

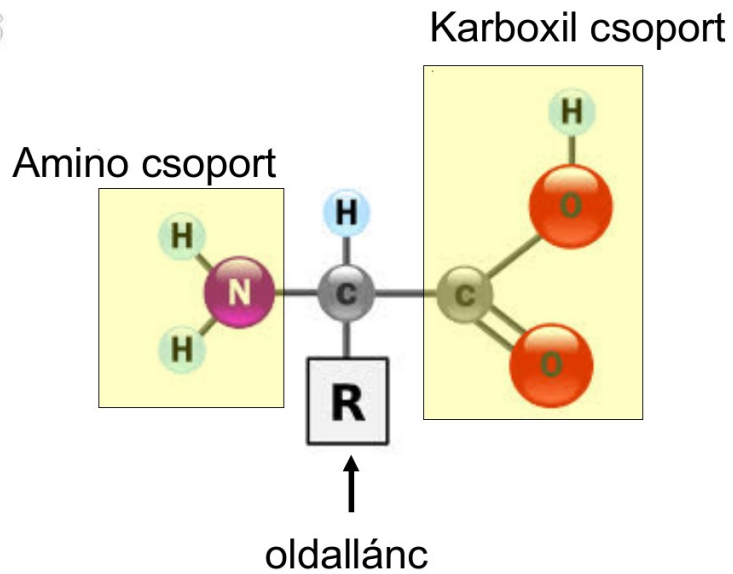
Szerkezeti bioinformatika

Dr. Dosztányi Zsuzsanna
MTA-ELTE Lendület Bioinformatika Csoport
2016. november 28.

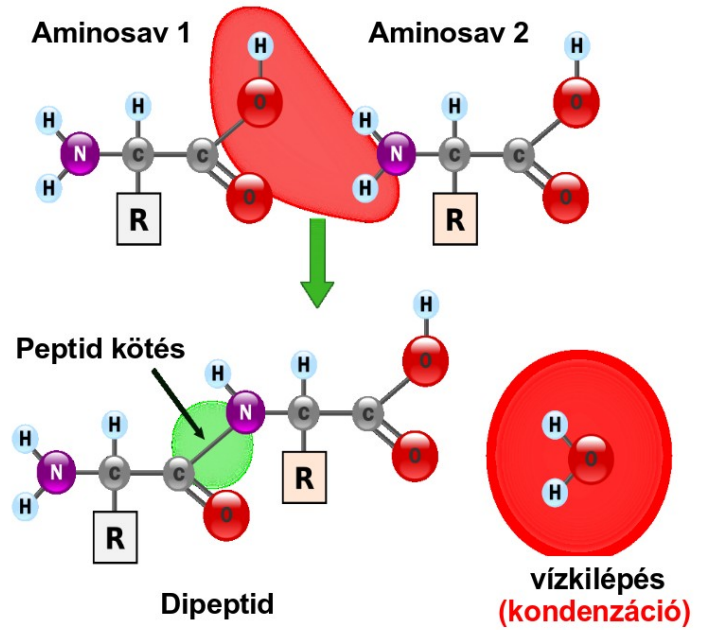
Gyakorlat: Erdős Gábor Dániel



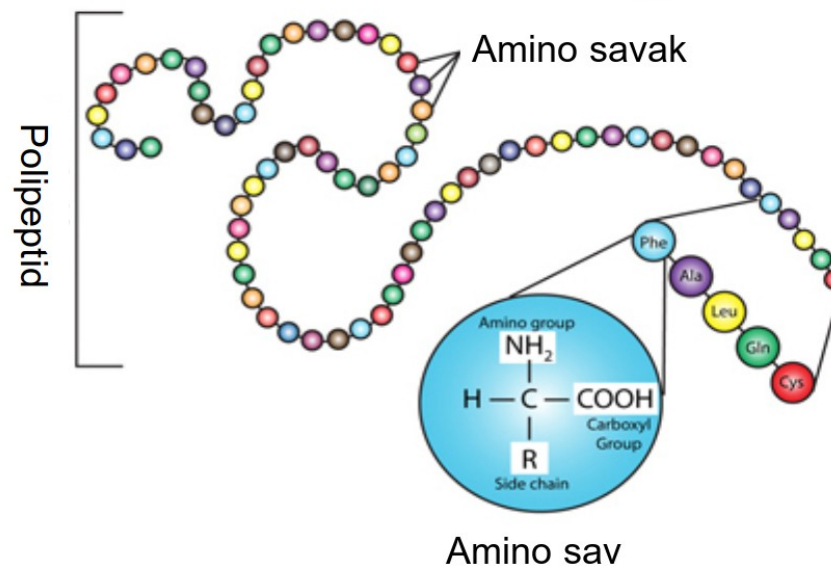
Aminosav



Peptidkötés



Polipeptidlánc



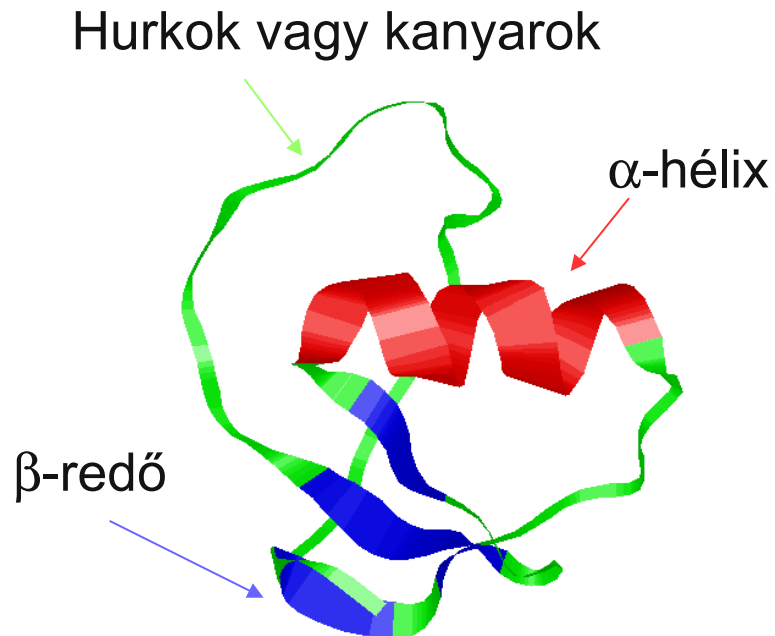
Globuláris fehérjék szerkezeti szintjei

Elsődleges szerkezet (=szekvencia):

VDCSEYPKPACPKDYRPVCGSDNKTYSNKCNCNAVV
ESNGTLTLNHFGKC
(turkey ovomucid inhibitor III domain)

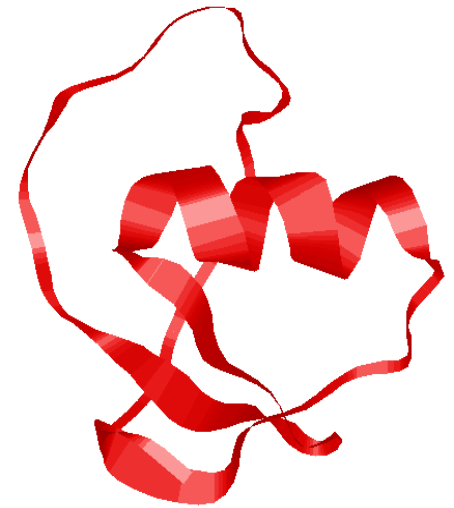
Másodlagos szerkezet:

