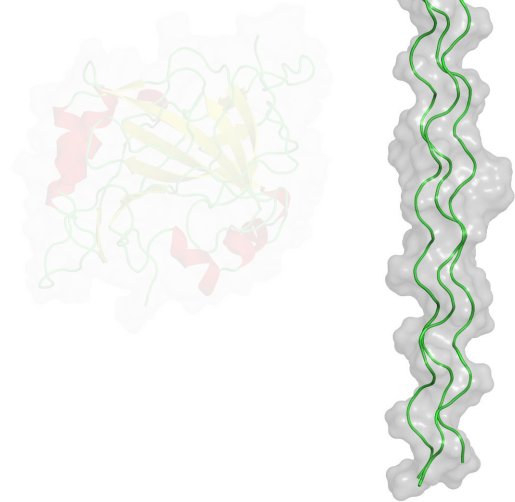


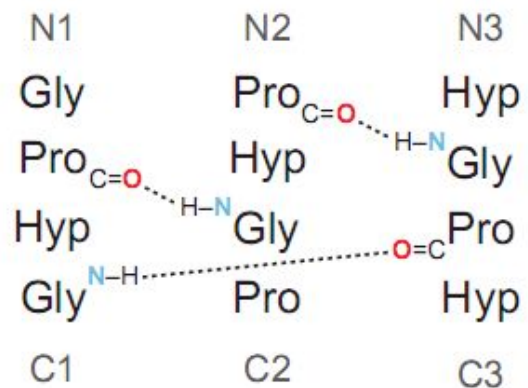
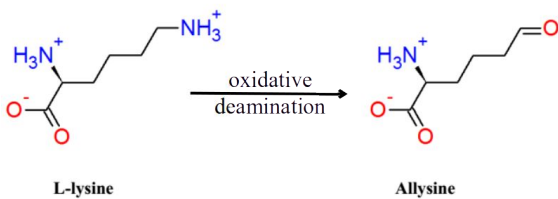
## Strukturalne rodzaje białek

- **Globularne**  
białko globularne, białko o kształcie kulistym lub sferoidalnym, zbudowane z gęsto pofalowanych lub pozwijanych łańcuchów peptydowych; wartość stosunku osiowego cząsteczki nie przekracza 3 : 1 lub 4 : 1; do b. g. należy ogromna większość białek, m.in. albuminy i globuliny oraz wszystkie enzymy.
- **Fibrylarne**  
białka proste o strukturze włóknkowej stanowiące podstawowy materiał budulcowy organizmów zwierzęcych. Są to typowe białka o budowie włóknistej, dzięki temu pełnią funkcje podporowe. Do tej grupy białek należy keratyna, kolagen, miozyna i fibroina.



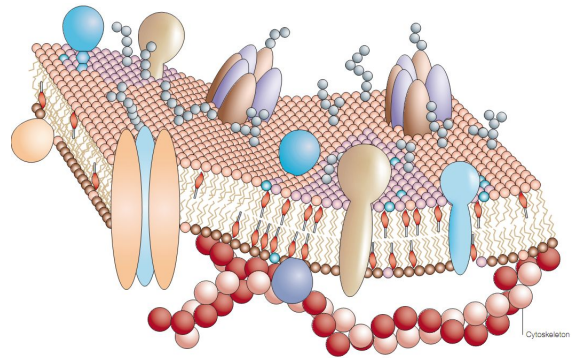
## Kolagen

- **Stabilność:**
  - wiązania wodorowe
  - cross-link związany ze starzeniem



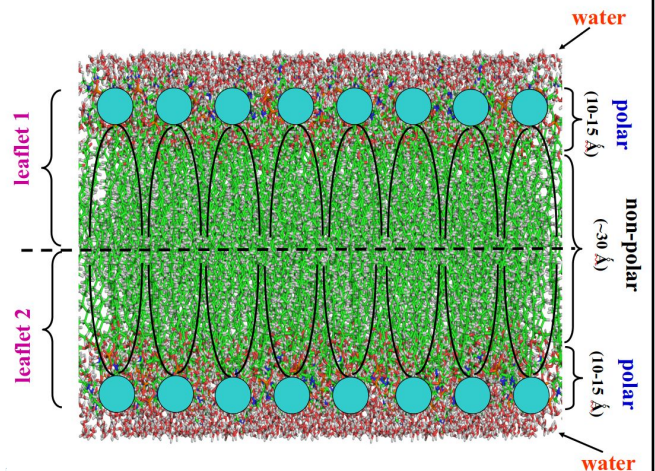
## Białka błonowe

- Białka błonowe to powszechne białka, które są częścią błon biologicznych lub wchodzą z nimi w interakcje.
- Białka błonowe dzielą się na kilka szerokich kategorii w zależności od ich lokalizacji.
  - Integralne białka błonowe są stałą częścią błony komórkowej i mogą przenikać przez błonę (transbłonowe) lub wiązać się z jedną lub drugą stroną błony (integralne monotopowe).
  - Peryferyjne białka błonowe są przejściowo związane z błoną komórkową.



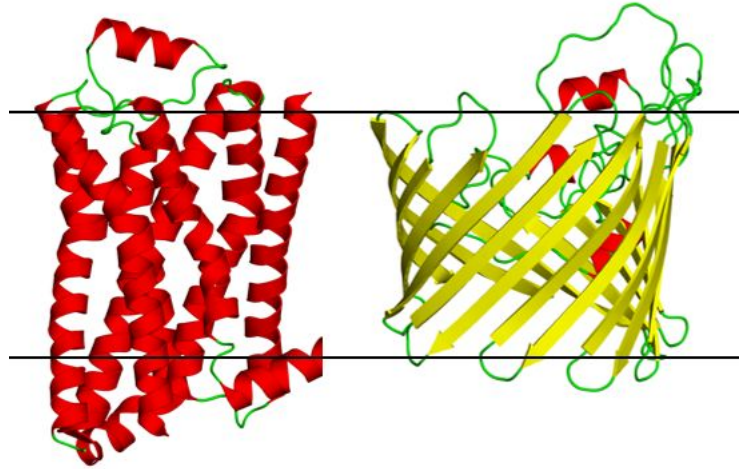
## Błona komórkowa

- Błona biologiczna, biomembrana lub błona komórkowa to selektywnie przepuszczalna błona, która oddziela wnętrze komórki od środowiska zewnętrznego lub tworzy przedziały wewnątrzkomórkowe, służąc jako granica między jedną częścią komórki a drugą.
- Błony biologiczne, w postaci błon komórek eukariotycznych, składają się z dwuwarstwy fosfolipidowej z wbudowanymi, integralnymi i peryferyjnymi białkami wykorzystywanymi w komunikacji i transporcie substancji chemicznych i jonów. Większość lipidów w błonie komórkowej zapewnia płynną matrycę dla białek, które mogą się obracać i dyfundować na boki w celu fizjologicznego funkcjonowania.
- Białka są przystosowane do wysokiej płynności środowiska dwuwarstwy lipidowej z obecnością pierścieniowej otoczki lipidowej, składającej się z cząsteczek lipidów ściśle związanych z powierzchnią integralnych białek błonowych. Błony komórkowe różnią się od izolujących tkanek utworzonych przez warstwy komórek, takich jak błony śluzowe, błony podstawne i błony surowicze.



## Białka błonowe struktura

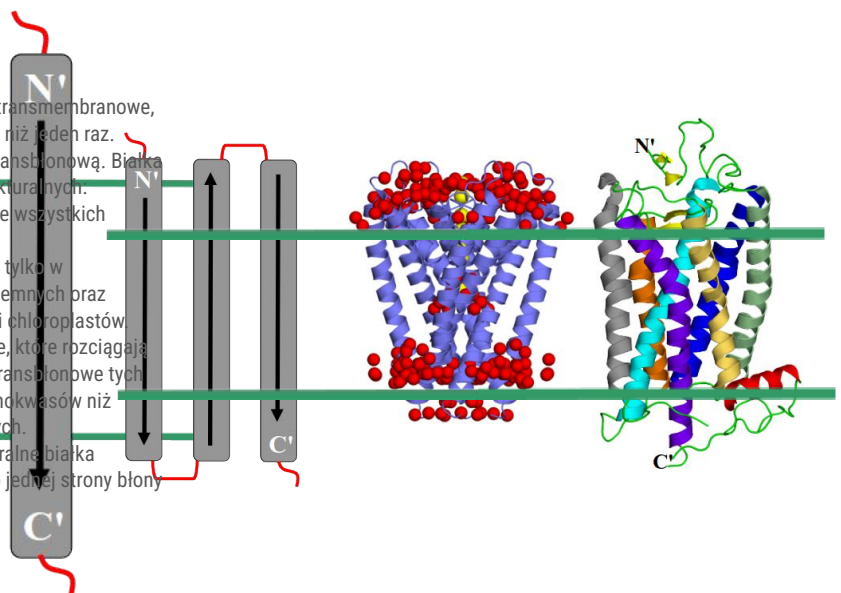
- >90 %:  $\alpha$  helikalne



## Białka błonowe struktura

### Bitopic

- >90 %:  $\alpha$  helikalne
- Integralne białka politopowe to białka transmembranowe, które rozciągają się przez błonę więcej niż jeden raz. Białka te mogą mieć różną topologię transbłonową. Białka te mają jedną z dwóch architektur strukturalnych.
- Białka wiązki helisy, które są obecne we wszystkich typach błon biologicznych;
- Białka beta-beczkowe, które występują tylko w zewnętrznych błonach bakterii Gram-ujemnych oraz zewnętrznych błonach mitochondriów i chloroplastów
- Białka bitopowe to białka transbłonowe, które rozciągają się w poprzek błony tylko raz. Helikty transbłonowe tych białek mają znacznie inny rozkład aminokwasów niż helikty transbłonowe białek politopowych.
- Integralne białka monotopowe to integralne białka błonowe, które są przyłączone tylko do jednej strony błony i nie rozciągają się na całej długości.



