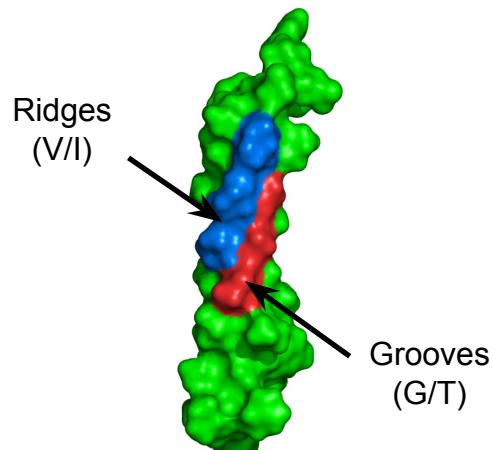
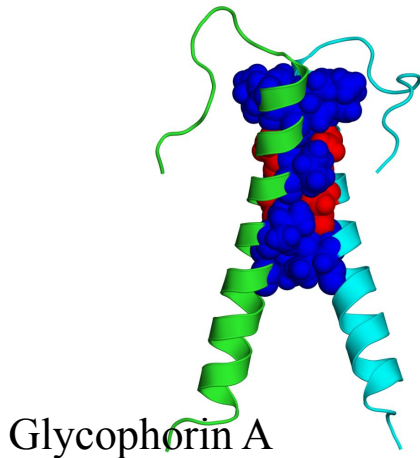


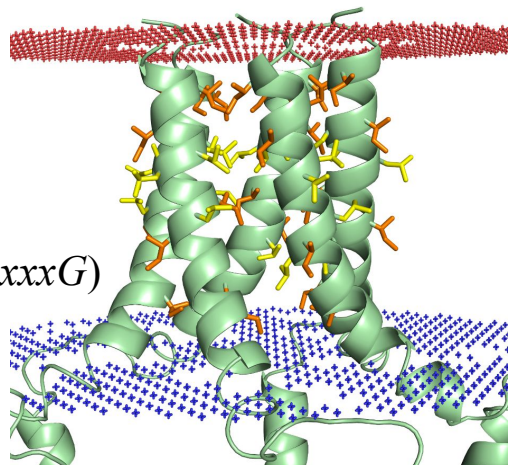
Efektywność upakowania w błonie

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



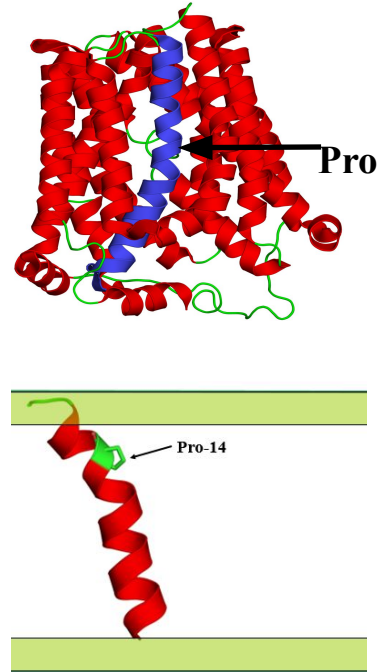
Efektywność upakowania w błonie

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



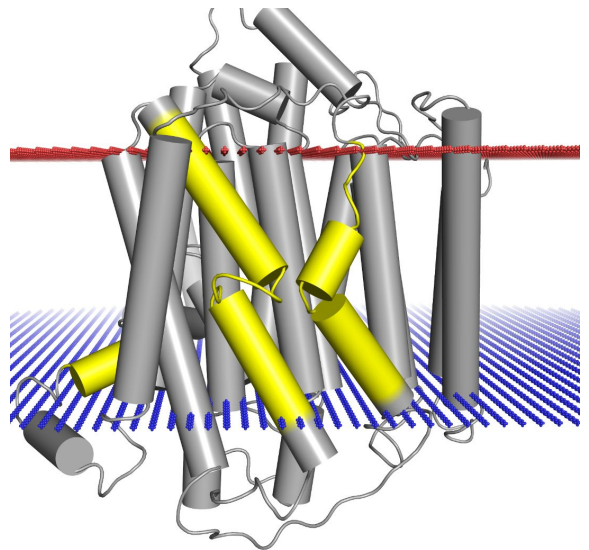
Elementary dynamiczne

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



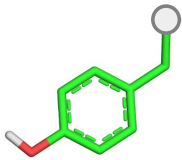
Elementary dynamiczne

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

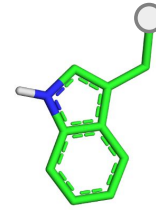
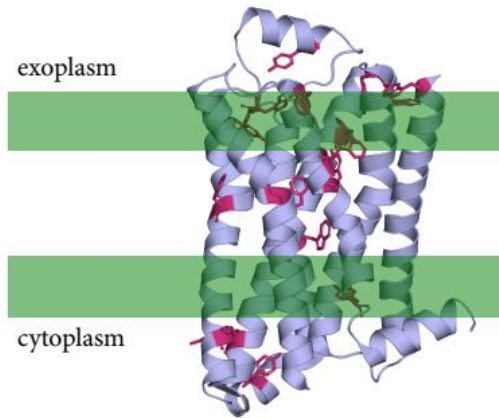


Skład jakościowy

- Pas aromatyczny - zakotwiczenie**



Tyrosine
(Tyr, Y)



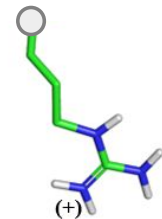
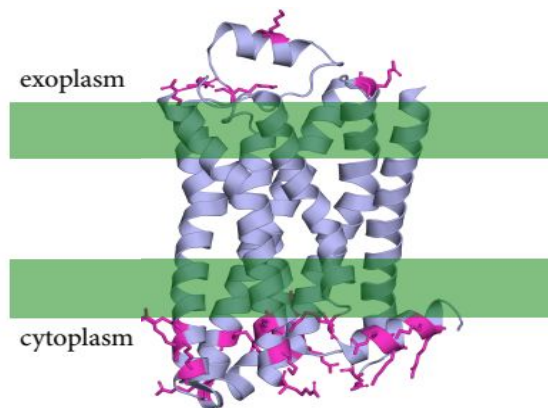
Tryptophan
(Trp, W)