A polipeptidlánc lokális konformációja

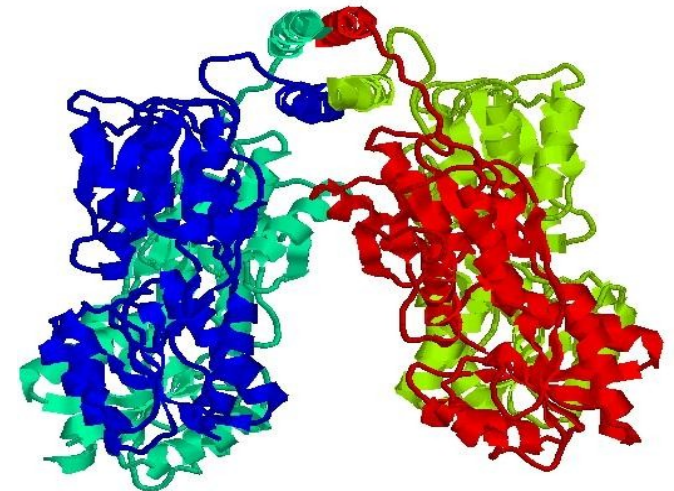


Harmadlagos szerkezet:

a molekula 3D-szerkezete

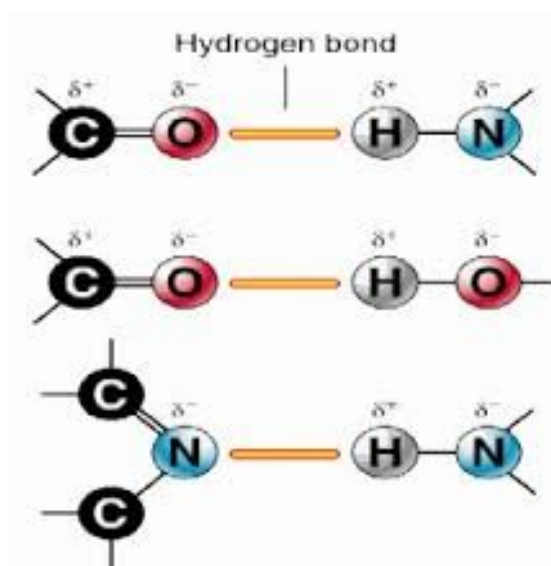


Negyedleges szerkezet: alegységösszetétel

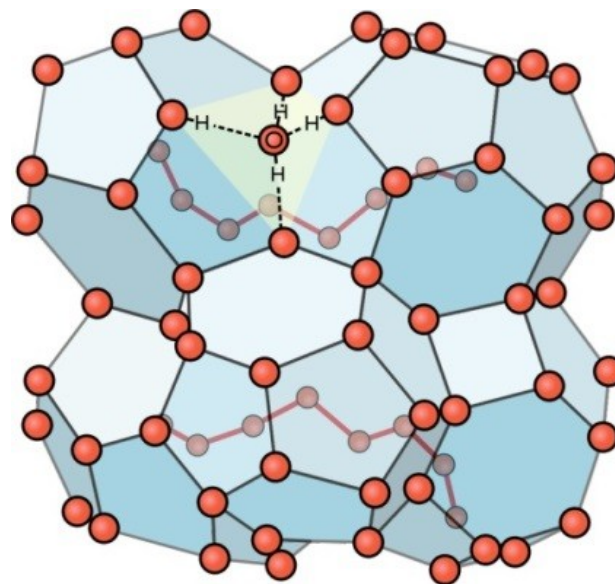
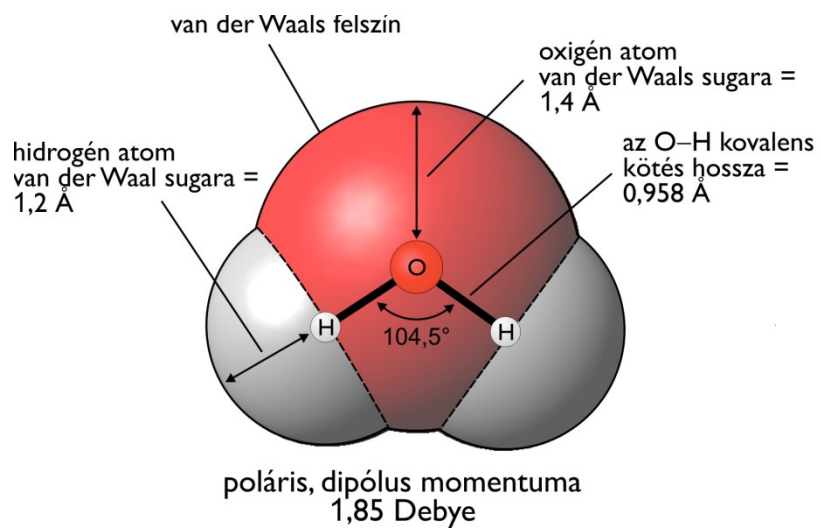


Hidrogén kötés

Hidrogénkötés (hidrogénhíd) a molekulában kötött H-atom hozza létre a kötést egy másik molekula **nagy elektronegativitású atomjával** (F, N, O), vagyis egy hidrogénatom létesít kötést két másik atom között.