## Keratyna

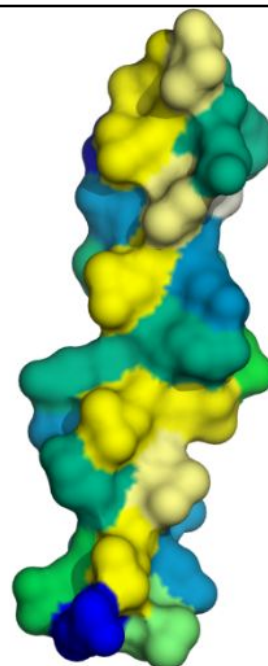
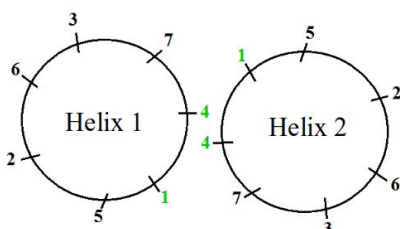
- Powtarzalny motyw: HxxHCxC

H-hydrofobowe

C-naładowane

P-polarne

a	b	c	d	e	f	g
H	P	P	H	C	P	C



L I F V A G C S T W P Y Q H K N E D R  
 Hydrophobic Hydrophilic

## Kolagen

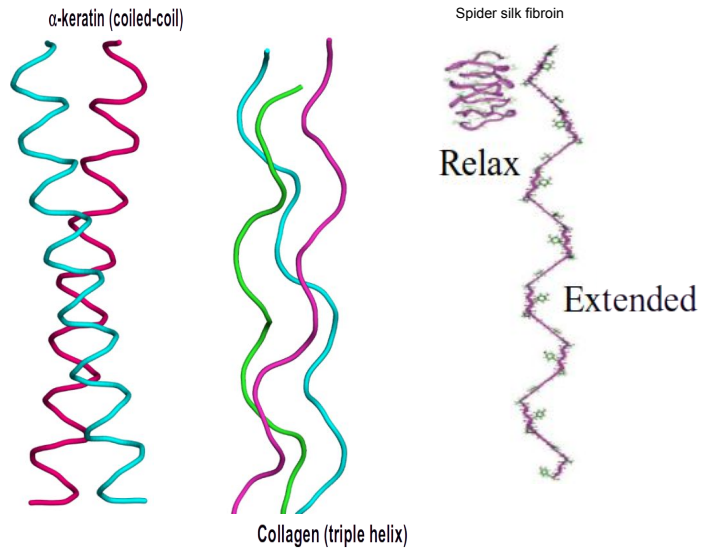
- Kolagen (/ 'kolədʒən/) jest głównym białkiem strukturalnym w macierzy zewnątrzkomórkowej występującej w różnych tkankach łącznych organizmu.
- Jako główny składnik tkanki łącznej, jest najobficiej występującym białkiem u ssaków, stanowiąc od 25% do 35% zawartości białka w całym organizmie.
- Kolagen składa się z aminokwasów połączonych ze sobą, tworząc potrójną helisę wydłużonych włókien znaną jako helisa kolagenowa. Występuje głównie w tkance łącznej, takiej jak chrząstka, kości, ścięgna, więzadła i skóra.
- Witamina C jest niezbędna do syntezy kolagenu, a witamina E zwiększa jego produkcję.

Typ	Budowa	Występowanie	Inne cechy
<b>Kolageny fibylarne (włókniste)</b>			
I	[ $\alpha 1(I)$ ] <sub>2</sub> [ $\alpha 2(I)$ ]	skóra, ścięgna, więzadła, kości, zębina, inne tkanki łączne niebędące chrząstką	włókno o dl. 300 nm
II	[ $\alpha 1(II)$ ] <sub>3</sub>	chrząstki, ciało szkliste, jądro miążdżyste (w krążku międzykręgowym)	włókno o dl. 300 nm
III	[ $\alpha 1(III)$ ] <sub>3</sub>	rozciągliwe tkanki łączne jak skóra, mięśnie, układ naczyniowy; często razem z typem I	włókno o dl. 300 nm
V	[ $\alpha 1(V)$ ] <sub>2</sub> [ $\alpha 2(V)$ ], [ $\alpha 1(V)$ ] <sub>3</sub>	rogówka, zęby, kości, łożysko, skóra, mięśnie gładkie; często razem z typem I	włókno o dl. 390 nm, N-końcowy fragment globularny
<b>Kolageny tworzące układy sieciowe</b>			
IV	[ $\alpha 1(IV)$ ] <sub>2</sub> [ $\alpha 2(IV)$ ]	wszystkie błony podstawne	
VIII	[ $\alpha 1(VIII)$ ] <sub>3</sub> , [ $\alpha 2(VIII)$ ] <sub>3</sub> , [ $\alpha 1(VIII)$ ] <sub>2</sub> [ $\alpha 2(VIII)$ ] <sup>35)</sup>	śródbłonek, błona Descemeta	tworzy heksagonalne układy sieciowe
X	[ $\alpha 1(X)$ ] <sub>3</sub>	chrząstka hipertroficzna	tworzy heksagonalne układy sieciowe
<b>Kolageny kotwiczące</b>			
VII	[ $\alpha 1(VII)$ ] <sub>3</sub>	pódesmosomy	najdłuższa domena trójhelikalna o dl. 420 nm
<b>Kolageny towarzyszące włóknom o przerywanej strukturze helisowej (FACIT)</b>			
IX	[ $\alpha 1(IX)$ ] <sub>2</sub> [ $\alpha 2(IX)$ ] <sub>3</sub> [ $\alpha 3(IX)$ ]	chrząstki, ciało szkliste; towarzyszy kolagenowi typu II	3 krótkie domeny kolagenowe zawierające 4 regiony globularne <sup>11)</sup>
XII	$\alpha 1(XII)$ <sub>3</sub> <sup>11)</sup>	towarzyszy kolagenowi typu I	
<b>Kolageny transblonowe</b>			
XIII	[ $\alpha 1(XIII)$ ] <sub>3</sub>	wiele tkanek, m.in. płytki nerwowo-mięśniowa	kolageny transblonowe mają N-końiec zatopiony wewnątrz komórki i domenę hydrofobową przechodzącą przez dwuwarstwą lipidową błony <sup>11)</sup>
XVII	[ $\alpha 1(XVII)$ ] <sub>3</sub>	nabłonek, hemidesmosomy	
<b>Multipleksyny – kolageny o licznych nieciągłych domenach o budowie helikalnej</b>			
XV	[ $\alpha 1(XV)$ ] <sub>3</sub>	występuje w proteoglikanach, gdzie stanowi rdzeń białkowy połączony z siarczanem chondroilny; towarzyszy kolagenom błony podstawnej	stabilizuje mięśnie szkieletowe i mikronaczynia
XVIII	[ $\alpha 1(XVIII)$ ] <sub>3</sub>	towarzyszy kolagenom błony podstawnej	utrzymuje strukturalną integralność błony podstawnej; bliski homolog kolagenu typu XV



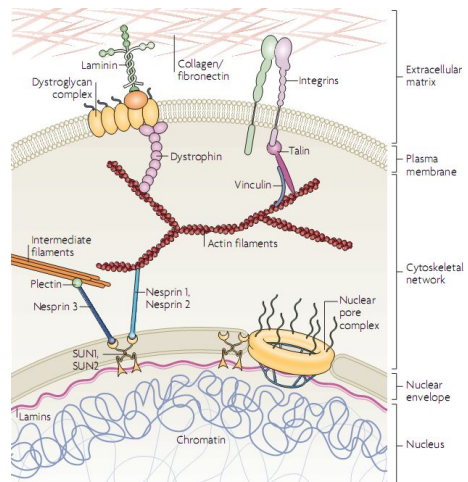
## Właściwości ogólne

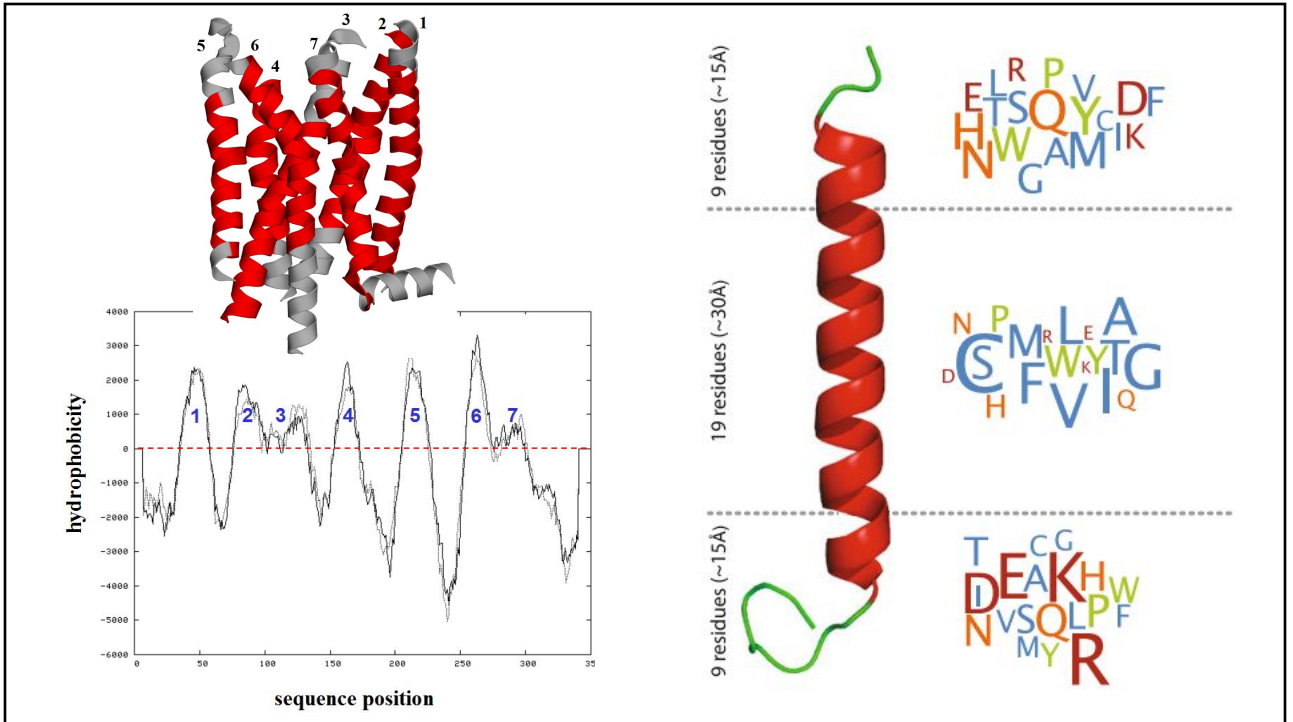
- Superstruktura drugorzędowa
  - nadrzędne elementy  $\alpha$  lub  $\beta$
- Hydrofobowość
- Zdolność tworzenia makrostruktur czwartorzędowych



## Wykorzystanie białek fibrylarnych

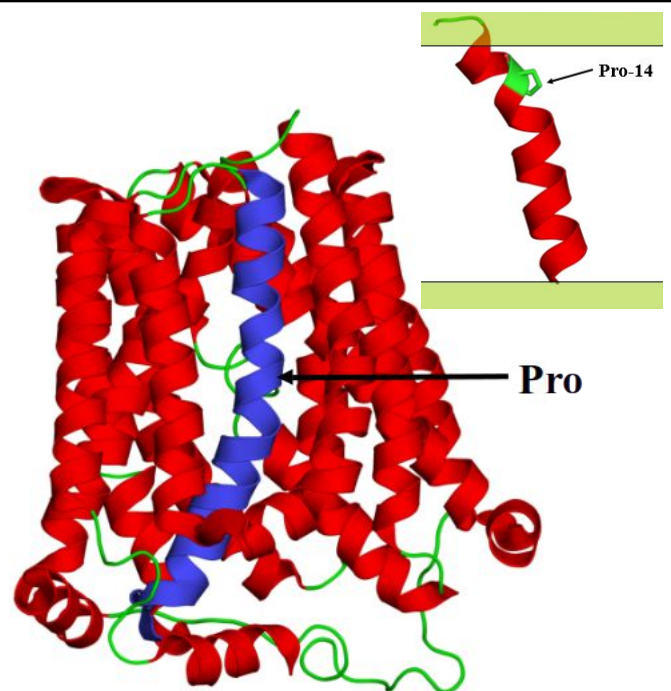
- Wsparcie mechaniczne i ochrona
  - cytoszkielet, np. keratyna
  - macierz pozakomórkowa, np. kolagen
- Budowa struktur komórkowych
- Komunikacja wewnątrzkomórkowa
- 'Ruch'
- Budowa struktur zewnętrznych, np. jedwab
- Funkcje regulacyjne
  - pigmentacja
  - wzrost włosów
  - regeneracja nabłonka