Skład jakościowy

- Arg, Lys – oddziaływania zwnąrbłonowe**



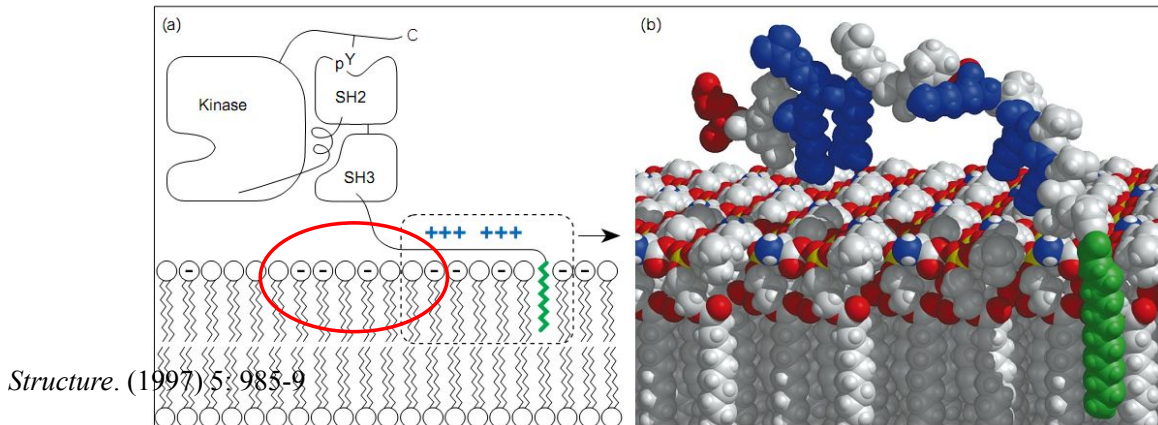
Lysine
(Lys, K)



Arginine
(Arg, R)

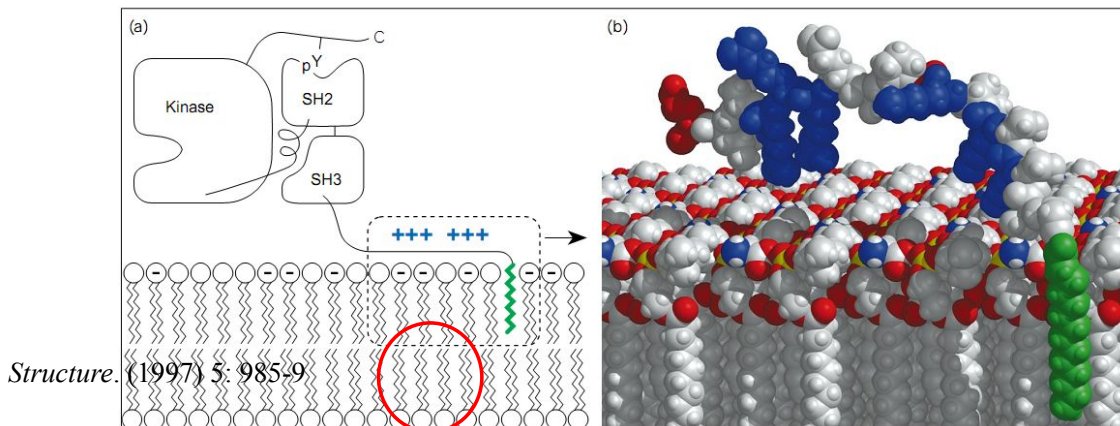
Białka
związane z błonami

- oddziaływania elektrostatyczne



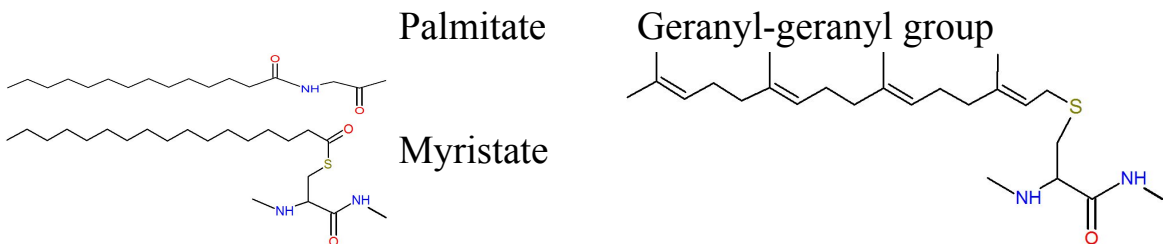
Białka
związane z błonami

- oddziaływania kowalencyjne



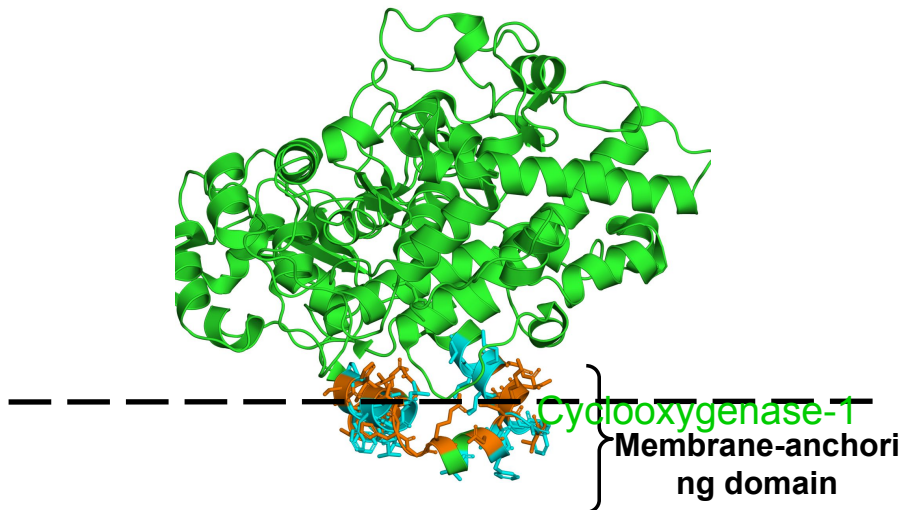
Białka
związane z błonami

- **oddziaływania kowalencyjne:**
 - kwas tłuszczowy związany z N-terminalną glicyną lub cysteiną/seryną
 - hydrofobowej reszty farnezylowej lub geranylogeranylowej przyłączona do cysteiny