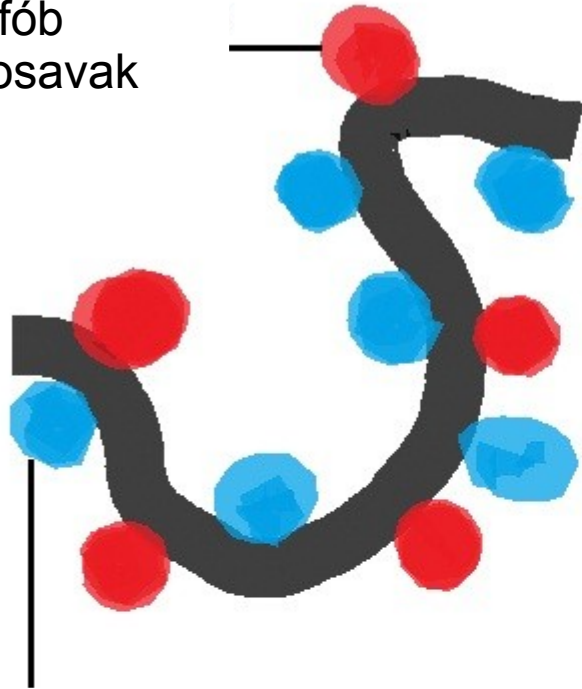
Víz molekula



Hidrofób effektus: entrópikus tényezők dominálnak

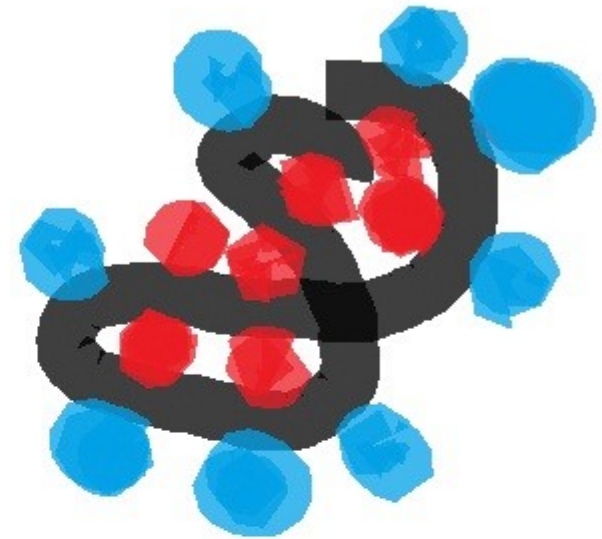
Hidrofóbicitás

Hidrofób
aminosavak



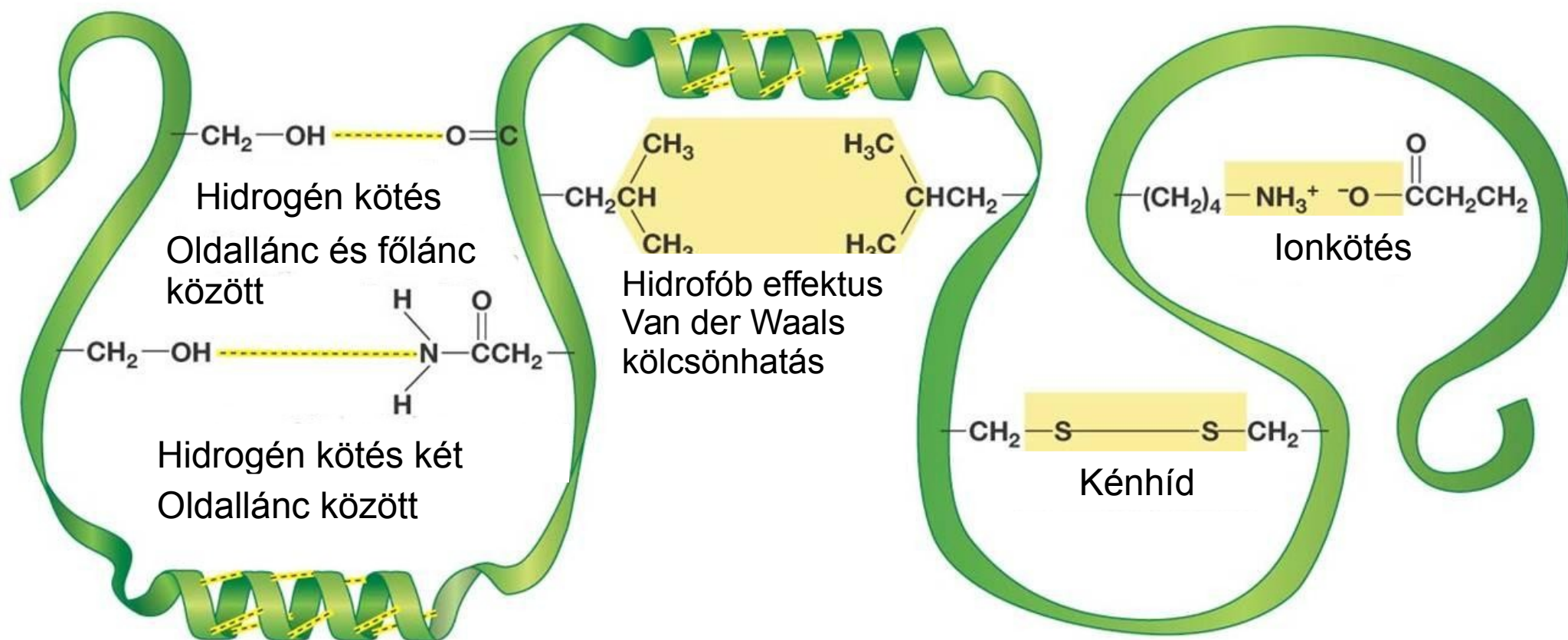
Hidrofil aminosavak

Izolált fehérje



A fehérje vizes közegben

Fehérjéket stabilizáló kölcsönhatások



Miért érdekes a fehérjék térszerkezete?



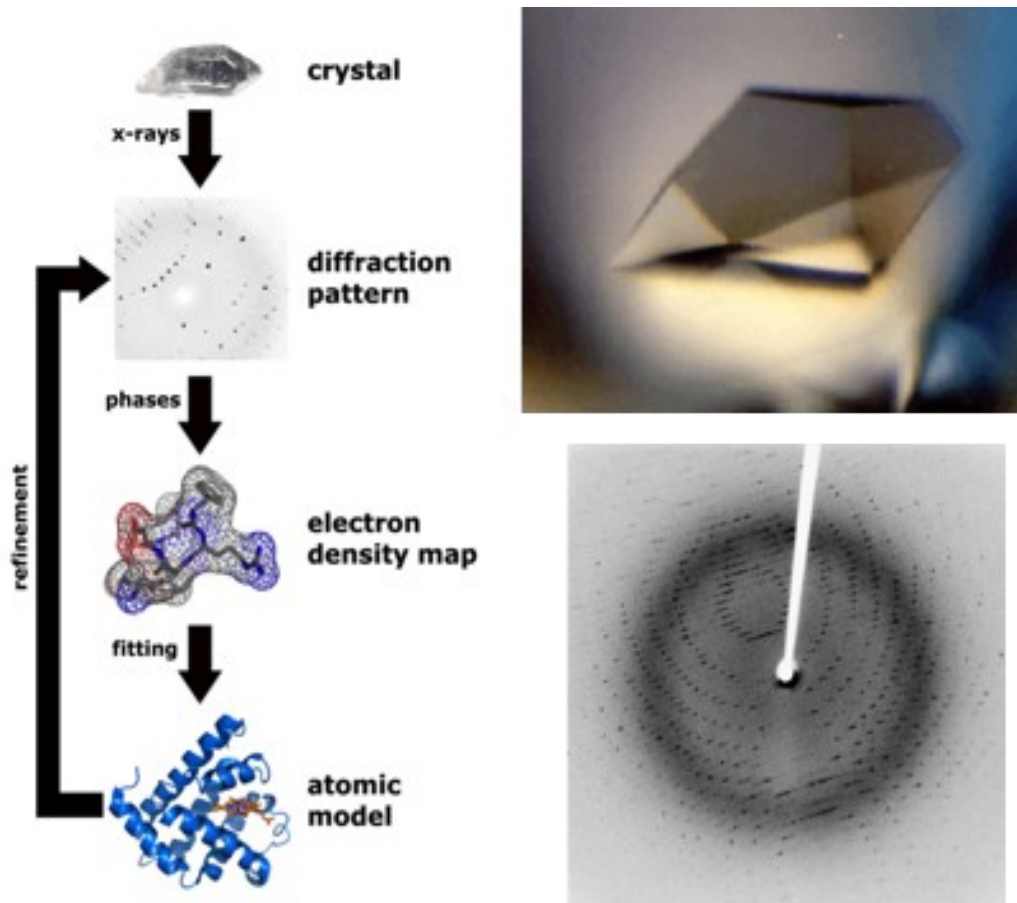
A funkció nagyban függ a fehérje alakjától

- Biológiai folyamatok atomi szintű megértése (DNA, RNS, enzimek, hormonok, receptorok)
- Betegségek molekuláris alapjainak megértése
- Gyógyszertervezés, fehérje-gyógyszer kölcsönhatás
- ✗ Nem tudja: pl. kötés erősséget

Térszerkezet meghatározási módszerek

- Röntgenkristallográfia
- NMR
- EM (elektron mikroszkópia)

Röntgenkristallográfia



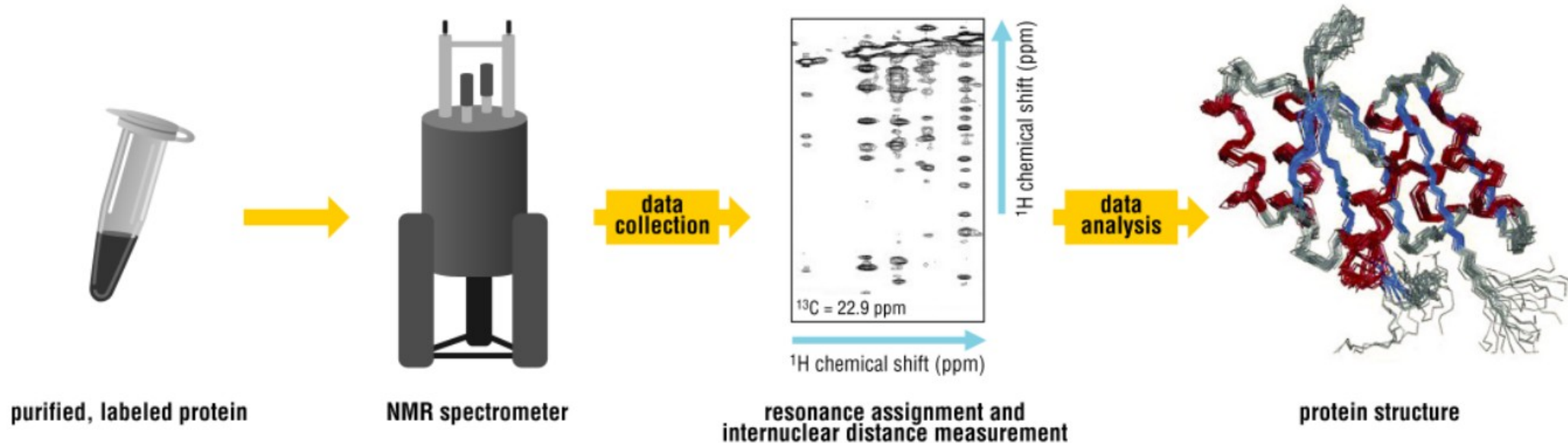
X-ray:

- A röntgensugarak hullámhossza rövid (ca. 1.5 Å), ez kell ahhoz, hogy az atomok közötti kis távolságokat mérni tudjuk
- elektron sűrűséget ad, amibe bele kell illeszteni a modellt
- kristallizációs artefaktumok
- nem fiziológiás körülmények
- Hidrogének nem látszanak