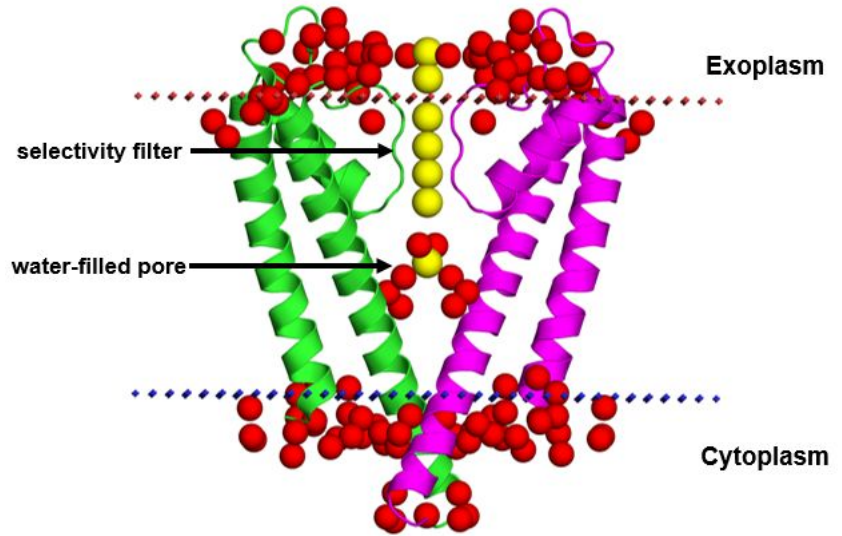
## Dynamika

- Zmiana konformacji związana z funkcją:
  - kanał/transporter musi wykazywać selektywność
  - receptor musi posiadać formę aktywną i nieaktywną



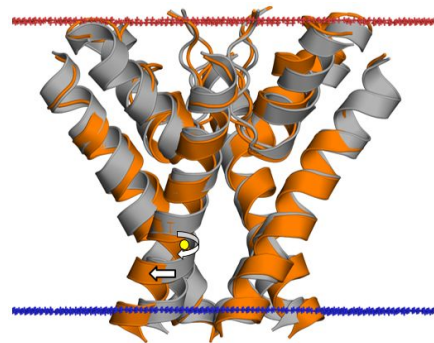
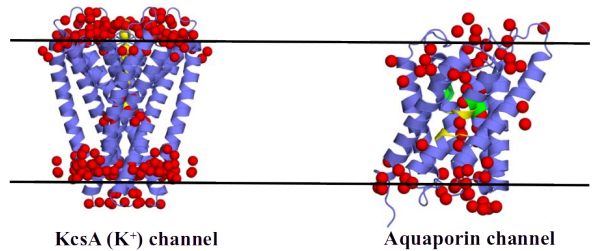
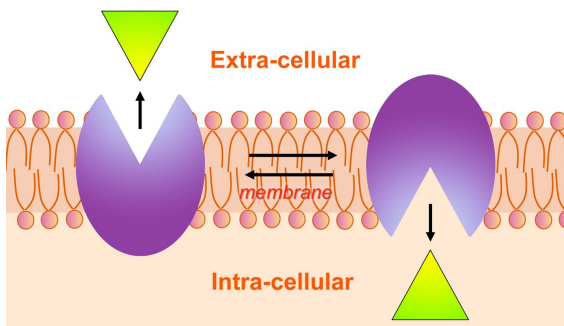
## Białka transportowe

- Selektywność



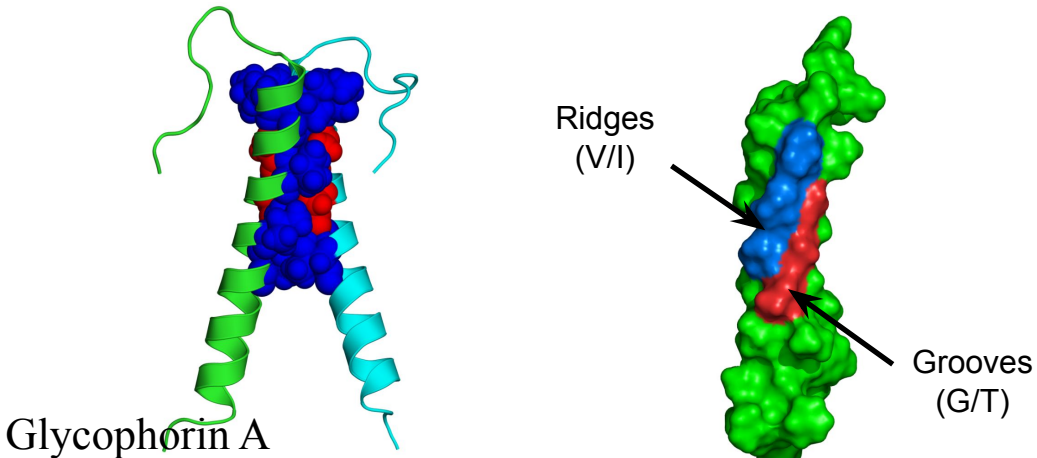
## Białka transportowe

- Reaktywność



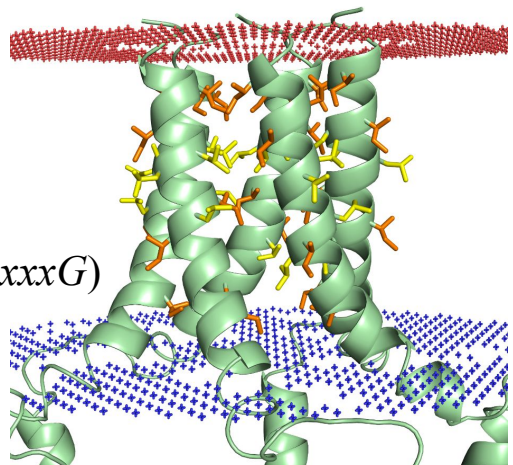
Efektywność upakowania w błonie

- **Grzebienie (V/L/I) & doliny (G/S/T/A)** [motyw **GxxG**]
- **Nachylenie (~20°)**



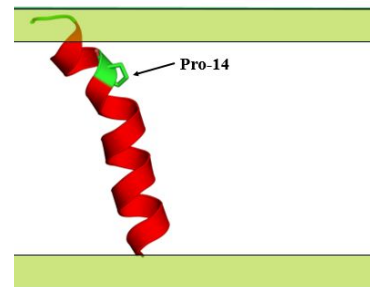
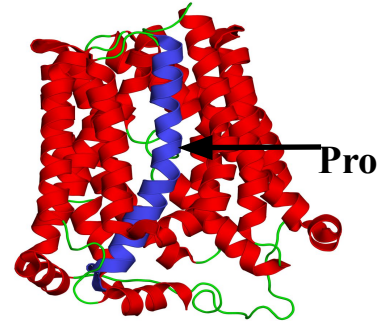
Efektywność upakowania w błonie

- **Inne motywy stabilizujące**
  - Suwaki leucynowe/izoleucynowe
  - Suwaki serynowe
  - Suwaki glicynowe (motyw *GxxxGxxxG*)



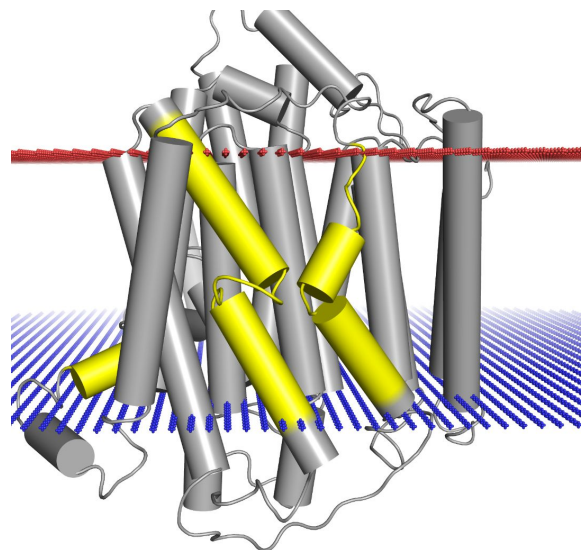
## Elementary dynamiczne

- **Gly, Pro** znajdujące się w załamaniach
- **Pro** w elementach transbłonowych (TM) służą:
  - rodzaj zawiasów
  - odpowiadają za aktywację funkcji transportowych
  - zmianę konformacji receptorów



## Elementary dynamiczne

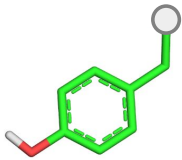
- **Nieciągłość helis (około 40%):**
  - helisy  $3_{10}$ ,
  - $\pi$ -helices,
  - przerwy.
- **Konsekwencje dudowy mechanicznej:**
  - optymalizacja odległości
  - tworzenie miejsc aktywnych



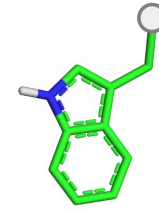
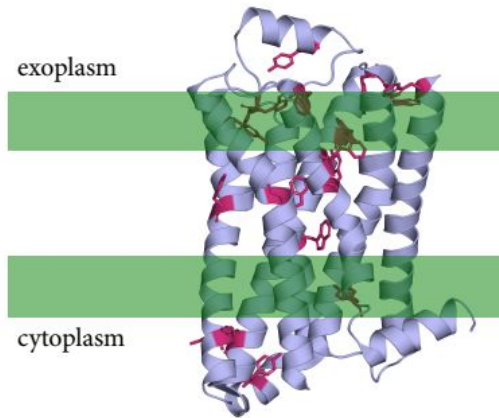


Skład jakościowy

- **Pas aromatyyczny - zakotwiczenie**



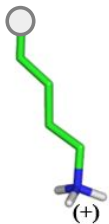
**Tyrosine**  
(Tyr, Y)



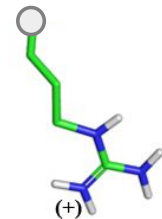
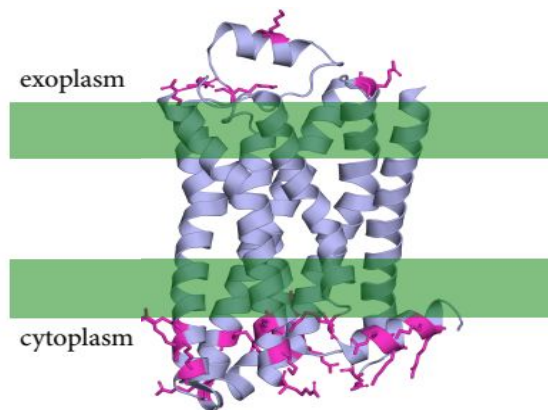
**Tryptophan**  
(Trp, W)

Skład jakościowy

- **Arg, Lys – oddziaływania zewnętrzne**



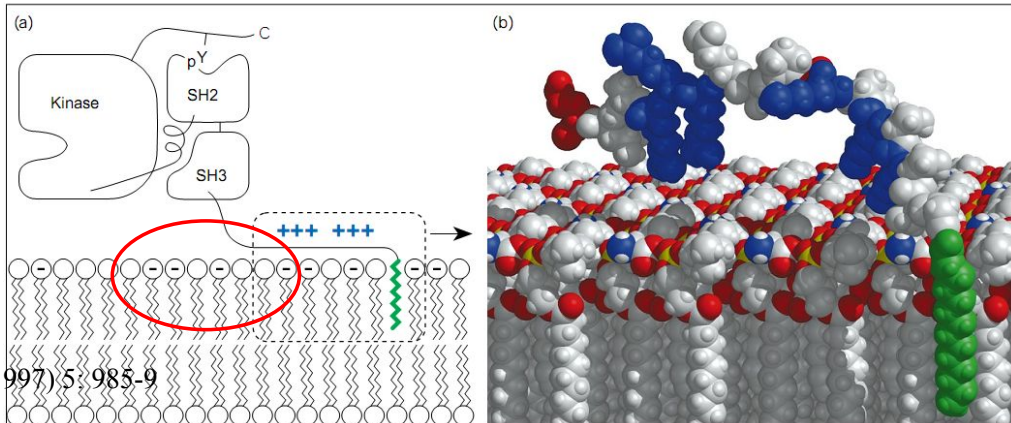
**Lysine**  
(Lys, K)