Kotwiczenie
za pomocą domeny

- **Nonpolar**

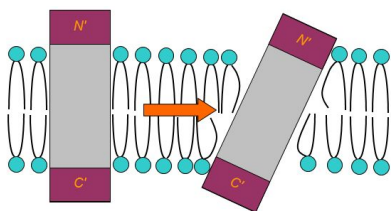


Interakcje białko-błona

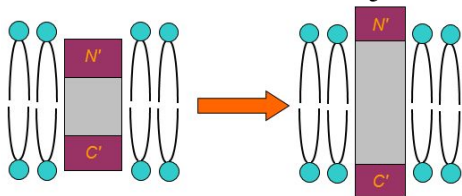
Wpływ błony na białko

- Niedopasowanie hydrofobowe**

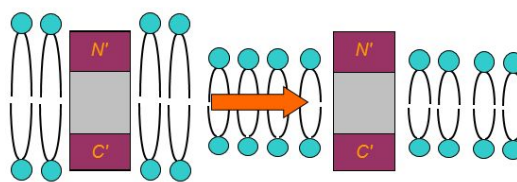
Skrzywienie



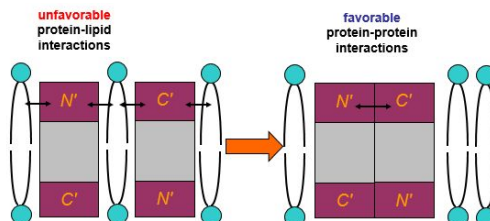
Zmiana konformacji



Dyfuzja pozioma

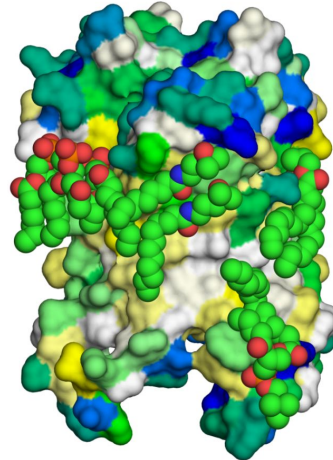


Oligomeryzacja



Wpływ błony na białko

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



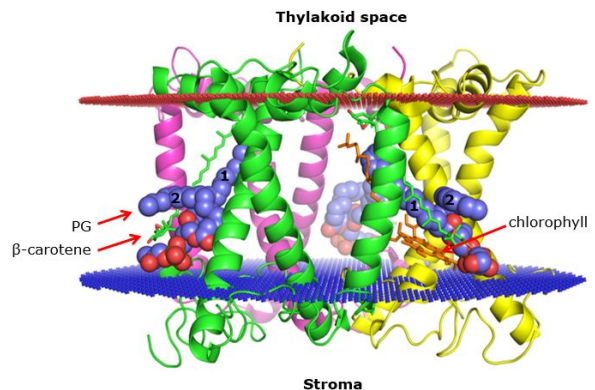
L I F V A G C S T W P Y Q H K N E D R

Hydrophobic

Hydrophilic

Wpływ kw. tłuszczowych na białko

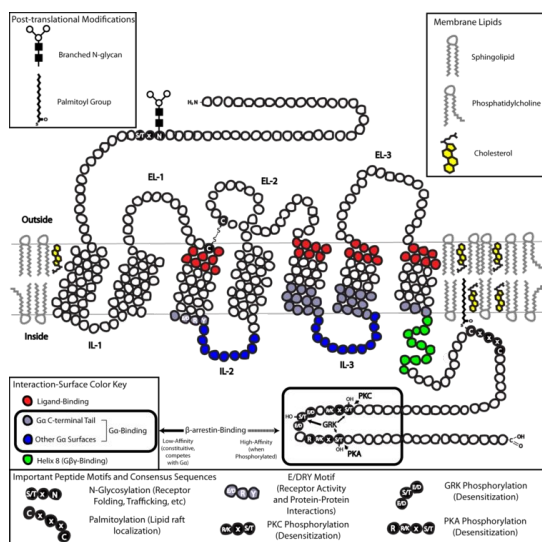
- Interakcje pomiędzy kwasami tłuszczowymi a białkami błonowymi są niezbędne dla utrzymania:
 - stabilności
 - poprawności fałdowania
 - aktywności biologicznej



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



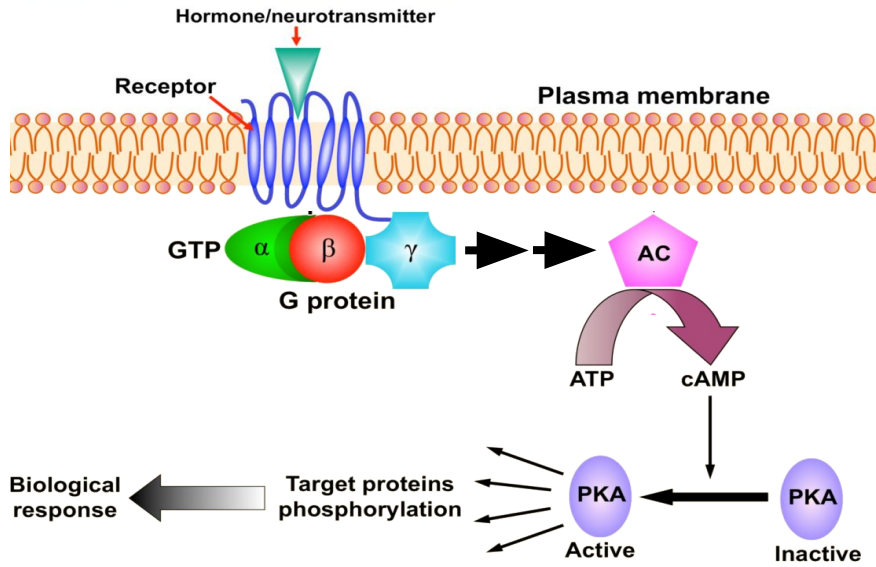
© The Nobel Foundation. Photo: U. Montan
Robert J. Lefkowitz
Prize share: 1/2



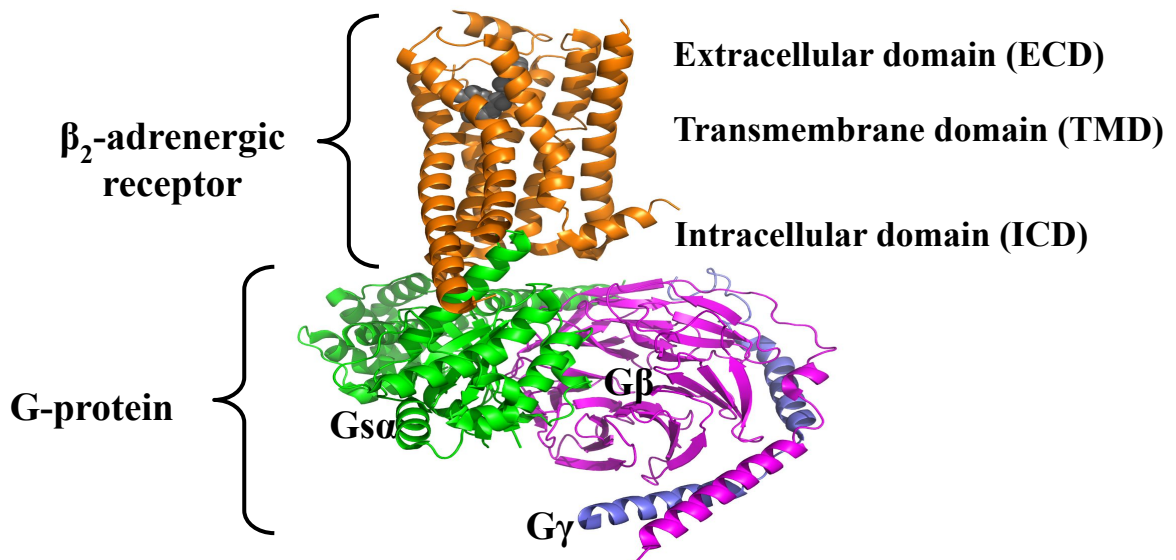
© The Nobel Foundation. Photo: U. Montan
Brian K. Kobilka
Prize share: 1/2