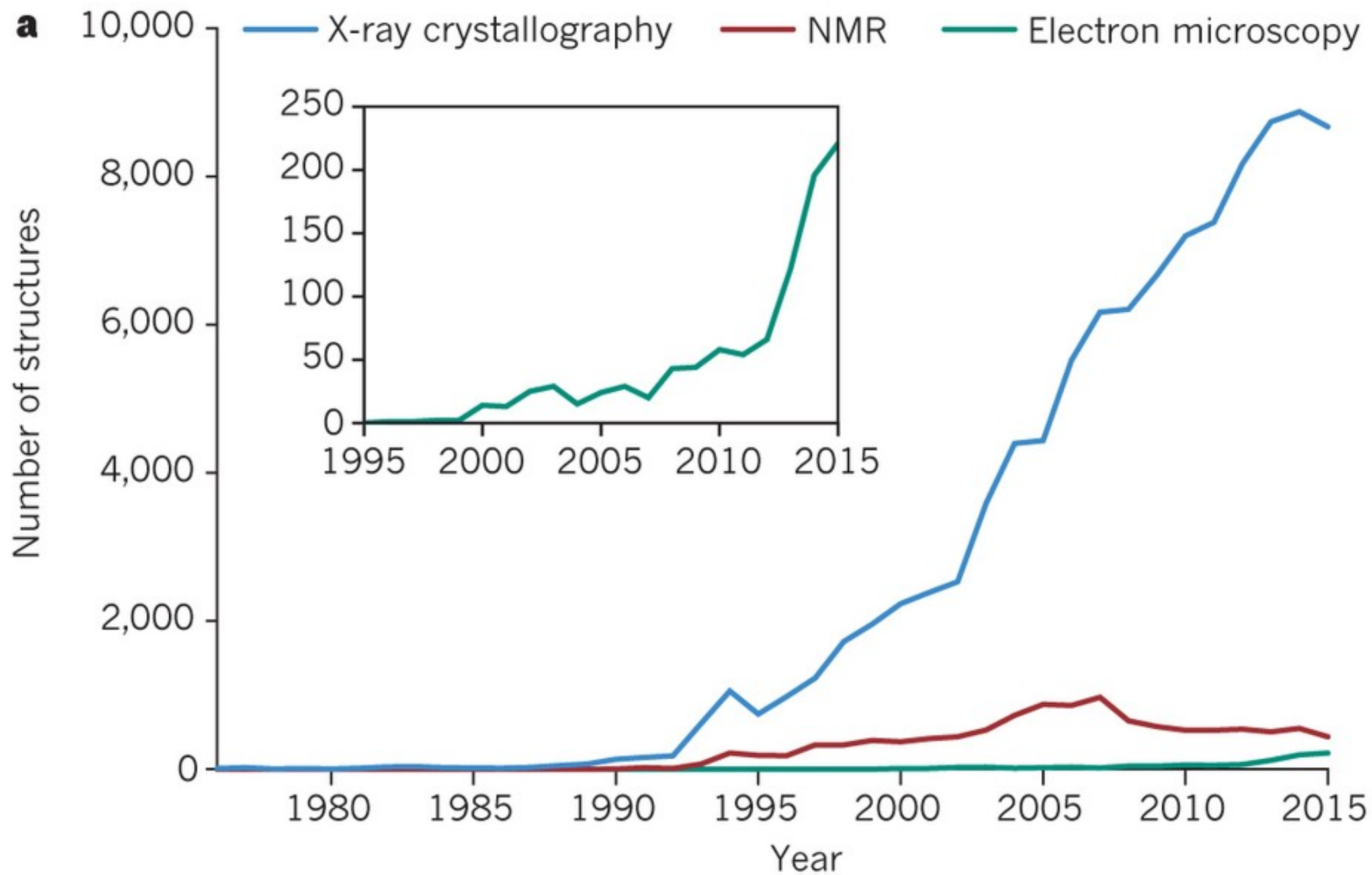
Figure 4: Left: Structure determination by X-ray crystallography. Work by Bragg and others connected spots on diffraction pattern with arrangement of atoms in the crystal to solve simple structures like salts. Work by Perutz, Rossmann, and Blow allowed automated processing of crystal data to solve complex structures. Right (top): typical protein crystal, less than one millimeter in size; Right (bottom): protein crystal diffraction pattern.

NMR

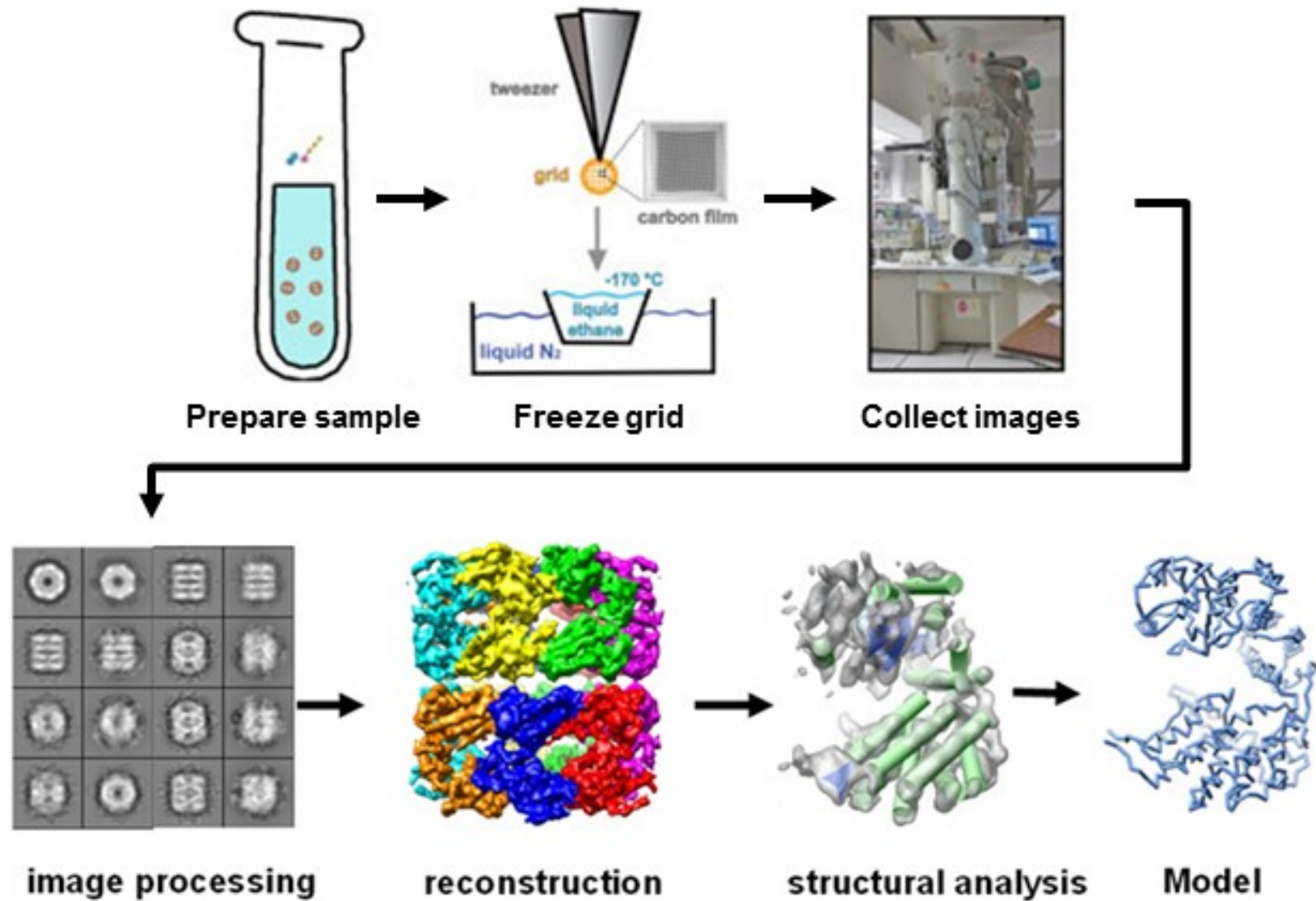
^1H -NMR (Proton Nuclear Magnetic Resonance)



- oldatban
- általában egy szerkezeti sokaságot ad vissza, amelyek kielégítik a kényszereket
- csak kis molukulák vizsgálhatóak
- kevésbé pontos modell
- flexibilisebb fehérjékre is alkalmazható



Cryo-EM (atomi felbontással)





A MEMBER OF THE PDB EMDDataBank

An Information Portal to Biological Macromolecular Structures

As of Tuesday Mar 18, 2014 at 5 PM PDT there are 98720 Structures | PDB Statistics

Search

Advanced

Browse

Everything

Author

Macromolecule

Sequence

Ligand



e.g., PDB ID, molecule name, author



Search History , Previous Results

Customize This Page

Biological Macromolecular Resource

Full Description

Learn: Featured Molecules

Hide

Structural View of Biology

List View of Archive By: Title | Date | Category



Molecule of the Month Neurotransmitter Transporters

Nerve cells communicate with one another in two ways. Some neurons send an electrical signal directly to their neighbors, which is very fast. Most neurons, however, use chemical signals to transmit their messages, releasing small neurotransmitter molecules that are recognized by receptors on neighboring neurons. Neurotransmitters have two important advantages: since thousands of molecules are released, they amplify the signal, and since many different types of neurotransmitters are used, they can encode a variety of different types of signals.