**Arginine**  
(Arg, R)

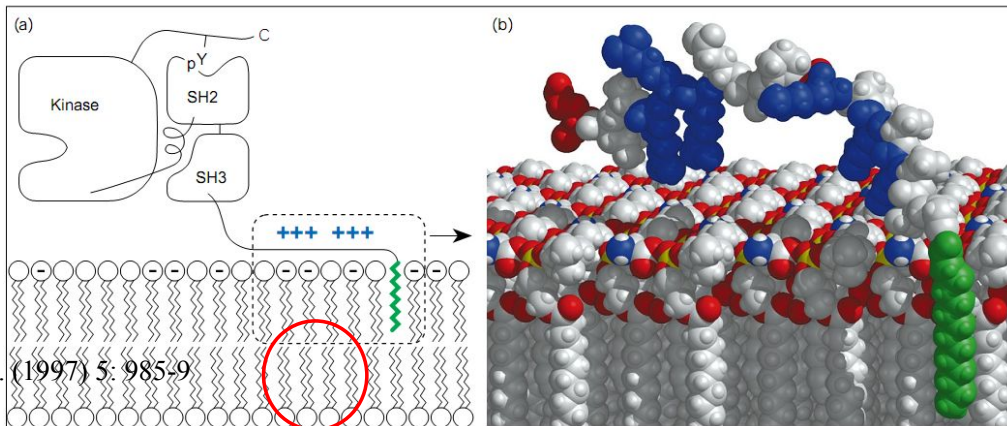
Białka  
związane z błonami

- oddziaływania elektrostatyczne



Białka  
związane z błonami

- oddziaływania kowalencyjne



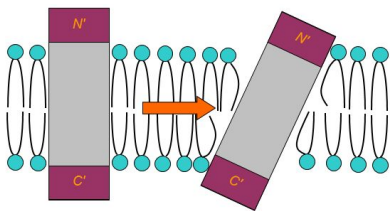


# Interakcje białko-błona

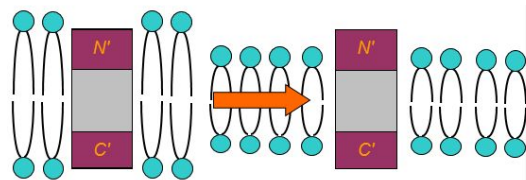
## Wpływ błony na białko

- **Niedopasowanie hydrofobowe**

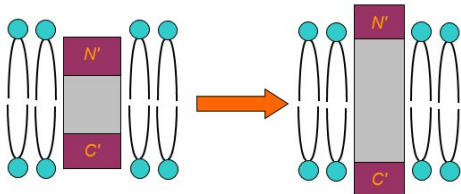
Skrzywienie



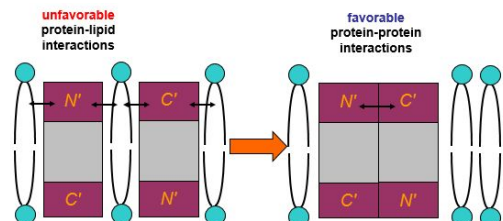
Dyfuzja pozioma



Zmiana konformacji

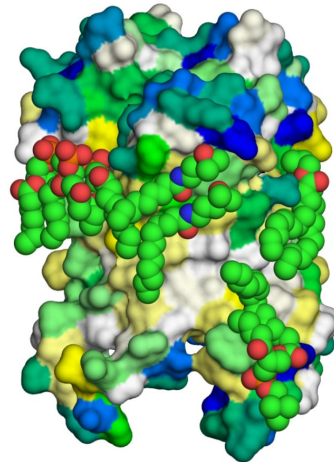


Oligomeryzacja



## Wpływ błony na białko

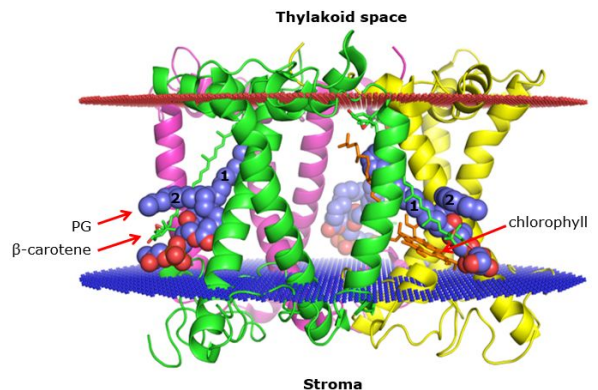
- **Pojedyncze komponenty błony mogą oddziaływać silnie z białkiem (np. artefakty krystalizacyjne).**



L I F V A G C S T W P Y Q H K N E D R  
Hydrophobic Hydrophilic

## Wpływ kw. tłuszczowych na białko

- **Interakcje pomiędzy kwasami tłuszczowymi a białkami błonowymi są niezbędne dla utrzymania:**
  - **stabilności**
  - **poprawności fałdowania**
  - **aktywności biologicznej**

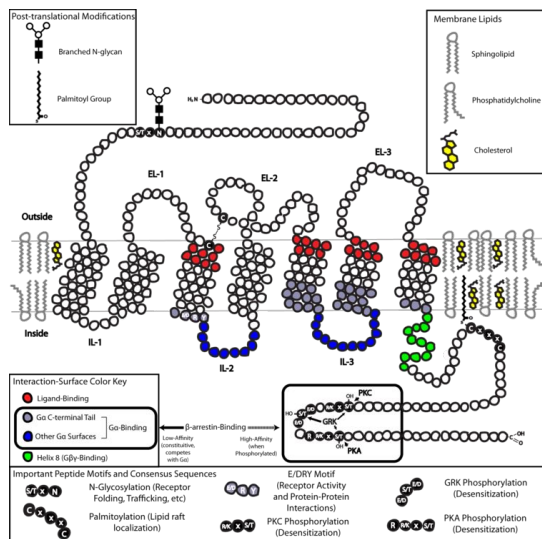




## Receptory sprzężone z białkami G

- GPCRs (ang. G protein coupled receptors) stanowią najliczniejszą i bardzo zróżnicowaną grupę białek błonowych odpowiedzialnych za przekazywanie sygnałów przez dwuwarstwą lipidową do miejsc efektorowych znajdujących się wewnątrz komórki.
- Sekwencjonowanie ludzkiego genomu ujawniło występowanie ok. 800 różnych typów receptorów należących do rodziny GPCR (geny kodujące receptory GPCR stanowią powyżej 3% ludzkiego genomu), a ponad połowa z nich wykazuje potencjalne znaczenie dla przemysłu farmaceutycznego.
- Analiza porównawcza sekwencji receptorów GPCR oraz badania funkcji poszczególnych typów receptorów doprowadziły do podziału tej rodziny białek na klasy, które oznaczono literami od A do F.
- Pierwsza, najliczniejsza klasa A (nazywana również klasą podobnych do rodopsyny – ang. rhodopsin like), obejmująca ponad 80% wszystkich GPCR, to klasa receptorów rodopsyno-podobnych. Receptory wchodzące w skład tej grupy są jednymi z najlepiej zbadanych. Zaliczamy do nich, obok rodopsyny, między innymi: receptory adrenergiczne, opioidowe, adenozynowe, kanabinoidowe, receptory chemokin, dopaminowe i histaminowe.
- Klasę B stanowią receptory sekretyno-podobne. Do klasy C zaliczamy receptory glutaminergiczne i feromonowe.
- Kolejne klasy D i E tworzą odpowiednio receptory feromonów grzybów oraz receptory cAMP.
- Ostatnia klasa F to receptory frizzled/smoothened. Klasyfikacja ta pokrywa się w większości z nową klasyfikacją GRAFS, opartą na badaniach filogenetycznych. Nazwa GRAFS pochodzi od pierwszych liter wyodrębnionych rodzin receptorów, do których należą odpowiednio: receptory Glutaminergiczne, Rodopsyno-podobne, Adhezyjne, Frizzled i smakowe oraz Sekretyno-podobne.

## GPCRs



## The Nobel Prize in Chemistry 2012



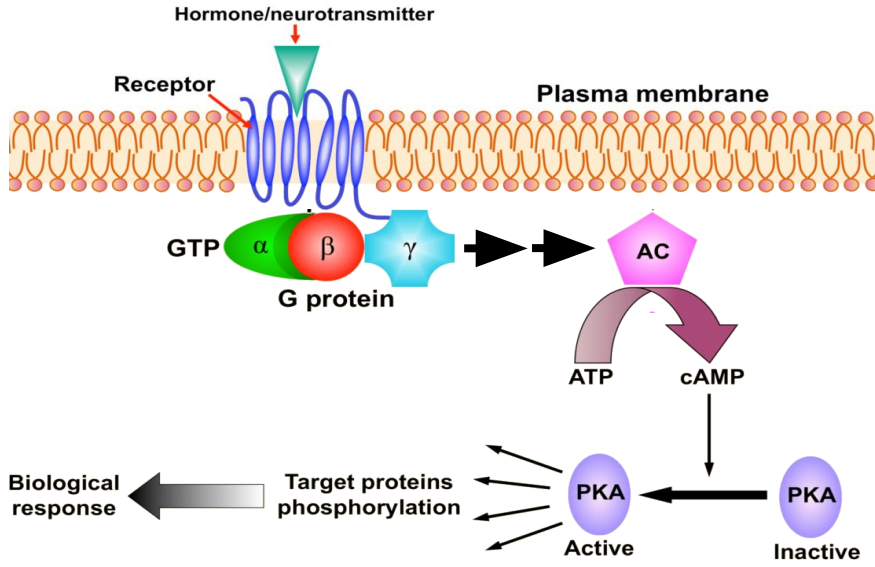
© The Nobel Foundation, Photo: U. Montan  
**Robert J. Lefkowitz**  
 Prize share: 1/2



