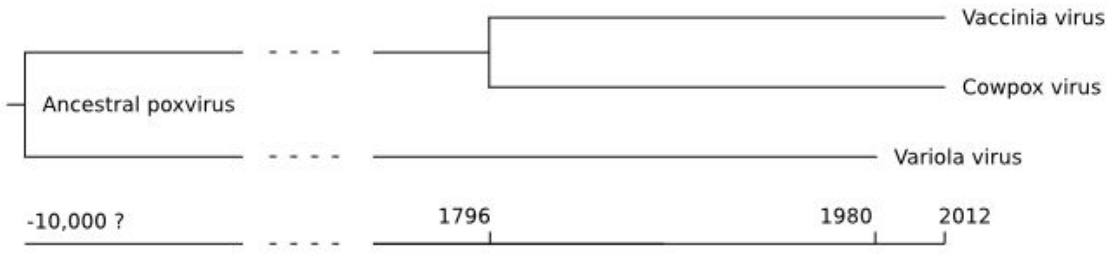
Im więcej sekwencji tym większy ‚ciężar' poprawnie dopasowanych fragmentów sekwencji.

Bez interwencji badacza automatyczna korekta dopasowanie jest niezwykle trudna.



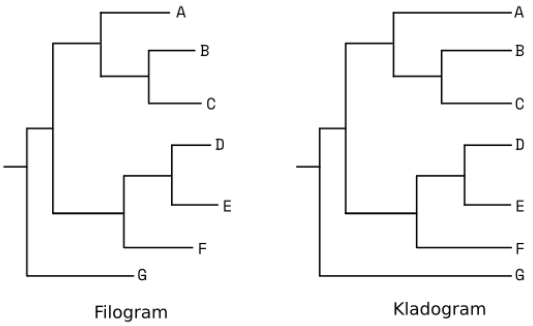

```
SeqA GARFIELD THE LAST FA-T CAT
SeqB GARFIELD THE FAST CA-T ---
SeqC GARFIELD THE VERY FAST CAT
SeqD -------  THE ---- FA-T CAT
```