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

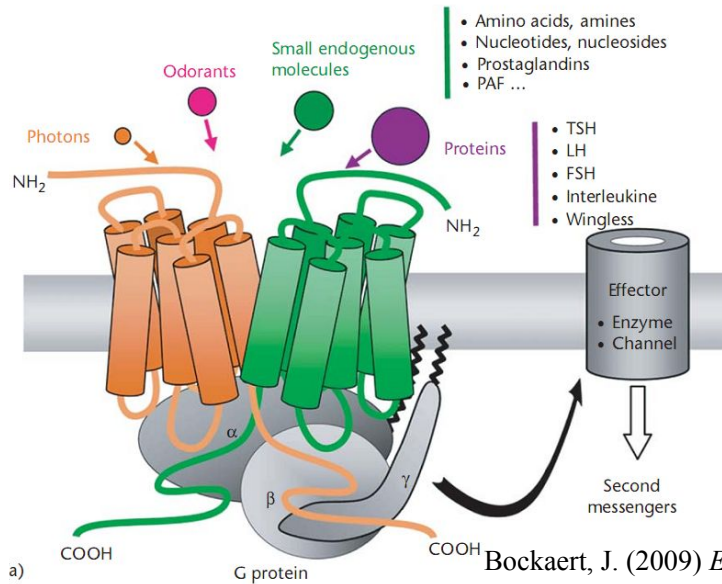
GPCRs ścieżka sygnalizacyjna



The GPCR-G protein complex



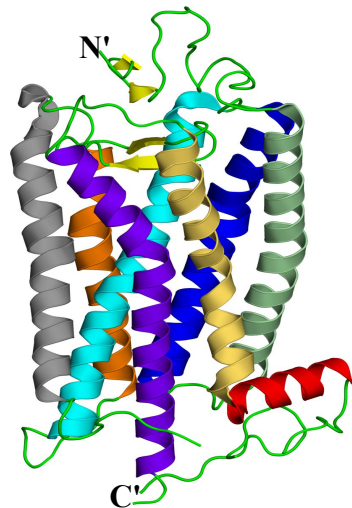
GPCRs ligandy i efektory



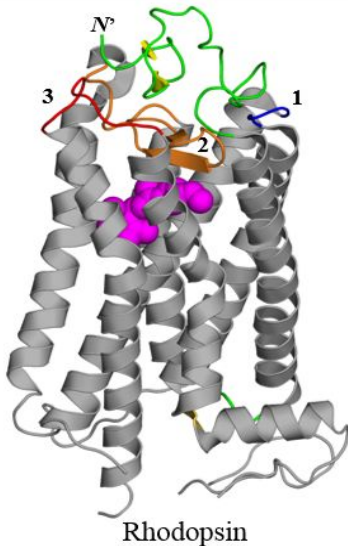
Bockaert, J. (2009) *ELS*. John Wiley & Sons, Ltd

GPCRs podobieństwo strukturalne

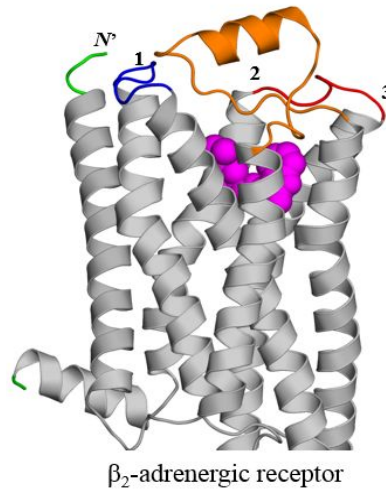
- **GPCRs mają podobne struktury pomimo niskiej zgodności sekwencji**
 - Rhodopsin & β_2 -AR: 21% zgodności,
 - TM r.m.s.d = 1.6 Å
- **TMD jest najlepiej zachowaną częścią z 7 TM helisami**



GPCRs różnice strukturalne



(ECL2 is long and blocks entrance)



(ECL2 is a short helix, entrance is open)

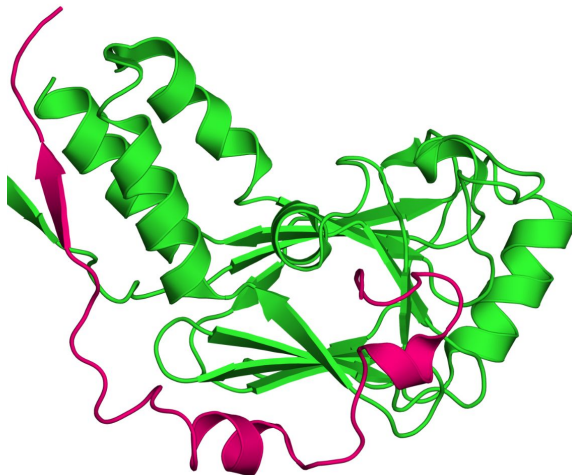
GPCRs

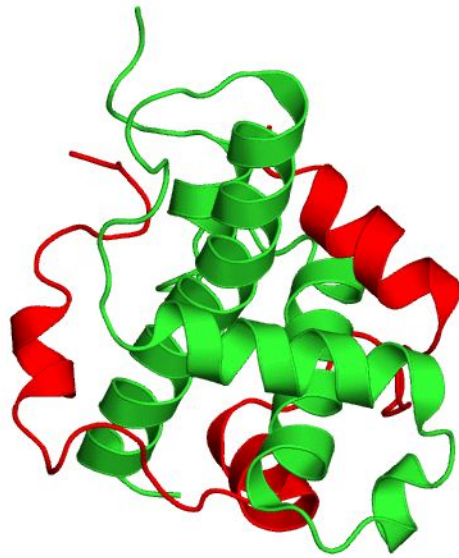
- **Specyficzne właściwości**
 - ECD: rozróżnia ligand
 - TMD:
 - ☐ specyficzna aktywacja w oparciu o motywy *NPxxY* & *CWxP*
 - ☐ kontrola allosteryczna (cholesterol)
 - ☐ dimeryzacja
 - ICD: stabilizacja stanu nieaktywnego poprzez motyw *D/ERY*