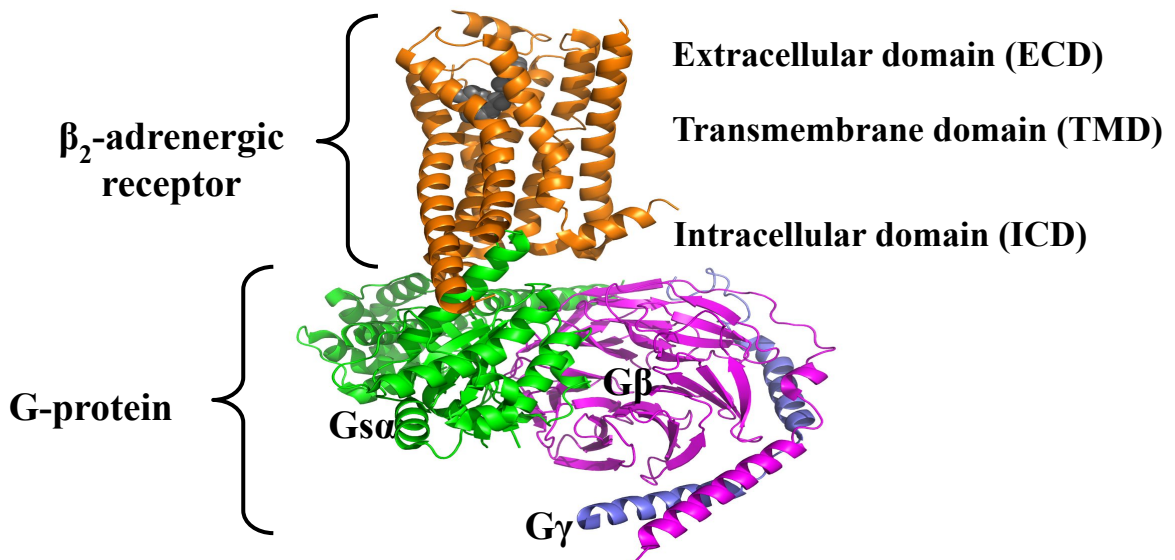
© The Nobel Foundation, Photo: U. Montan  
**Brian K. Kobilka**  
 Prize share: 1/2

The Nobel Prize in Chemistry 2012 was awarded jointly to Robert J. Lefkowitz and Brian K. Kobilka "for studies of G-protein-coupled receptors"

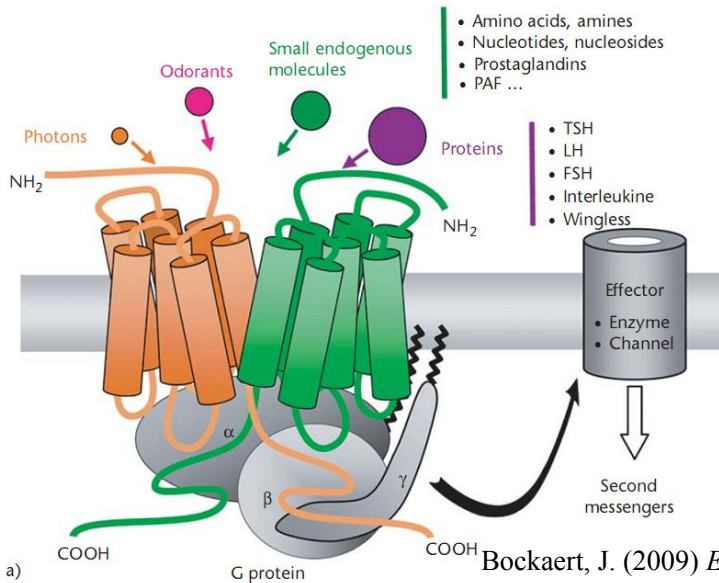
## GPCRs ścieżka sygnalizacyjna



## The GPCR-G protein complex



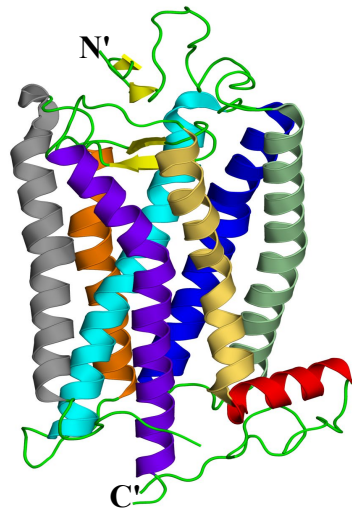
## GPCRs ligandy i efektery



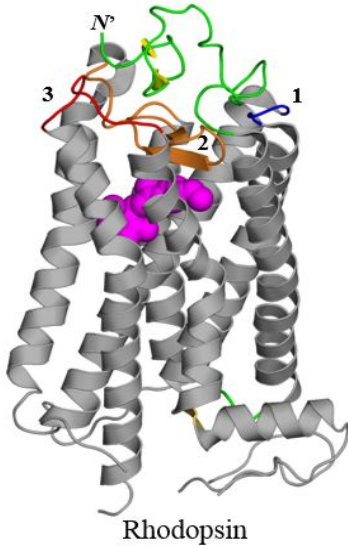
Bockaert, J. (2009) *ELS*. John Wiley & Sons, Ltd

## GPCRs podobieństwo strukturalne

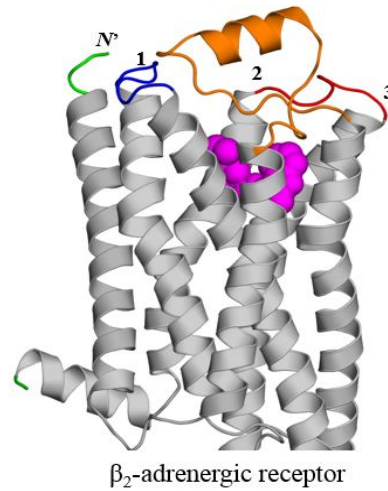
- **GPCRs mają podobne struktury pomimo niskiej zgodności sekwencji**
  - Rhodopsin &  $\beta_2$ -AR: 21% zgodności,
  - TM r.m.s.d = 1.6 Å
- **TMD jest najlepiej zachowaną częścią z 7 TM helisami**



## GPCRs różnice strukturalne



(ECL2 is long and blocks entrance)



(ECL2 is a short helix, entrance is open)

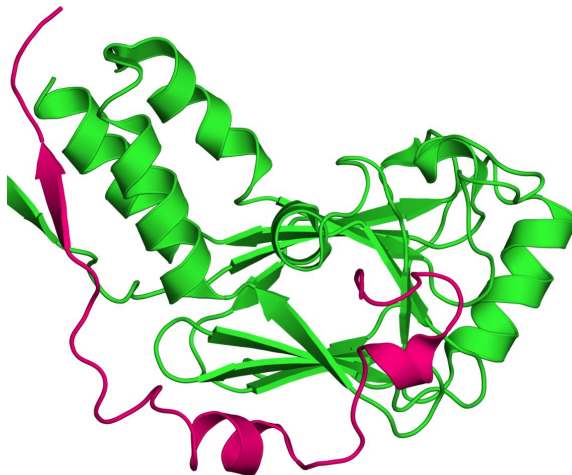
## GPCRs

- **Specyficzne właściwości**
- ECD: rozróżnia ligand
- TMD:
  - specyficzna aktywacja w oparciu o motywy *NPxxY* & *CWxP*
  - kontrola allosteryczna (cholesterol)
  - dimeryzacja
- ICD: stabilizacja stanu nieaktywnego poprzez motyw *D/ERY*

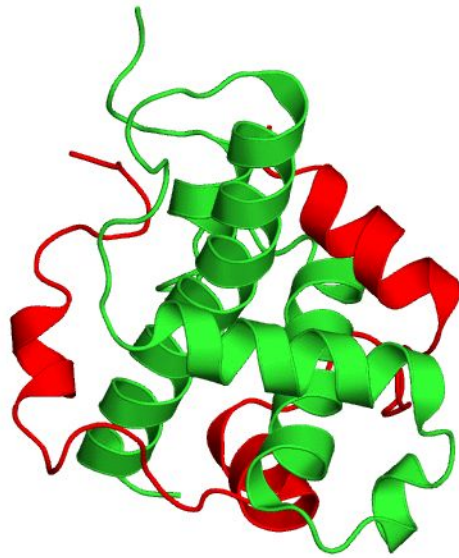
# Natywnie nie pofałdowane białka (*intrinsically unstructured proteins, IUPs*)

Brak organizacji strukturalnej w białkach

- **Regionalizacja (IDRs)**

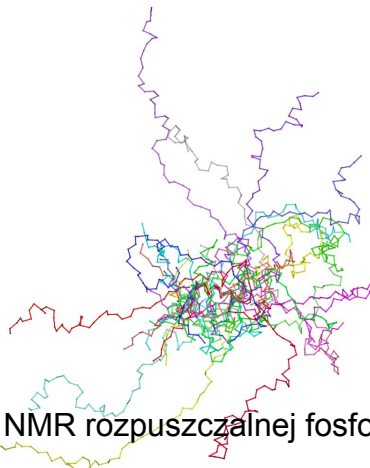






Brak organizacji strukturalnej w białkach

- **Całe białka (IUP)**



20 struktur NMR rozpuszczalnej fosfoproteiny



Rola IUPs w  
cyklu życiowym komórki

- **Regulacja cyklu komórkowego i ekspresji genów**
- **Pakowanie DNA**
- **Transport komórkowy**
- **Komunikacja i sygnalizacja**
- **Regulacja**
- **Patogeneza (np. nowotwory)**

## Metodyka badań IUPs

- **Techniki eksperymentalne**

- ~~Krystalografia~~

- Spektroskopia: NMR/EPR, CD, FRET

- High-speed atomic force microscopy (mikroskopia sił atomowych)

- Raman optical activity

- Time-resolved small angle X-ray scattering (TR-SAXS)

- Metody obliczeniowe:

- IUPred, DISOPRED (prediction)

- Molecular dynamics (characterization)

## Wybrane funkcje

- **Regulacja funkcji i procesów biologicznych:**

- Komunikacja

- Sygnalizacja

- Cykle komórkowe

- Ekspresja genów

- **Specyficzne tryby aktywności:**

- rozpoznawanie cząsteczek

- kontrola aktywności i entropii systemu

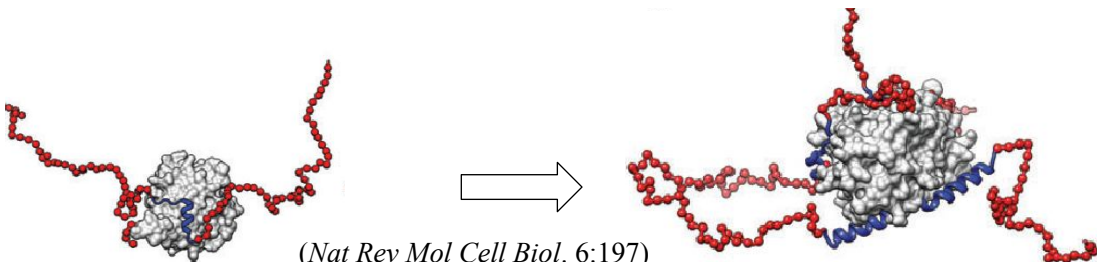
## Funkcje

### • Rozpoznawanie molekularne

- **Assemblers:** act as scaffolds that spatially direct the assembly of large protein complexes (ribosome, cytoskeleton, chromatin)
- **Effectors:** regulate the activity of other proteins or other regions of the same protein (e.g. p21 and p27 in cell cycle regulation)
- **Scavengers:** store and/or neutralize small ligands (e.g. casein)
- **Display sites (IDRs):** targets for post-translational modifications
- **Chaperones:** assist the folding of proteins and RNA molecules

## Rozpoznawanie molekularne

- Wiązanie z partnerem powoduje częściową organizację IUPs
- Pewne fragmenty pozostają elastyczne prawdopodobnie
  - powodując zmiany strukturalne u partnera
  - zmieniając właściwości w miejscu wiązania



## Funkcje

- **Kontrola aktywności i entropii systemu**

- IUPs stanowią siłę przeciwną do uporządkowanych zmian strukturalnych lub mogą wpływać na organizację i orientację domen wewnątrz kompleksów białkowych
- Nawet w trakcie aktywności pozostają nie pofałdowane
- Rodzaje:
  - **Sprężyny**: generują pasywną siłę (powrót napiętych mięśni do stanu spoczynkowego)
  - **Separator**: utrzymuje niezbędne odległości
  - **Łącznik**: łączy domeny w nowe konfiguracje

## Charakterystyka IUPs

- **Struktura szczątkowa**

- Helisy **PPII** (polyprolinne II helices) (10-20% aminokwasów)
- Motywy odpowiedzialne za wiązanie:
  - (1) **MoREs/MoRFs** (ang. molecular recognition elements/featurea, 10-70 AA, często amifatyczne),
  - (2) **SLiMs** (ang. short linear motif).