[Full Article](#)

Protein Structure Initiative Featured System

Viroporins

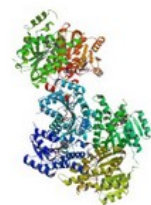


Viruses always seem to find streamlined ways of doing things. Their enzymes are often multifunctional and more compact than their cellular counterparts, and overlapping genes allow them to pack a lot of genetic information into a small space. They have also found a very parsimonious way of creating membrane channels. These simple channels, termed viroporins, are stripped to the bare minimum, with just enough molecular machinery to perform their membrane-spanning function.

[Full Article](#) | [Archive](#) | [PSI Structural Biology Knowledgebase](#)

Latest Structures

Hide



4BMS : Short chain alcohol dehydrogenase from Ralstonia sp. DSM 6428 in complex with NADPH

Structures of Alcohol Dehydrogenases from Ralstonia and Sphingobium Spp. Reveal the Molecular Basis for Their Recognition of 'Bulky-Bulky' Ketones

View in 3D (Jmol)

check out all of the latest structures released.



PDB-101 Hide

Structural View of Biology
Understanding PDB Data
Molecule of the Month
Educational Resources
Author Profiles

MyPDB Hide

Login to your Account
Register a New Account
MyPDB Help Page

Home Hide

News & Publications
Usage/Reference Policies
Deposition Policies
Website FAQ
Deposition FAQ
Contact Us
About Us
Careers
External Links
New Website Features

Deposition Hide

All Deposit Services
Electron Microscopy
X-ray | NMR
Validation Server
BioSync Beamlines/Facilities
Related Tools

Tools Hide

A .pdb fájlformátum

A fehérje adatbázis (Protein Data Bank) fájl formátum egy text formátum, ami az adatbázisban található molekulák térszerkezetét írja le.

- A fehérjék és nukleinsavak térszerkezetének leírása és annotációja, beleértve az atomi koordinátákat, oldallánc rotamereket, másodlagos szerkezeti elemeket és atomi konnektivitást.
- A szerkezetek gyakran tartalmazznak egyéb molekulákat is, mint víz, ionok, ligandumok, stb. Ezek is megadhatóak pdb formátumban.

A formátum leírása:

<http://www.wwpdb.org/documentation/format33/v3.3.html>

PDB ID: azonosító kód

A Protein Data Bank-ben minden atomi koordináta fájlhoz tartozik egy egyedi azonosító kód, ami pontosan négy karakterből áll. Az első karakter mindig szám, a következő három karakter lehet szám vagy betű.

Több, mint 400,000 4-jegyű PDB azonosító kód lehetséges (419,904 vagy 466,560 ha "0" is lehet az első karakter). Jelenlegi kb 120.000 bejegyzés van

Példák:

- 1mbn - a 1973 az első fehérje térszerkezeti modell, a **myoglobin**
- 1tna - a 1975 élesztő phenylalanine transfer RNS, az első RNS szerkezet
- 1bna - az első B-formájú **DNS** kettős hélix egy teljes csavart tartalmazó röntgen szerkezetét 1980-ban oldották meg, (27 évvel korábbi 1953-as Watson & Crick elméleti modeljét megerősítve)
- 2hhd - human **hemoglobin**, (deoxy form)
- 9ins - **insulin**

A .pdb fájlformátum

```
HEADER      EXTRACELLULAR MATRIX                      22-JAN-98    1A3I
TITLE       X-RAY CRYSTALLOGRAPHIC DETERMINATION OF A COLLAGEN-LIKE
TITLE       2 PEPTIDE WITH THE REPEATING SEQUENCE (PRO-PRO-GLY)
...
EXPDTA      X-RAY DIFFRACTION
AUTHOR      R.Z.KRAMER,L.VITAGLIANO,J.BELLA,R.BERISIO,L.MAZZARELLA,
AUTHOR      2 B.BRODSKY,A.ZAGARI,H.M.BERMAN
...
REMARK 350  BIOMOLECULE: 1
REMARK 350  APPLY THE FOLLOWING TO CHAINS: A, B, C
REMARK 350    BIOMT1    1  1.000000  0.000000  0.000000          0.00000
REMARK 350    BIOMT2    1  0.000000  1.000000  0.000000          0.00000
...
SEQRES      1  A      9  PRO PRO GLY PRO PRO GLY PRO PRO GLY
SEQRES      1  B      6  PRO PRO GLY PRO PRO GLY
SEQRES      1  C      6  PRO PRO GLY PRO PRO GLY
...
ATOM         1  N      PRO A      1          8.316  21.206  21.530  1.00 17.44      N
ATOM         2  CA     PRO A      1          7.608  20.729  20.336  1.00 17.44      C
ATOM         3  C      PRO A      1          8.487  20.707  19.092  1.00 17.44      C
ATOM         4  O      PRO A      1          9.466  21.457  19.005  1.00 17.44      O
ATOM         5  CB     PRO A      1          6.460  21.723  20.211  1.00 22.26      C
...
HETATM      130  C      ACY      401          3.682  22.541  11.236  1.00 21.19      C
HETATM      131  O      ACY      401          2.807  23.097  10.553  1.00 21.19      O
HETATM      132  OXT    ACY      401          4.306  23.101  12.291  1.00 21.19      O
...
```

A .pdb fájlformátum

ATOM	1	N	ARG	A	774	-31.629	7.797	92.108	1.00	71.22		N
ATOM	2	CA	ARG	A	774	-31.385	8.882	91.101	1.00	71.34		C
ATOM	3	C	ARG	A	774	-29.888	9.183	90.975	1.00	70.56		C
ATOM	4	O	ARG	A	774	-29.474	10.339	91.042	1.00	71.34		O
ATOM	5	CB	ARG	A	774	-32.139	10.161	91.504	1.00	71.23		C
ATOM	6	CG	ARG	A	774	-33.305	10.546	90.582	1.00	74.01		C
ATOM	7	CD	ARG	A	774	-34.689	10.154	91.158	1.00	79.24		C
ATOM	8	NE	ARG	A	774	-35.050	10.873	92.400	1.00	83.50		N
ATOM	9	CZ	ARG	A	774	-36.024	10.514	93.259	1.00	84.49		C
ATOM	10	NH1	ARG	A	774	-36.947	9.596	92.924	1.00	80.65		N
ATOM	11	NH2	ARG	A	774	-36.117	11.134	94.447	1.00	84.25		N
ATOM	12	N	ASP	A	775	-29.076	8.139	90.843	1.00	70.24		N
ATOM	13	CA	ASP	A	775	-27.627	8.302	90.918	1.00	71.14		C
ATOM	14	C	ASP	A	775	-27.042	8.263	89.530	1.00	68.46		C
ATOM	15	O	ASP	A	775	-26.321	9.177	89.118	1.00	67.65		O