## graficzna reprezentacja zmian ewolucyjnych

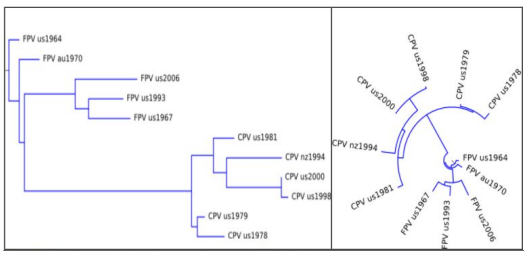


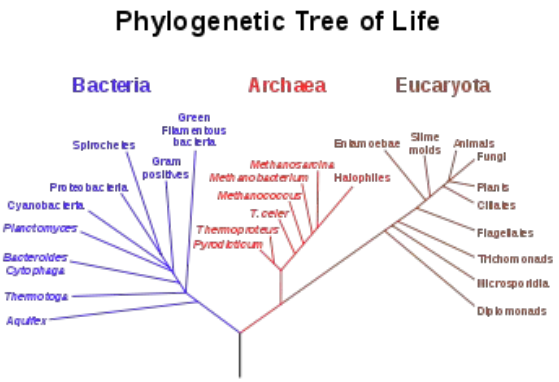
## drzewo filogenetyczne | topologia

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

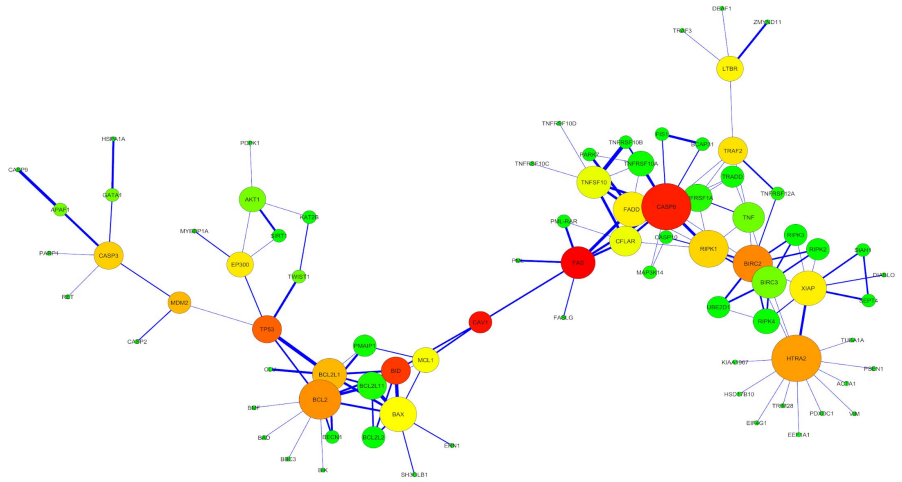



## drzewa filogenetyczne | reprezentacja

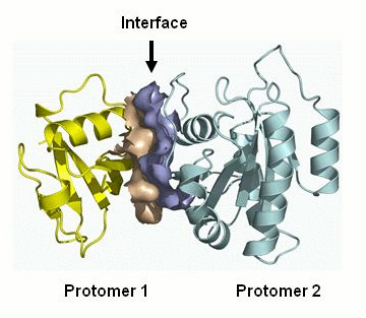


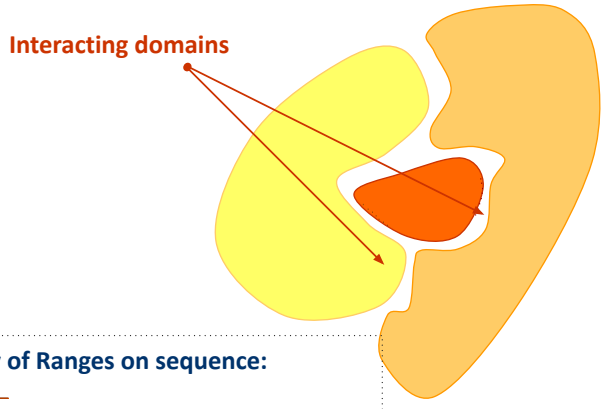



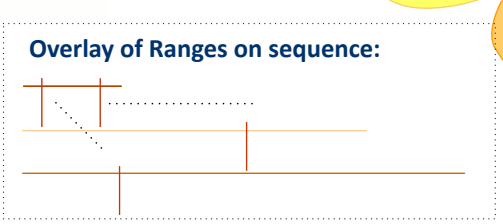
# PPI: efekt centralizacji




# Representing protein-protein interactions: interacting domains



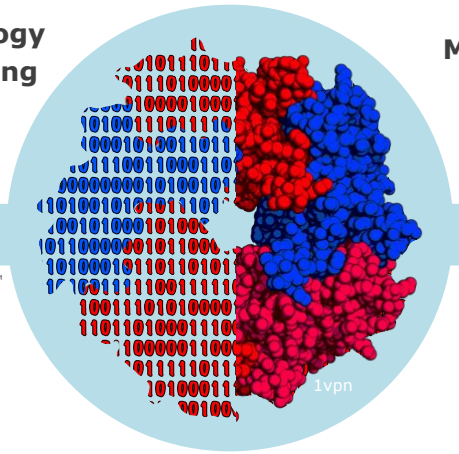
# Biologia strukturalne | interakcje

Docking

Homology Modeling

Molecular Dynamics

Threading




SAS

MS

X-Ray

FRET

EPR

Cryo-EM

NMR

Experiment



DeepMind

2021-10-04

**Protein complex prediction with AlphaFold-Multimer**

Richard Evans[1*], Michael O'Neill[1*], Alexander Pritzel[1*], Natasha Antropova[1*], Andrew Senior[1], Tim Green[1], Augustin Žídek[1], Russ Bates[1], Sam Blackwell[1], Jason Yim[1], Olaf Ronneberger[1], Sebastian Bodenstein[1], Michal Zielinski[1], Alex Bridgland[1], Anna Potapenko[1], Andrew Cowie[1], Kathryn Tunyasuvunakool[1], Rishub Jain[1], Ellen Clancy[1], Pushmeet Kohli[1], John Jumper[1*] and Demis Hassabis[1*]

[1]DeepMind, London, UK, [*]These authors contributed equally

High-throughput computation vs. High-resolution experiments

computational models are often not trusted by the experimental community