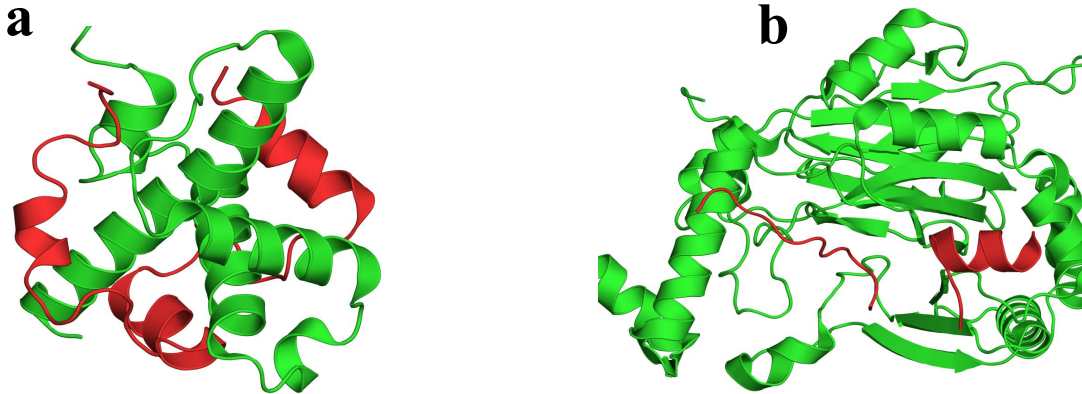
- **Preferencje sekwencji**

- **Pro** → formuje i stabilizuje strukturę drugorzędową
- **Polarne** → blokuje fałdowanie
- **Cys** → brak wiązań disarczkowych
- **SLiMs**: niskie powinowactwo, szybka wymiana
- **Brak konserwacji ewolucyjnej**



## Zależność struktura-funkcja IUPs

- **Elastyczność konformacyjna prowadzi do elastyczności wiązania w zależności od warunków.**



The hypoxia-inducible factor-1α (red) bound to different targets

## Zależność struktura-funkcja IUPs

- **Powinowactwo wiązania:**

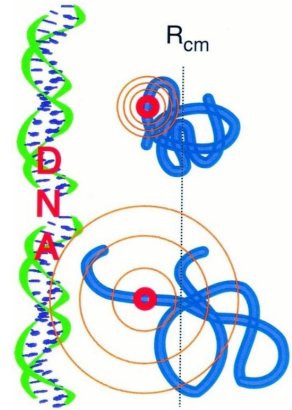
$$\Delta G_{bind} = \Delta H_{bind} - T\Delta S_{bind}$$

- IUPs → **interakcje wiązanie-faldowanie** → wzrost  $\Delta S_{bind}$  w porównaniu do białek pofaldowanych → **słabsze wiązanie** →
  1. **większa częstotliwość wymiany partnerów** (zaleta dla białek sygnalizacyjnych)
  2. komponent  $\Delta H_{bind}$  wyłącza katalizę → **enzymy są ZAWSZE sfaldowane**
- Niepolarny interfejs → większe  $\Delta H_{bind}$  kompensuje utratę entropii

## Zależność struktura-funkcja IUPs

- **Kinetyka mechanizmu wiązania-fałdowanie**

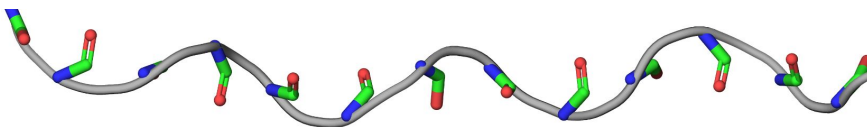
1. Wstępne słabe wiązanie poprzez szcążkowe fałdowanie
2. Wysokie powinowactwo po identyfikacji partnera i osiągnięciu optymalnej konfiguracji.



## Zależność struktura-funkcja IUPs

- **Specyficzność substratowa (...lub jej brak)**

- brak specyficznych wiązań wodorowych w strukturze drugorzędowej**
  - **helisy PPII nie posiadają wiązań wodorowych**
- specyficzne wiązanie wykorzystuje dostępne wolne wiązania**



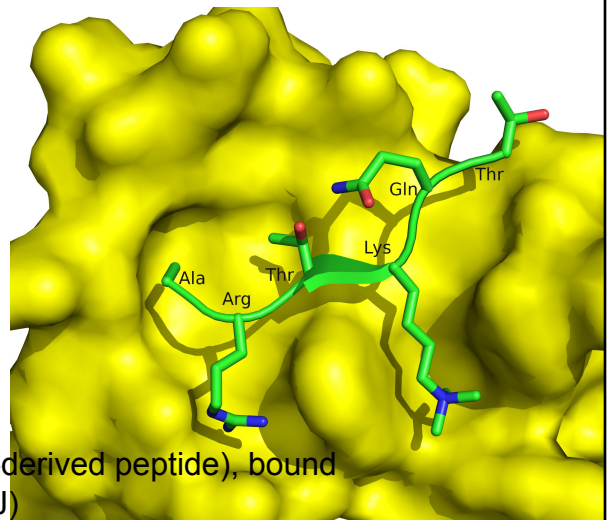
PPII helix with unpaired backbone polar groups

## Zależność struktura-funkcja IUPs

- **Reulacja poprzez modyfikacje potranslacyjne**
  - Brak struktury
  - Rozwinięty łańcuch
  - Łatwość enzymatycznej fosforylacji, glikozylacji, acylacji, hydroxylacji...
  - SLiMs są łatwo rozpoznawane przez enzymy PTM

## Zależność struktura-funkcja IUPs

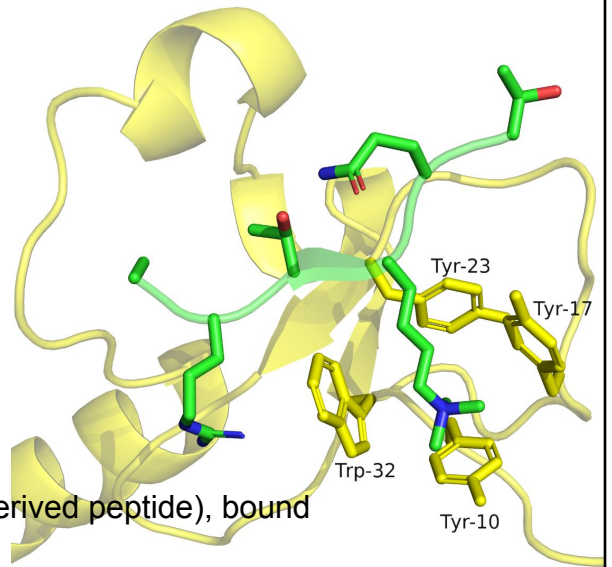
- **Regulacja poprzez modyfikacje potranslacyjne**



Methylated Lys in short linear motif (histone-derived peptide), bound to human BPTF (2F6J)

## Zależność struktura-funkcja IUPs

- **Regulacja poprzez modyfikacje potranslacyjne**



Methylated Lys in short linear motif (histone-derived peptide), bound to human BPTF (2F6J)

## Metody przewidywania struktury białek

## **Eksperymentalne metody wyznaczania struktury 3D**

- **Metody dyfrakcyjne:**

- X-ray diffraction/scattering
- Neutron scattering
- Electron microscopy/crystallography

- **Metody spektroskopowe**

- Nuclear magnetic resonance (NMR) spectroscopy
- Electron paramagnetic resonance (EPR) spectroscopy
- FTIR, Raman spectroscopy, circular dichroism, mass spectrometry

## **Przewidywanie struktury**

- **Metody eksperymentalne są dokładne ale...**

- są czasochłonne
- są kosztowne
- białka błonowe i kompleksy są kłopotliwe

- Tylko niewielki odsetek białek posiada eksperymentalne struktury.

- **Metody obliczeniowe są mniej dokładne ale szybsze!**

# Przegląd stosowanych rozwiązań

- **Podejście fizyczne (*ab initio*)**

1. Wykorzystanie pól siłowych do obliczenia energii potencjalnej
2. Próbkowanie konfiguracji
3. Wykorzystanie teorii pól średnich

- **Podejście oparte o szablony (*comparative*)**

1. Modelowanie homologiczne
2. Rozpoznawanie foldu (struktury drugorzędowej)

- **Metody zintegrowane:**

1. Kierowane eksperymentalnie modelowanie komputerowe
2. Metody ewolucyjne (skorelowane mutacje)

## The Nobel Prize in Chemistry 2013



© Nobel Media AB. Photo: A. Mahmoud  
Martin Karplus  
Prize share: 1/3

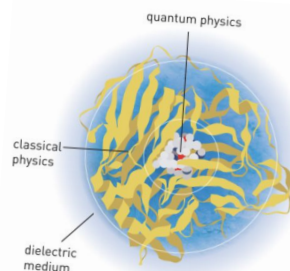


© Nobel Media AB. Photo: A. Mahmoud  
Michael Levitt  
Prize share: 1/3



© Nobel Media AB. Photo: A. Mahmoud  
Arieh Warshel  
Prize share: 1/3

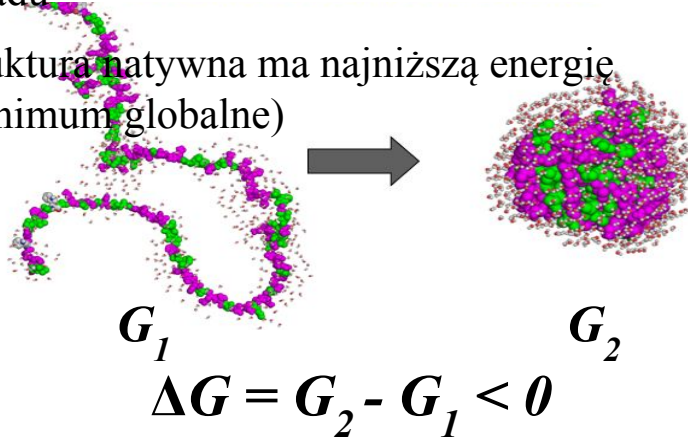
The Nobel Prize in Chemistry 2013 was awarded jointly to Martin Karplus, Michael Levitt and Arieh Warshel "for the development of multiscale models for complex chemical systems"



## Podejście fizyczne

- **Założenia**

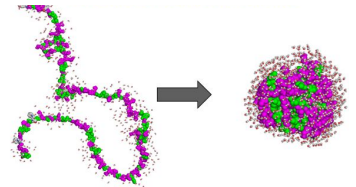
1. Fałdowanie białek bierze się z obniżeniem energii układu
2. Struktura natywna ma najniższą energię (minimum globalne)