atom
sorszáma, neve

aminosav típusa, láncazonosítója, sorszáma

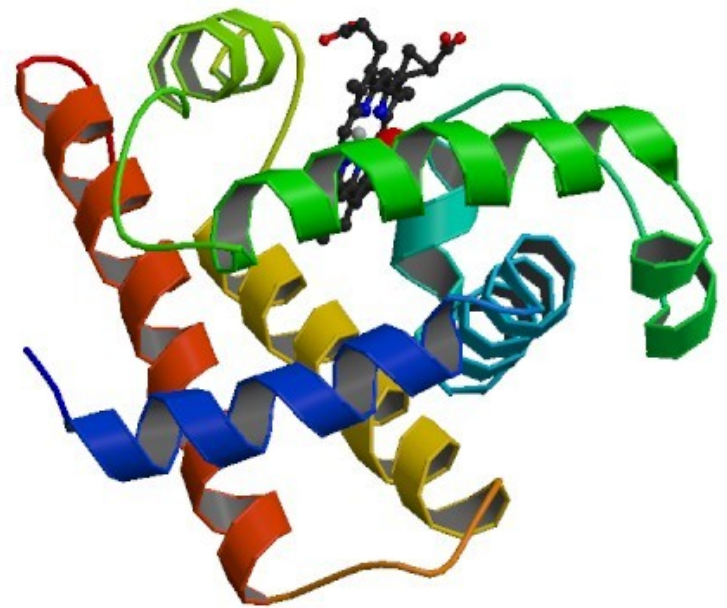
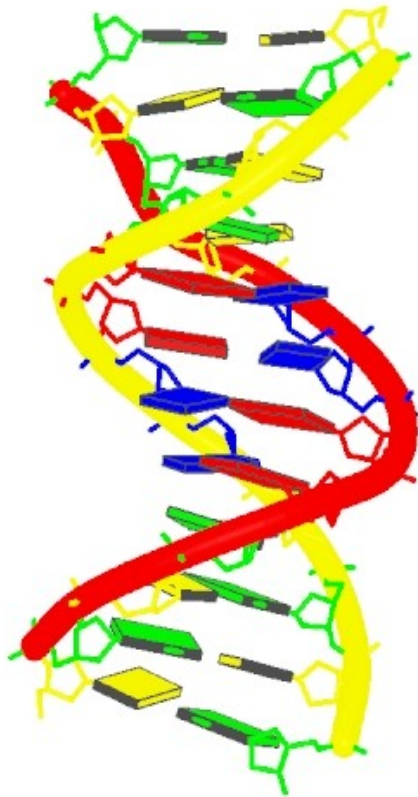
3D koordináták

betöltöttség

B-faktor

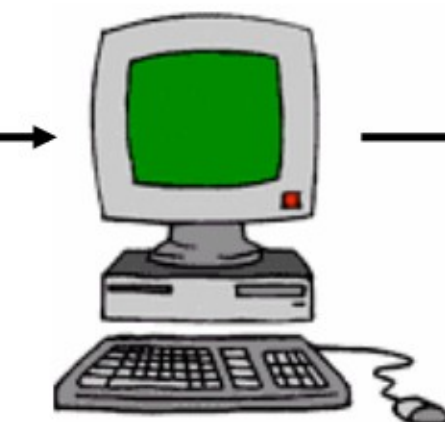
atom
típusa

DNS vs Fehérje

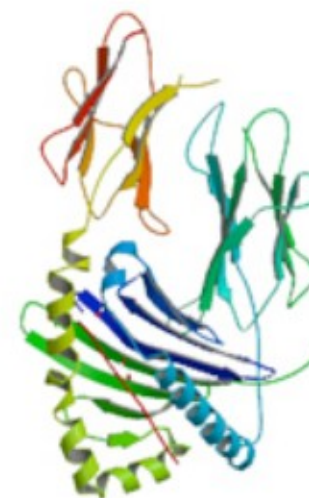


Fehérje szerkezetek megjelenítése

1. PDB koordináta fájl



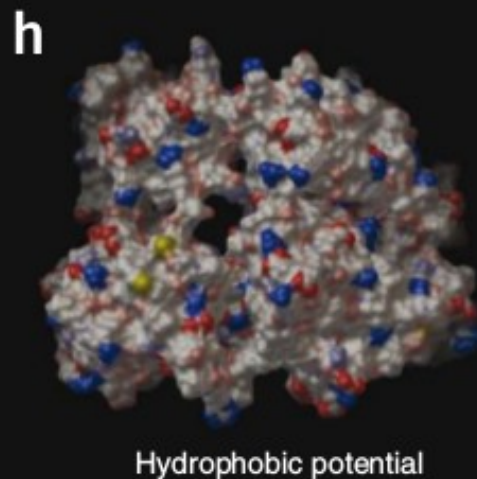
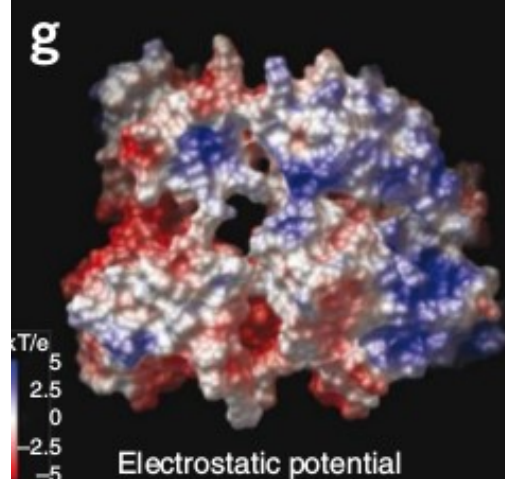
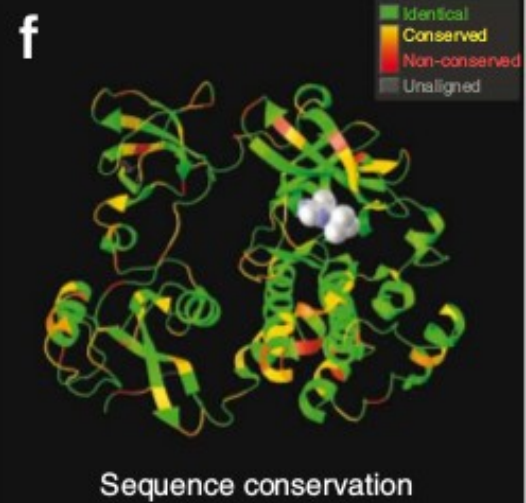
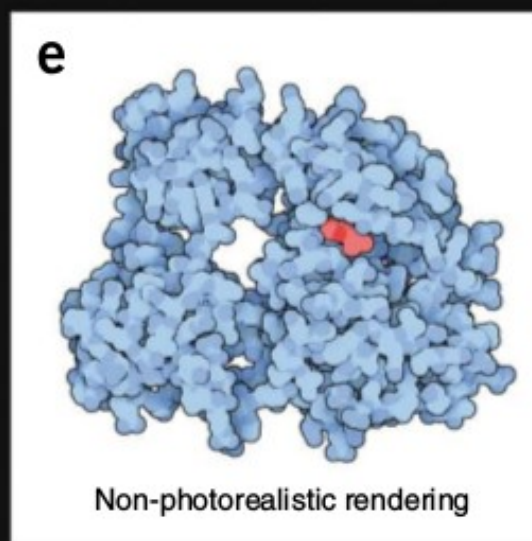
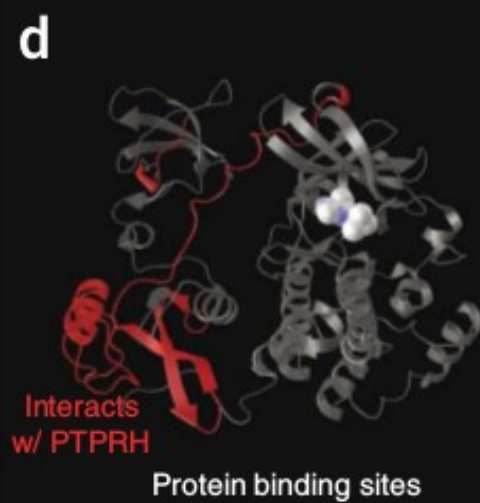
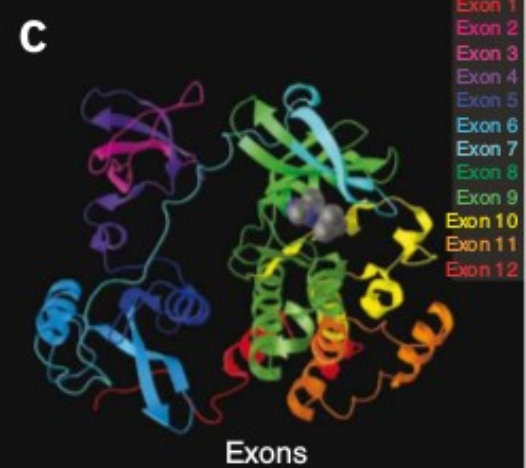
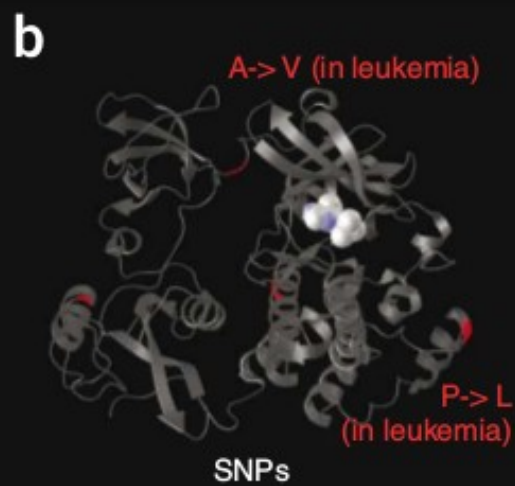
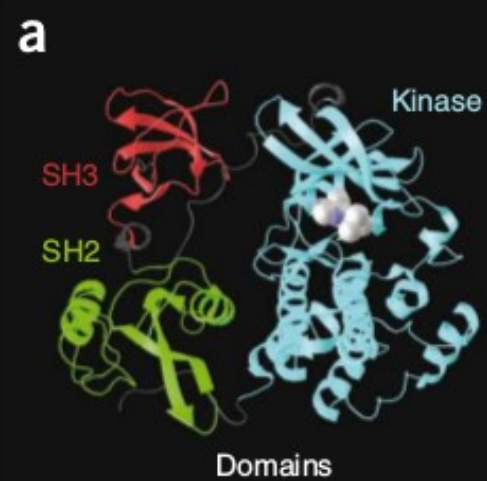
3. Számítógép