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

Brak organizacji strukturalnej w białkach

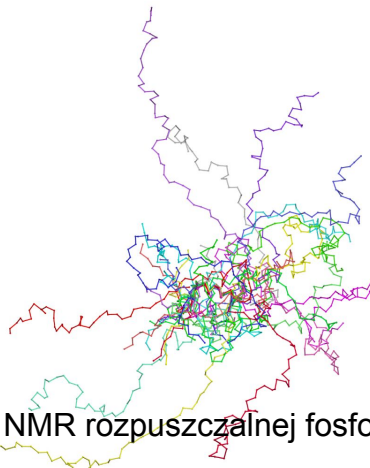
- **Regionalizacja (IDRs)**





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

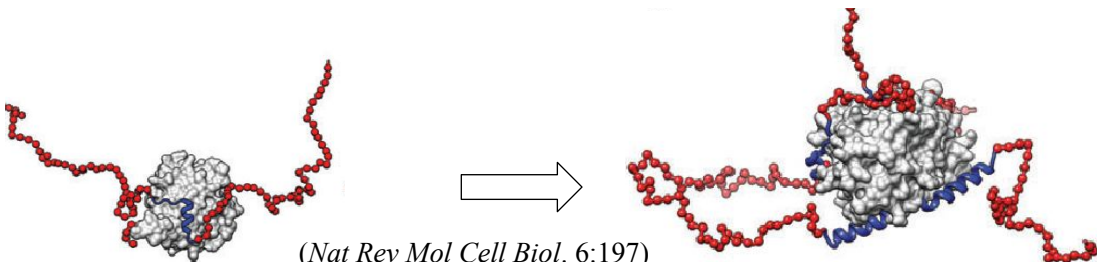
Funkcje

• 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**

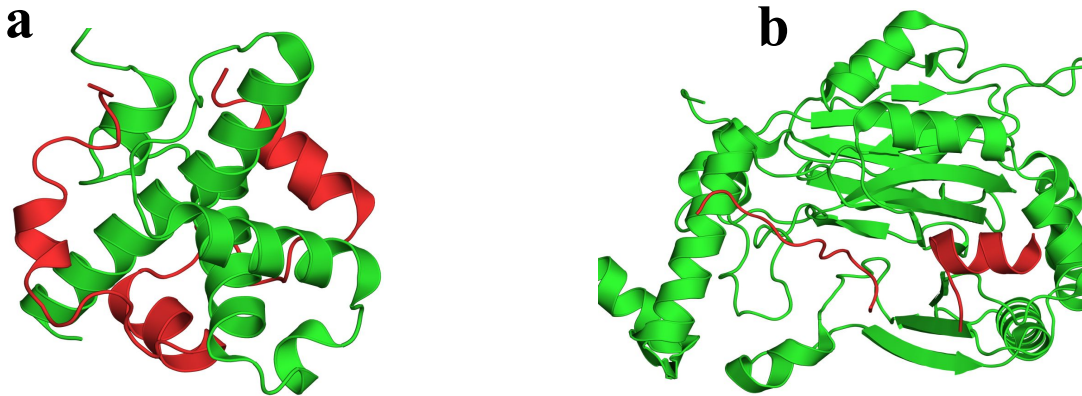
- Helisy **PII** (polyprolinne II helices) (10-20% aminokwasów)
- Motywy odpowiedzialne za wiązanie:
 - (1) **MoREs/MoRFs** (ang. molecular recognition elements/featurea, 10-70 AA, często amifatyczne),
 - (2) **SLiMs** (ang. short linear motif).

- **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:**

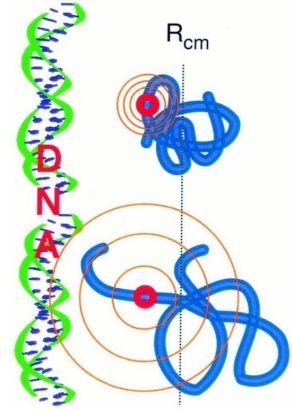
$$\Delta G_{bind} = \Delta H_{bind} - T\Delta S_{bind}$$

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

Zależność struktura-funkcja IUPs

- Kinetyka mechanizmu wiązania-fałdowanie**

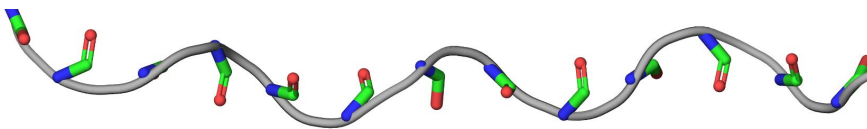
1. Wstępne słabe wiązanie poprzez szczątkowe fałdowanie
2. Wysokie powinowactwo po identyfikacji partnera i osiągnięciu optymalnej konfiguracji.



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

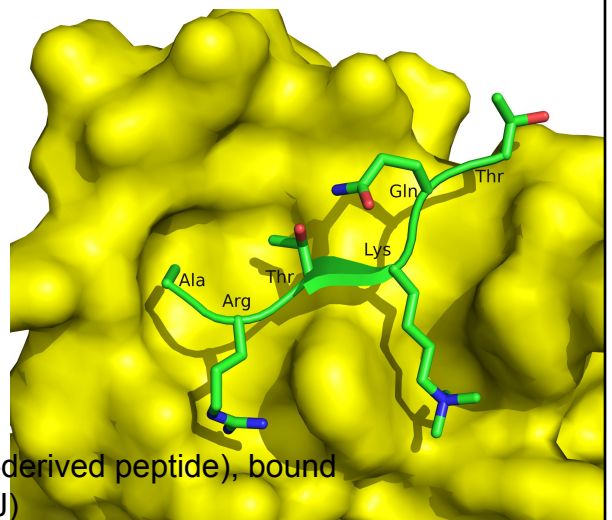
Zależność struktura-funkcja IUPs

- **Regulacja poprzez modyfikacje potranslacyjne**

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

Zależność struktura-funkcja IUPs

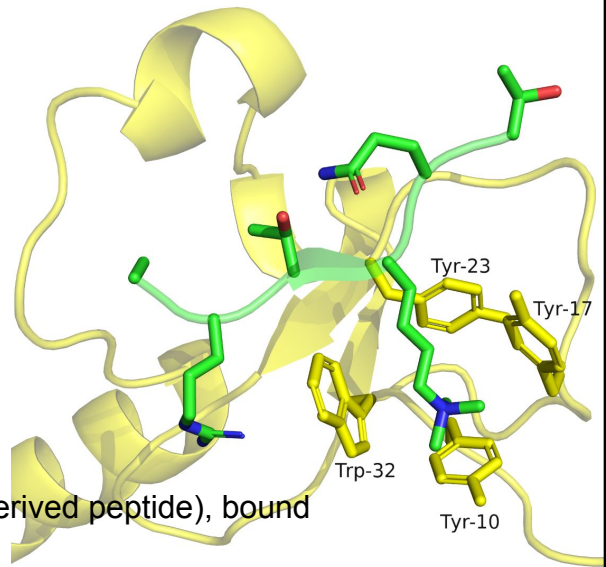
- **Regulacja poprzez modyfikacje potranslacyjne**