## Podejście fizyczne

3. Proces fałdowania można opisać poprzez:

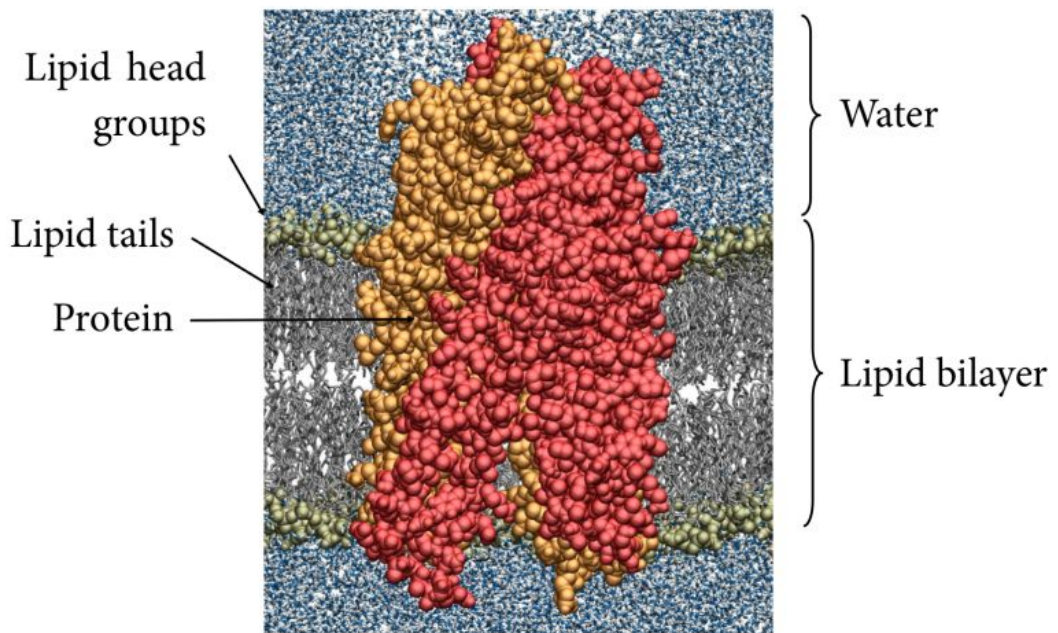
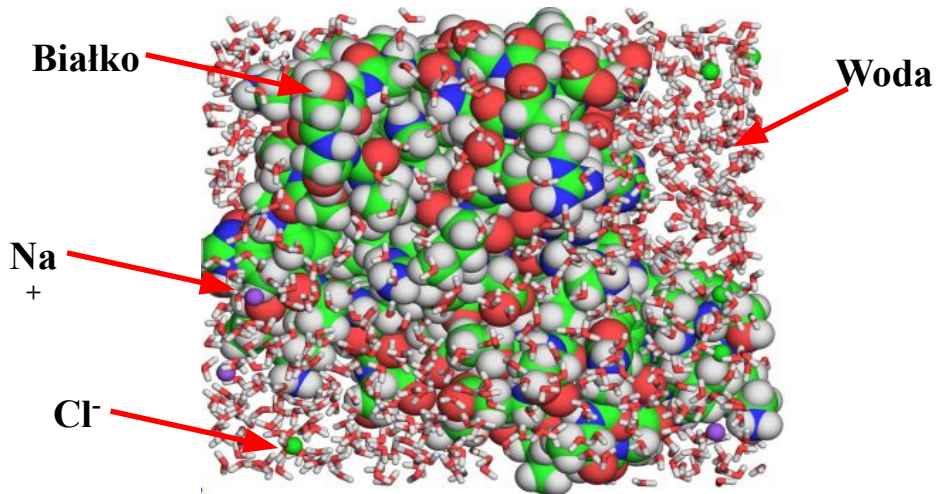
- Bezpośrednie (*explicit*) opisanie zmian struktury niepofałdowanej i jej otoczenia (wody, jonów, tłuszczu)
- Matematyczny opis całkowitej energii systemu i konfiguracji (*force field*)
- Algorytm testujący lokalne struktury w celu odzyskania tej o najniższej energii (*natywnej*)





## Podejście fizyczne

- Bezpośredni (*explicit*) opis systemu





## Podójście fizyczne

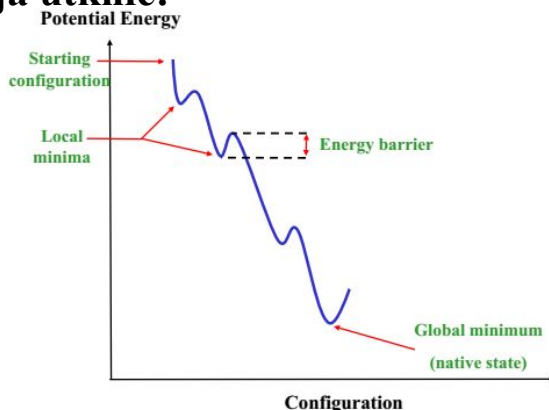
- **Próbkiwanie konfiguracji**

- **Opcja 1:** wygenerowanie wszystkich możliwych konfiguracji i identyfikacja optymalnej - oblicznowo niepraktyczne
- **Opcja 2: minimalizacja energii**
  1. Oszacowanie potencjału niepołałdowanego białka
  2. Wprowadzenie lokalnych zmian w lokalizacji atomów i (1)
  3. Niższa energia kontynuuj, brak zmian w energii lub wzrost wróć do (1)

## Podójście fizyczne

- **Próbkiwanie konfiguracji**

- **Jeżeli symulacja natrafi na lokalne minimum symulacja utknie!**



## Metody oparte o szablony

- **Struktura jest generowana w oparciu o szablony (templates), które:**
  1. Wykazują podobieństwo na poziomie sekwencji (homologs)
  2. Posiadają właściwości fizykochemiczne zgodne z analizowaną sekwencją (query)
- **Podstawowe strategie:**
  1. Modelowanie homologiczne
  2. Rozpoznawanie foldu (threading)

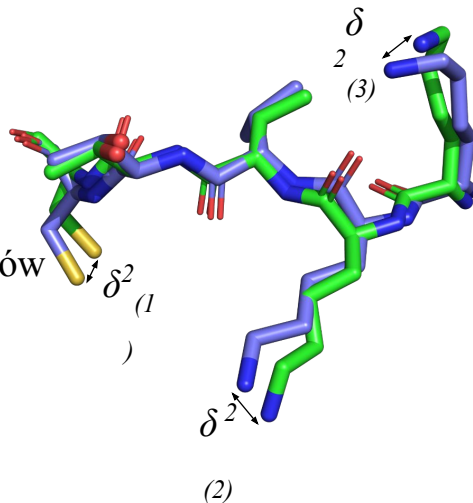
## Zgodność strukturalne

- **R.M.S.D - root-mean-square deviation**

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2}$$

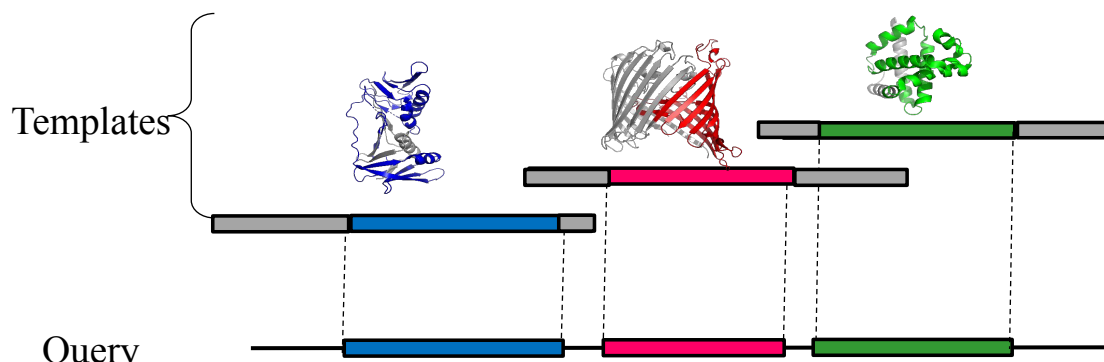
- $N$  - porównywane atomy
- $\delta_i$  - odległość [Å] pomiędzy parą atomów  $i$ :

$$\delta^2 = (x_a - x_b)^2 + (y_a - y_b)^2 + (z_a - z_b)^2$$



## Metody oparte o szablony

- Białka o podobnej **sekwencji** mają podobną **strukturę**
- Strukturę białka można przewidzieć po identyfikacji homologii na poziomie sekwencji



## Metody oparte o szablony

1. Identyfikacja szablonów > 30% zgodności (psi-BLAST)
2. Dopasowanie sekwencji
3. Integracja współrzędnych zgodnych aminokwasów
4. Optymalizacja energii (minimalizacja)
5. Ocena jakości symulacji (WHATIF, Verify3D)

## I. Identyfikacja szablonów

30% zgodności  
w dopasowaniu  
80 lub więcej  
aminokwasów.

```
>THN_DENCL
KSCCPTTAARNQYNICRLPGTFRFVCAALSGCKIISGTGCPFGYRH-
>1crnA
TTCCPSIVARSNFNVCLPGTPEALCATYIGCIIIPGATCPGDYAN-
```

Target	MAVHQIGEGGLVMYVWTFG--LMAFSALAFVMTFRPLNKRSHGYYITLA	48
3am6.1.A	-----ETGMIAQWIVFALM--RRAIAPVAVHFRPSELKS--AYYINI	46
Target	IVTIAAIYYAMAASGGKALVSNPDGNLRDIYYARYIDWFFTPPLLLLDI	98
3am6.1.A	AICTIAATAYYMAVINYQDLTNGE---RQVYARYIDW/LTTPLLLLLD	93

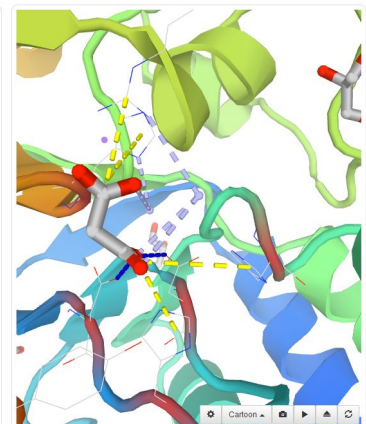
Sequence Identity: 33.80

## II. Przygotowanie szablonu

SWISS-MODEL  
Template Library  
(SMTL)

SMTL ID : 5inb.1  
RepoMan-PP1g (protein phosphatase 1, gamma isoform)  
holoenzyme complex

Coordinates	PDB Format
Method	X-RAY DIFFRACTION 1.30 Å
Oligo State	hetero-1-1-mer
Ligands	1 x ML1 <sup>12</sup> : MALONATE ION (Non-functional Binders) ^
	ML1: 6 residues within 4Å ^
	<input checked="" type="checkbox"/> Chain A: D.61, H.63, D.89, R.218, H.245, Y.269
	9 PLIP interactions: ^
	9 interactions with chain A
	<input checked="" type="checkbox"/> Hydrogen bonds: A.D.61, A.D.89
	<input checked="" type="checkbox"/> Water bridges: A.R.218, A.R.218, A.H.245
	<input checked="" type="checkbox"/> Salt bridges: A.H.63, A.H.122, A.R.218, A.H.245
Links	RCSB <sup>12</sup> PDBe <sup>12</sup> PDBj <sup>12</sup> PDBsum <sup>12</sup> CATH <sup>12</sup> PLIP <sup>12</sup>
Citation	Kumar, G.S. et al., The Ki-67 and RepoMan mitotic phosphatases assemble via an identical, yet novel mechanism. <i>Elife</i> (2016) <a href="#">DOI</a>
Release Date	2016-10-05
Peptides	Serine/threonine-protein phosphatase PP1-gamma catalytic subunit: A     Cell division cycle-associated protein 2: B
SMTL-PDB	SMTL Chain id: A B