4. Molekula kép

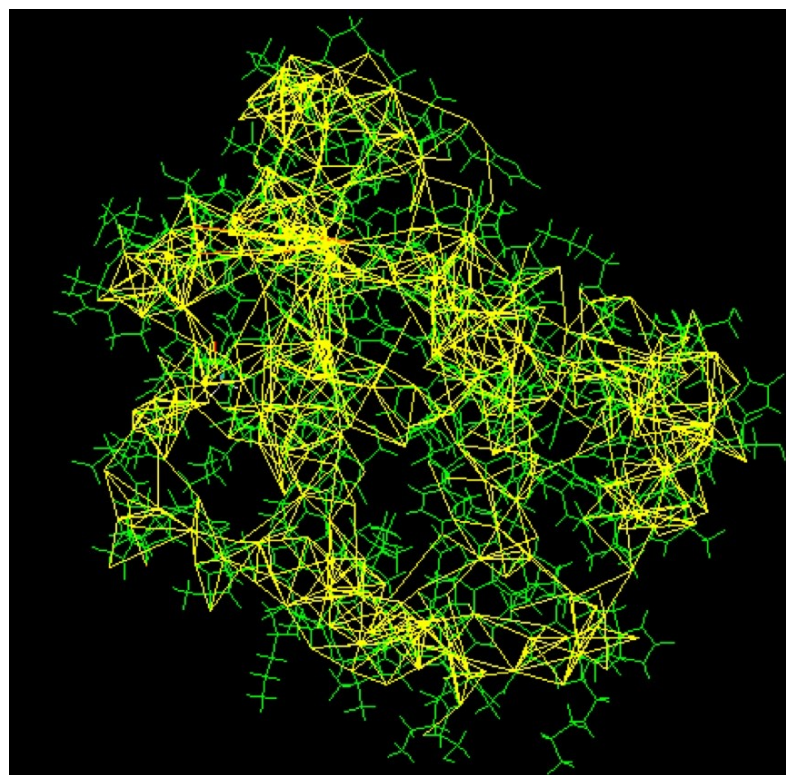
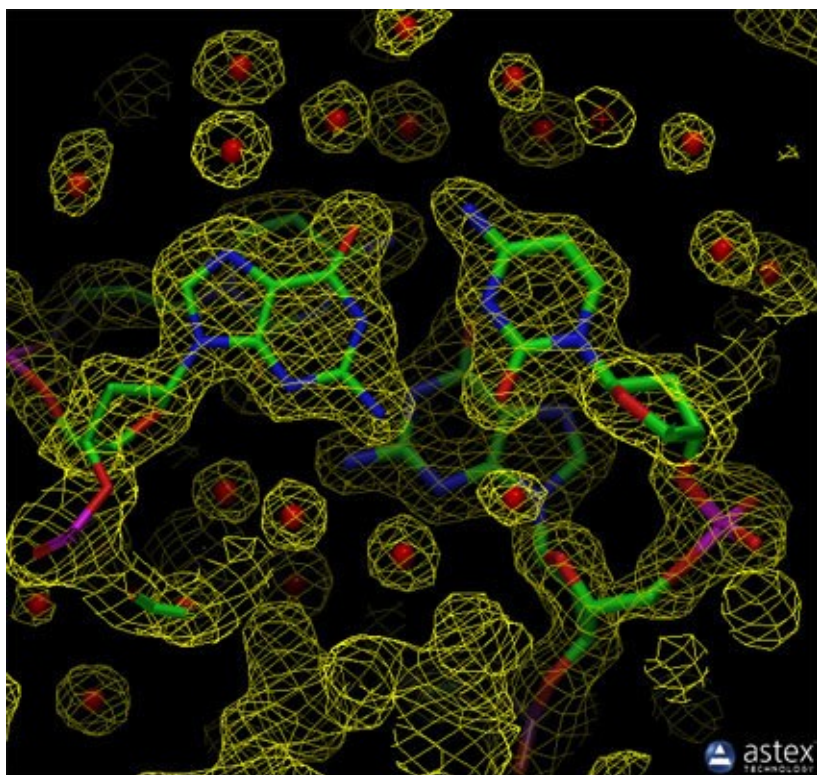
2. Vizualizációs program

Pl: Rasmol, Pymol, Chimera,
VMD, Jmol, Swiss PDB viewer



Modell

Minden fehérje szerkezet modell! A szerkezeteket a kísérletekből nem direkt módon határozzák meg, hanem az összegyűjtött adatokból olyan modellt generálnak, amely minél inkább összhangban van az adatokkal.



Felbontás (X-ray)

- Jellemezi, hogy mennyire bízunk meg az egyes atomi koordinátákban

Nagyon alacson: <4Å

Az egyedi koordináták nem értelmezhetőek

Alacsony: 3.0-4.0.0Å

Fold felismerhető

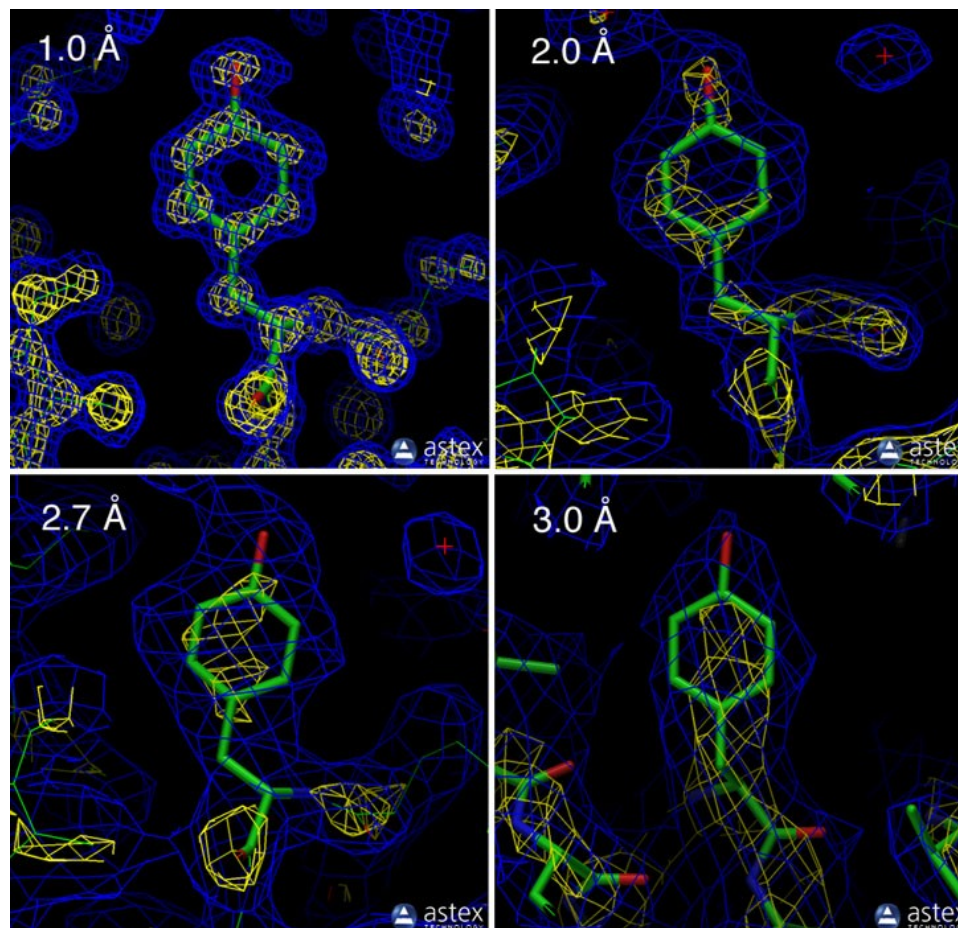
Átlagos: 1.8-3.0Å

Szerkezet nagy része korrekt,
de vannak benne rossz rotamerek,
felszíni loopok konformációja nem megbízható

Jó: 1.0 – 1.8Å

Atomi szintű: >1.0Å

A felbontás pozícionként változik!