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

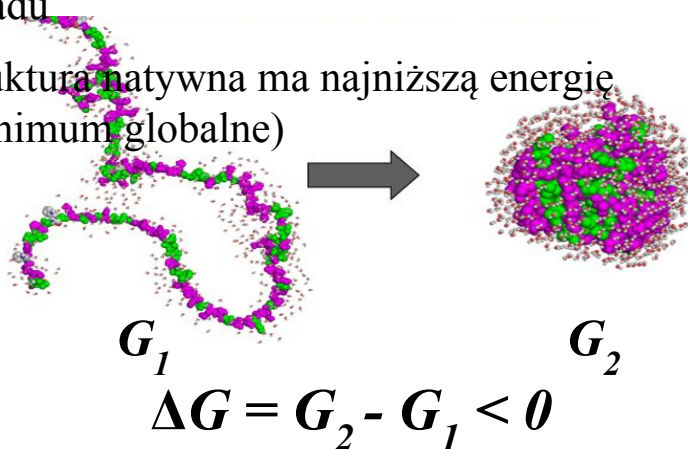
The Nobel Prize in Chemistry 2013 was awarded jointly to Martin Karplus, Michael Levitt and Arieh Warshel "for the development of multiscale models for complex chemical systems"



Podejście fizyczne

- **Założenia**

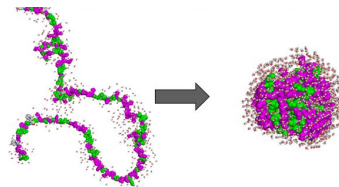
1. Fałdowanie białek bierze się z obniżeniem energii układu
2. Struktura natywna ma najniższą energię (minimum globalne)



Podejście fizyczne

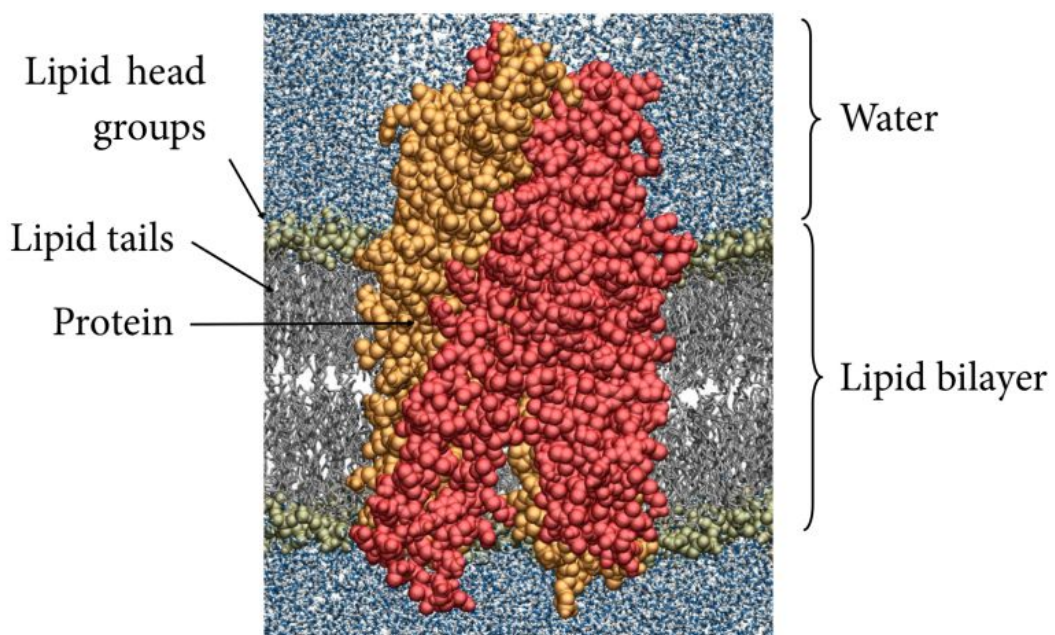
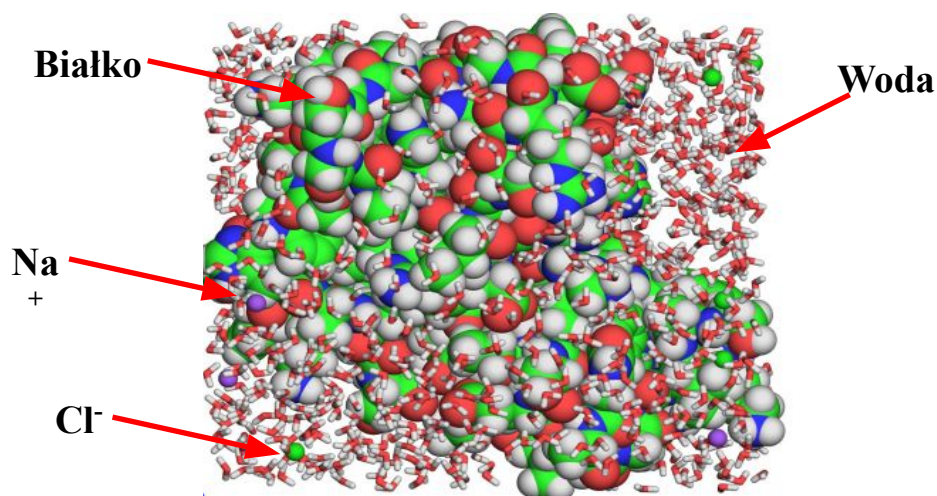
3. Proces fałdowania można opisać poprzez:

- 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 odnalezienia tej o najniższej energii (*natywnej*)



Podejście fizyczne

- Bezpośredni (*explicit*) opis systemu



Podójście fizyczne

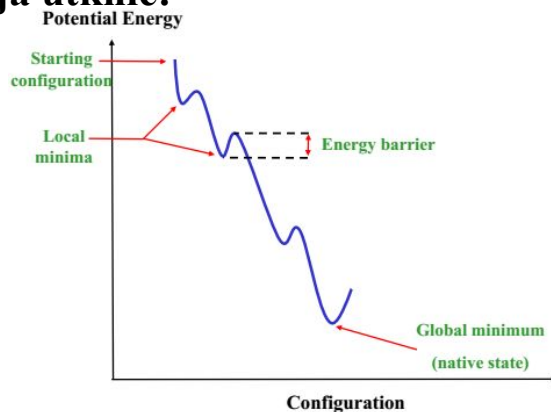
- **Próbkowanie konfiguracji**

- **Opcja 1:** wygenerowanie wszystkich możliwych konfiguracji i identyfikacja optymalnej - oblicznowo niepraktyczne
- **Opcja 2: minimalizacja energii**
 1. Oszacowanie potencjału niepołałowanego 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