### III. Identyfikacja optymalnego szablonu

The screenshot shows the I-TASSER interface with the following details:

- Templates:** Quaternary Structure, Sequence Similarity, Alignment, More
- ITSort:** Coverage, GMQE, QSQE, Identify, Method, Oligo State, Ligands
- Template 1:** 319y.1.A Superoxide dismutase [Cu-Zn]. GMQE: 0.80, QSQE: 0.91, IDDT: 68.83, Method: X-ray, 2.0Å, Oligo State: homo-dimer, Ligands: 2 x Zn<sup>2+</sup>.
- Template 2:** 319e.1.A Superoxide dismutase [Cu-Zn]. GMQE: 0.89, QSQE: 0.92, IDDT: 67.97, Method: X-ray, 2.0Å, Oligo State: homo-dimer, Ligands: 2 x Zn<sup>2+</sup>.
- Template 3 (Selected):** 319y.1.A Superoxide dismutase [Cu-Zn]. GMQE: 0.80, QSQE: 0.92, IDDT: 67.97, Method: X-RAY DIFFRACTION 1.80 Å, Found By: HHblits, Seq Similarity: 0.51, Bionut Oligo State: homo-dimer, Target Prediction: It is possible to build a homo-dimer. The target model is also predicted to be a homo-dimer. Build a homo-dimer (selected) or monomer.
- Build Models:** A 3D ribbon model of the protein structure is shown.
- Sequence:**

```

Target: M V V K A V C V I N G D A C G V F F E Q S S Q T P V Y G E V C G I A K G L R G F R V R E F D D I T S C H S S G F P F R
319y.1.A: M V V K A V C V I N G D A C G V F F E Q S S Q T P V Y G E V C G I A K G L R G F R V R E F D D I T S C H S S G F P F R
Target: V G K R G A P F E R E G L D R I C A Q D C F P K V R I T G S C F I L D A D D S I G R I V Y W N A D D S Q G G
319y.1.A: V G K R G A P F E R E G L D R I C A Q D C F P K V R I T G S C F I L D A D D S I G R I V Y W N A D D S Q G G
Target: I E E F Y T G R A R E S C C V Y T A T V
319y.1.A: I E E F Y T G R A R E S C C V Y T A T V

```

### IV. Ocena wyników

GMQE (Global Model Quality Estimate) is a quality estimate which combines properties from the target-template alignment and the template structure. They are combined using a multilayer perceptron trained to predict the IDDT score of the resulting model. The GMQE is available before building an actual model and thus helpful in selecting optimal templates for the modeling problem at hand. Once a model is built, the GMQE (1) in the figure above) gets updated for this specific case by also taking into account the OMEANDISCo global score of the obtained model in order to increase reliability of the quality estimation.

OMEANDISCo global score (Studer et al., (2) in the figure above) is the average per-residue OMEANDISCo score (see below) which has been found to correlate well with the IDDT score (Mariani et al.). The provided error estimate is based on OMEANDISCo global scores estimated for a large set of models and represents the root mean squared difference (i.e. standard deviation) between OMEANDISCo global score and IDDT (the ground truth). As the reliability of the prediction depends on model size, the provided error estimate is calculated based on models of similar size to the input.

OMEAN Z-score analysis (Benkert et al.) is deprecated and the GMQE and OMEANDISCo global scores should be consulted for global model quality estimates instead. It is based on 4 statistical potentials of mean force and their linear combination: the "OMEAN" score. All scores, 5 in total, are compared with what one would expect from experimentally determined structures of similar size using Z-scores (4) in the figure above). In other words: How many standard deviations from the mean is my model score given a score distribution from a large set of experimentally determined structures. Z-scores around 0.0 therefore reflect a "native-like" structure and, as a rule of thumb, a "OMEAN" Z-score below -4.0 indicates a model with low quality. This is illustrated by the "Comparison" plot (5) in the figure above). The x-axis shows protein length (number of residues). The y-axis is the "OMEAN" score. Every dot represents one experimental protein structure. Black dots are experimental structures with a "OMEAN" score within 1 standard deviation of the mean (Z-score) between 0 and 1), experimental structures with a [Z-score] between 1 and 2 are grey. Experimental structure that are even further from the mean are light grey. The actual model is represented as a red star.

The screenshot shows the Model Results page with the following details:

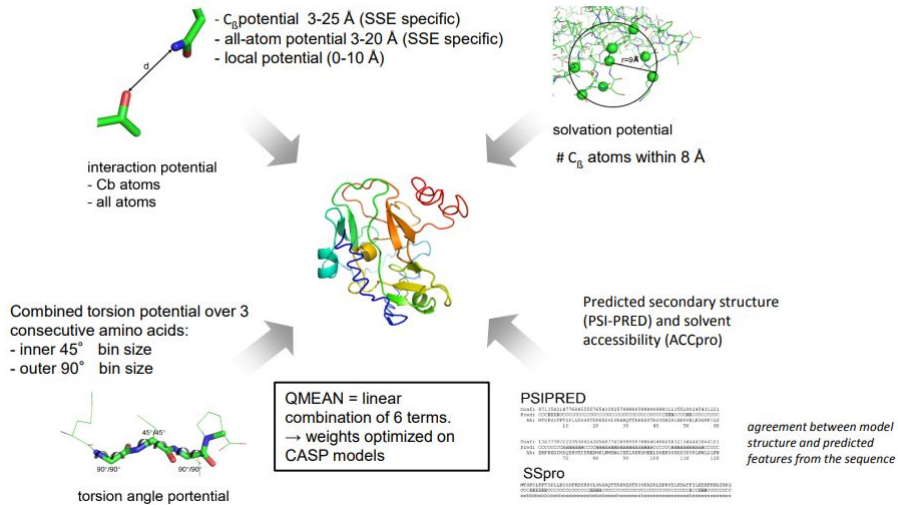
- Model 01:** Structure Assessment
- Oligo-State:** Homo-dimer
- Ligands:** 2 x CU<sup>2+</sup>, 2 x ZN<sup>2+</sup> (matching prediction)
- GMQE:** 0.88
- OMEANDISCo Global:** 0.85 ± 0.05
- 2 x COPPER (II) ION:** CU.2: 5 residues within 4Å. Chain A: H.45, H.47, H.62, V.117, H.119. 3 PLIP interactions, 3 interactions with chain A, Metal complexes: A.H.45, A.H.47, A.H.119.
- 2 x ZINC ION:** CU.4: 5 residues within 4Å. 3 PLIP interactions.
- OMEANDISCo Local:** Local Quality Estimate plot (3).
- OMEAN Z-Scores:** QMEAN: 0.37, CE: -0.17, AA Atom: -1.31, Solvation: 0.42, Torsion: -0.42 (4).
- Comparison Plot:** Protein Size (Residues) vs OMEAN score (5).
- Template:** 319y.1.A, Seq Identity: 67.97%, Coverage: [bar chart]
- Description:** Superoxide dismutase [Cu-Zn]. Crystal structures of holo and Cu-deficient CuZnSOD from the silkworm Bombyx mori and the implications in Amyotrophic lateral sclerosis.
- Model-Template Alignment:**

```


Model_01:A: M V V K A V C V I N G D A C G V F F E Q S S Q T P V Y G E V C G I A K G L R G F R V R E F D D I T S C H S S G F P F R
Model_01:B: M V V K A V C V I N G D A C G V F F E Q S S Q T P V Y G E V C G I A K G L R G F R V R E F D D I T S C H S S G F P F R
319y.1.A: M V V K A V C V I N G D A C G V F F E Q S S Q T P V Y G E V C G I A K G L R G F R V R E F D D I T S C H S S G F P F R
Model_01:A: M V V K A V C V I N G D A C G V F F E Q S S Q T P V Y G E V C G I A K G L R G F R V R E F D D I T S C H S S G F P F R
Model_01:B: M V V K A V C V I N G D A C G V F F E Q S S Q T P V Y G E V C G I A K G L R G F R V R E F D D I T S C H S S G F P F R
319y.1.A: M V V K A V C V I N G D A C G V F F E Q S S Q T P V Y G E V C G I A K G L R G F R V R E F D D I T S C H S S G F P F R

```

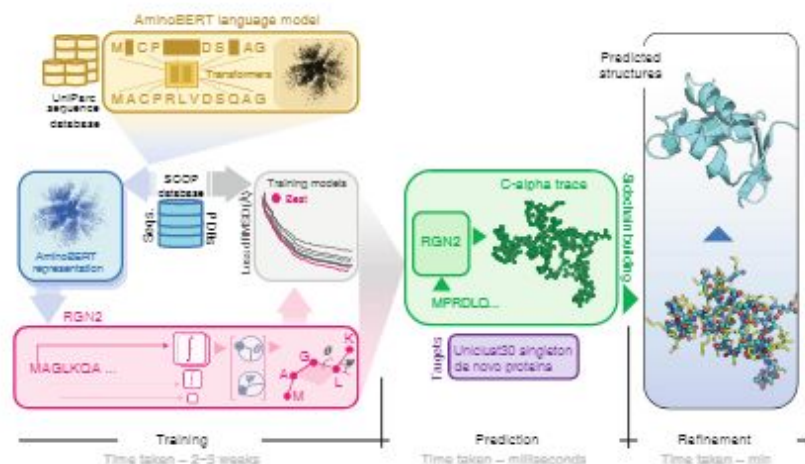
# QMEAN composite scoring function



## Single-sequence protein structure prediction using a language model and deep learning

Ratul Chowdhury<sup>1,8</sup>, Nazim Bouatta<sup>1,8</sup> , Surojit Biswas<sup>2,3,8</sup>, Christina Floristean<sup>4,8</sup>, Anant Kharkar<sup>4</sup>, Koushik Roy<sup>4</sup>, Charlotte Rochereau<sup>5</sup>, Gustaf Ahdrizt<sup>6</sup>, Joanna Zhang<sup>4</sup>, George M. Church<sup>1,2</sup>, Peter K. Sorger<sup>1,7</sup>  and Mohammed AlQuraishi<sup>4,6</sup> 

AlphaFold2 and related computational systems predict protein structure using deep learning and co-evolutionary relationships encoded in multiple sequence alignments (MSAs). Despite high prediction accuracy achieved by these systems, challenges remain in (1) prediction of orphan and rapidly evolving proteins for which an MSA cannot be generated; (2) rapid exploration of designed structures; and (3) understanding the rules governing spontaneous polypeptide folding in solution. Here we report development of an end-to-end differentiable recurrent geometric network (RGN) that uses a protein language model (AminoBERT) to learn latent structural information from unaligned proteins. A linked geometric module compactly represents C<sub>α</sub> backbone geometry in a translationally and rotationally invariant way. On average, RGN2 outperforms AlphaFold2 and RoseTTAFold on orphan proteins and classes of designed proteins while achieving up to a 10<sup>4</sup>-fold reduction in compute time. These findings demonstrate the practical and theoretical strengths of protein language models relative to MSAs in structure prediction.



**Fig. 1 | Organization and application of RGN2.** RGN2 combines a transformer-based protein language model (AminoBERT, yellow) with an RGN that uses Frenet-Serret frames to generate the backbone structure of a protein (green). After initial construction of the sidechains and hydrogen-bonded networks, refinement of the structure is subsequently performed using AF2Rank (blue).