- **Próbkowanie 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

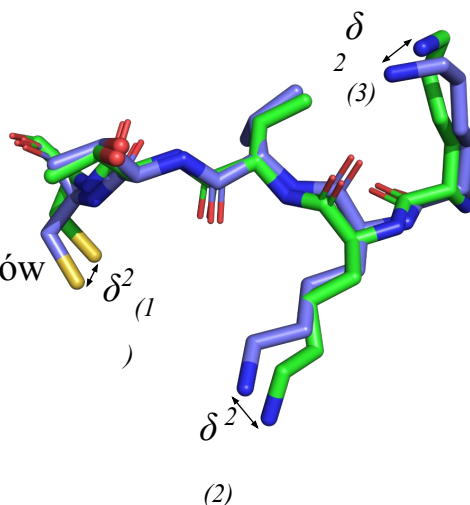
- **R.M.S.D - root-mean-square deviation**

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

□ N - porównywane atomy

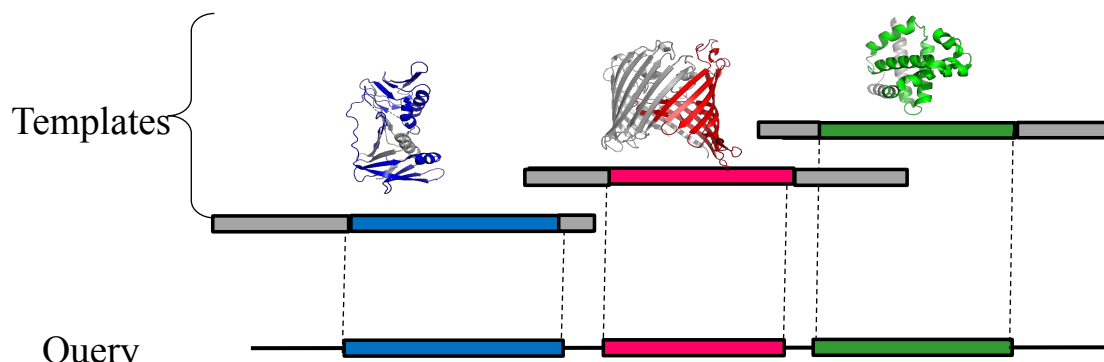
□ δ_i - odległość [Å] pomiędzy parą atomów i :

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



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. Optimalizacja energii (minimalizacja)
5. Ocena jakości symulacji (WHATIF, Verify3D)

I. Identyfikacja szablonów

30% zgodności
w dopasowaniu
80 lub więcej
aminokwasów.

	>THN_DENCL	
	KSCCPTTAARNQYNICRLPGTFRFVCAALSGCKIISGTGCPFGYRH-	
	>1crnA	
	TTCCPSIVARSNFNVCRLPGTPEALCATYTGCIIPGATCPGDYAN-	

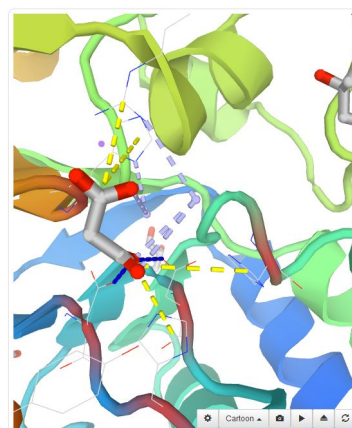
Target	MAVHQIGEGGLVMYWWTFG--LMAFSALAFVMTFRPLMKRSHGYITLA	48
3am6.1.A	-----ETGMLAQWIVFAIM-- MA AIAFQVAVHFRPSELNS-AYYINI	46
Target	IVTIAAIYYAMAASGGKALVSNPDGNLRDIYARYIDWFFTPPLLLLDI	98
3am6.1.A	AICTIAATAYYAMAVNYQDLTHNGE---RQVYARYIDW/LTTPLLLLD	93

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	POB Format
Method	X-RAY DIFFRACTION 1.30 Å
Oligo State	hetero-1-1-mer
Ligands	1 x MU ¹² : MALONATE ION (Non-functional Binders)
	ML1: 6 residues within 4Å
	☒ Chain A: D.61, H.63, D.89, R.218, H.245, Y.269
	9 PLIP interactions:
	9 interactions with chain A
	☒ Hydrogen bonds: A.D.61, A.D.89
	☒ Water bridges: A.R.218, A.R.218, A.H.245
	☒ Salt bridges: A.H.63, A.H.122, A.R.218, A.H.245
Links	RCSB ¹² PDBe ¹² PDBj ¹² PDBsum ¹² CATH ¹² PLIP ¹²
Citation	Kumar, G.S. et al., The Ki-67 and RepoMan mitotic phosphatases assemble via an identical, yet novel mechanism: Elife (2016)
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

Templates

Quaternary Structure

Sequence Similarity

Alignment

More

ITSort

Coverage

GMQE

OSQE

Identity

Method

Oligo State

Ligands

☐ 319e.1.A Superoxide dismutase [Cu-Zn]
Crystal structures of holo and Cu-deficient CuZnSOD from the silkworm Bombyx mori and the implications in Amyotrophic lateral sclerosis

☐ 319e.1.A Superoxide dismutase [Cu-Zn]
Crystal structures of holo and Cu-deficient CuZnSOD from the silkworm Bombyx mori and the implications in Amyotrophic lateral sclerosis

☒ 319y.1.A Superoxide dismutase [Cu-Zn]
Crystal structures of holo and Cu-deficient CuZnSOD from the silkworm Bombyx mori and the implications in Amyotrophic lateral sclerosis

Method

Found By

Seq Similarity

Bionut Oligo State

Target Prediction

X-RAY DIFFRACTION 1.80 Å

HHblits

0.51

homo-dimer

It is possible to build a homo-dimer. The target model is also **predicted** to be a homo-dimer.
Build a homo-dimer ☒ monomer ☐

Build Model

Target

319y.1.A

319y.1.A

319y.1.A

319y.1.A

Build Models

Clear Selection

319y.1.A

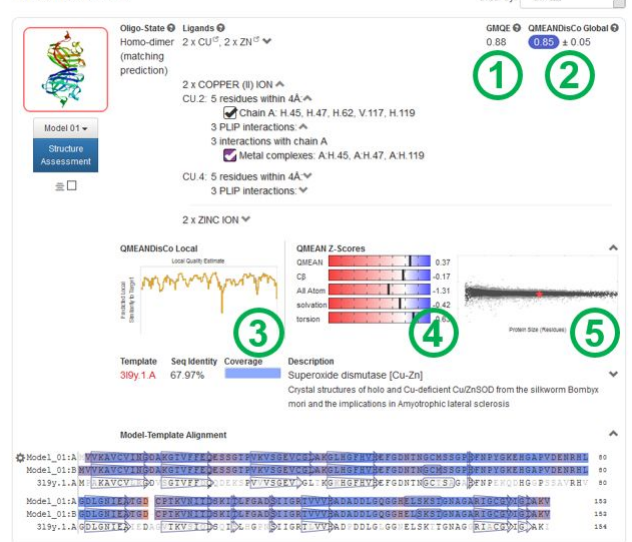
IV. Ocena wyników

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 modelling 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.

Model Results



2 x ZINC ION

OMEANDisCo Local

Local Quality Estimate

OMEAN Z-Scores

QMEAN

CE

All Atom

Solvation

Torsion

Template

Seq Identity

Coverage

Description

319y.1.A

67.97%

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

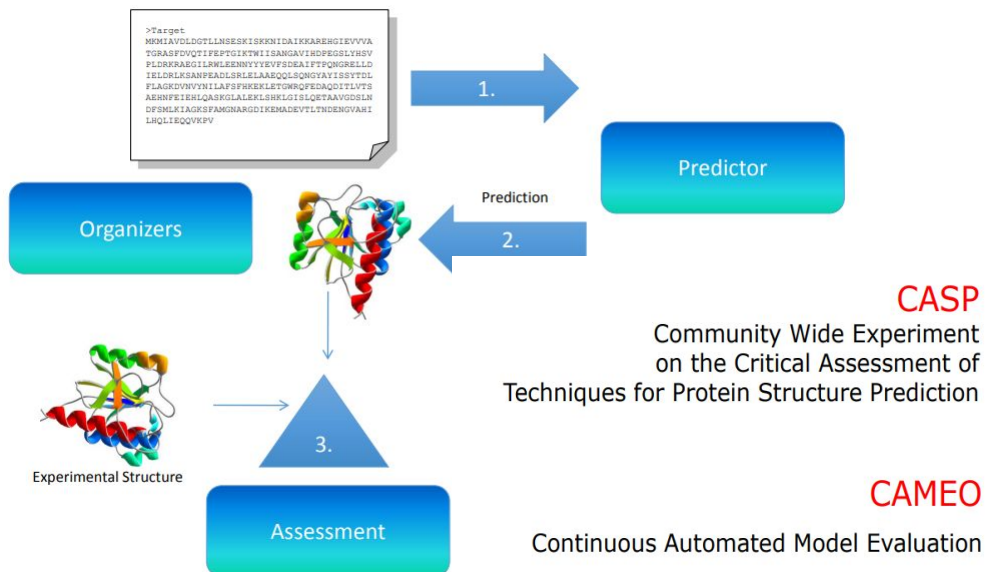
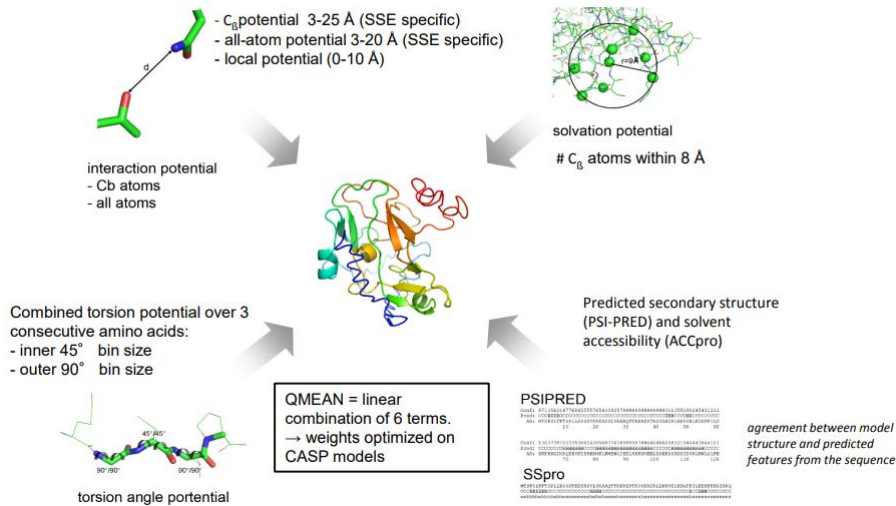
Model_01:1.A

Model_01:1.B

Model_01:1.C

319y.1.A

QMEAN composite scoring function



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

Ratul Chowdhury^{1,8}, Nazim Bouatta^{1,8}✉, Surojit Biswas^{2,3,8}, Christina Floristean^{4,8}, Anant Kharkar⁴, Koushik Roy⁴, Charlotte Rochereau⁵, Gustaf Ahdriz⁶, Joanna Zhang⁴, George M. Church^{1,2}, Peter K. Sorger^{1,7}✉ and Mohammed AlQuraishi^{4,6}✉

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_α 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 10^4 -fold reduction in compute time. These findings demonstrate the practical and theoretical strengths of protein language models relative to MSAs in structure prediction.

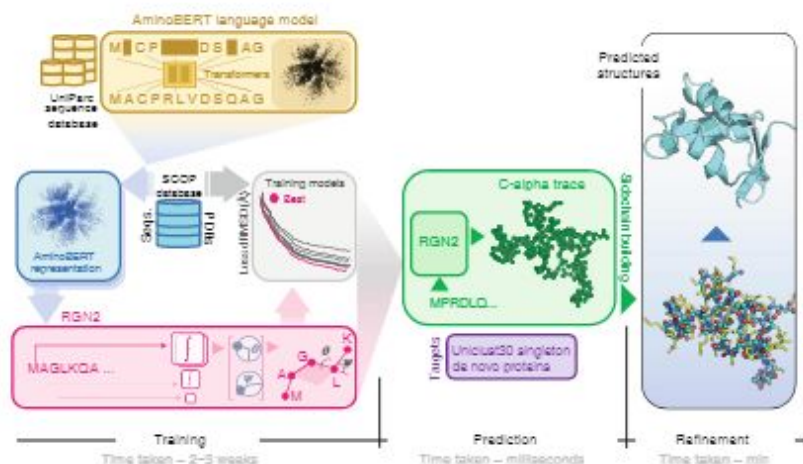


Fig. 1 | Organization and application of RGN2. RGN2 combines a transformer-based protein language model (AminoBERT, yellow) with an RGN that uses Frenet-Serret frames to generate the backbone structure of a protein (green). After initial construction of the sidechains and hydrogen-bonded networks, refinement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