Szerkezetek minőség jellemzése

Várt eloszlás :

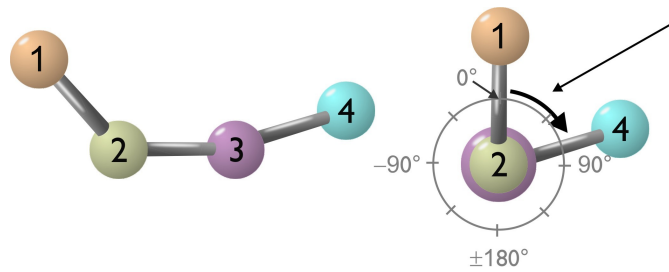
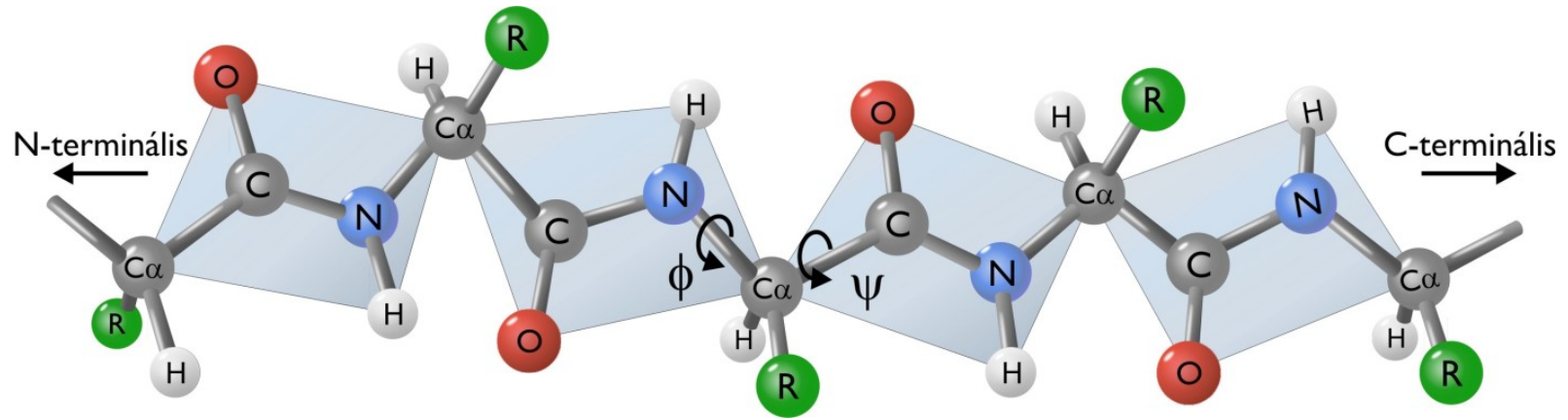
Kismolekulák alapján

Ismert, jó felbontású szerkezetek alapján

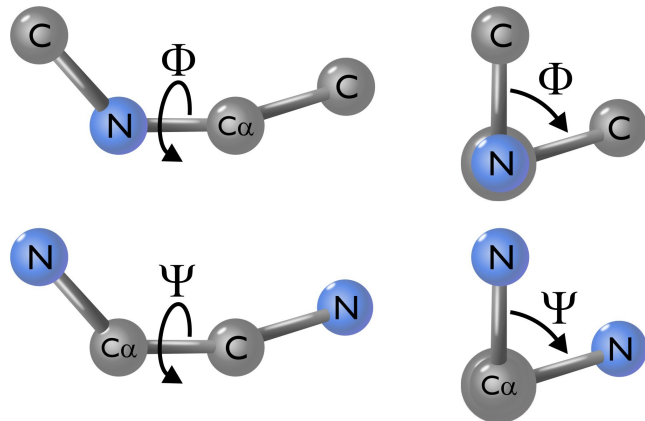
Lehetséges paraméterek

- Kötéshosszak, kötésszögek megfelelő volta
- Ne legyenek atom-atom ütközések
- Eltemetett amidcsoportok többsége H-kötésben legyen
- Főlánc konformációs tulajdonságai alapján

A fehérje főlánc konformációja



Torziós szög
óramutató járásával
megegyező (+)
ellentétes (-)



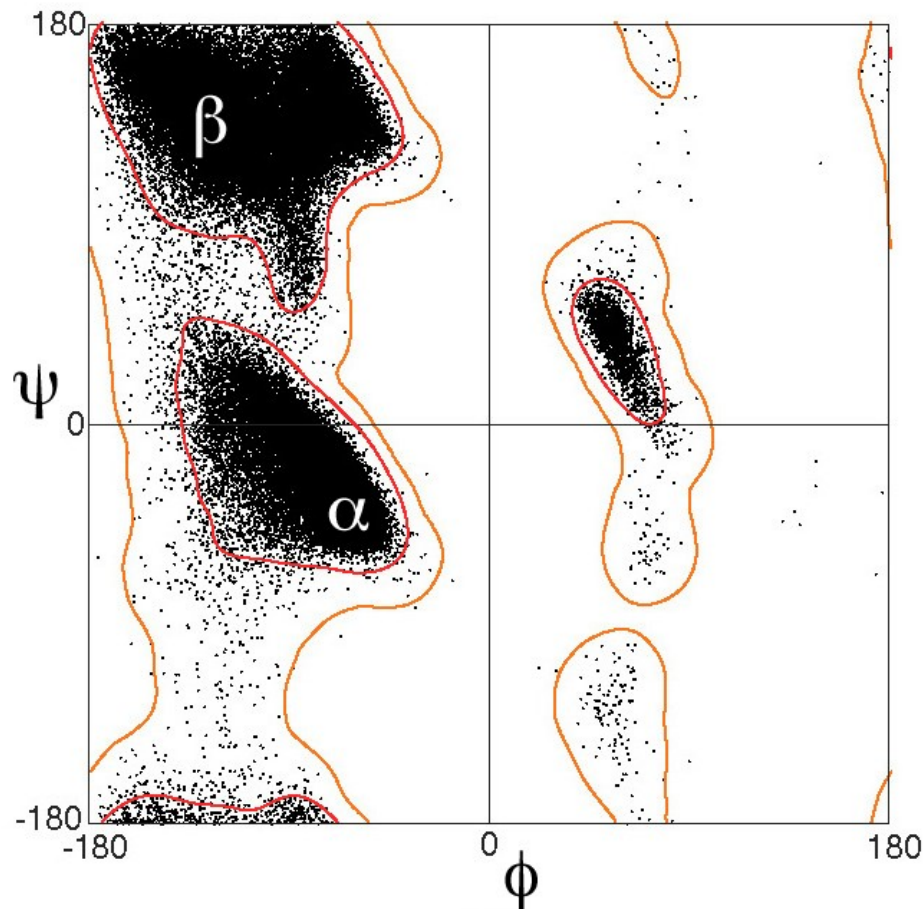
N-C α körüli szög: phi

C α -C' körüli szög: psi

A fehérjékben az aminosavak molekulagerincének ϕ és ψ torziós szögei nem vehetnek fel akármilyen értékeket, vannak preferált konformációk.

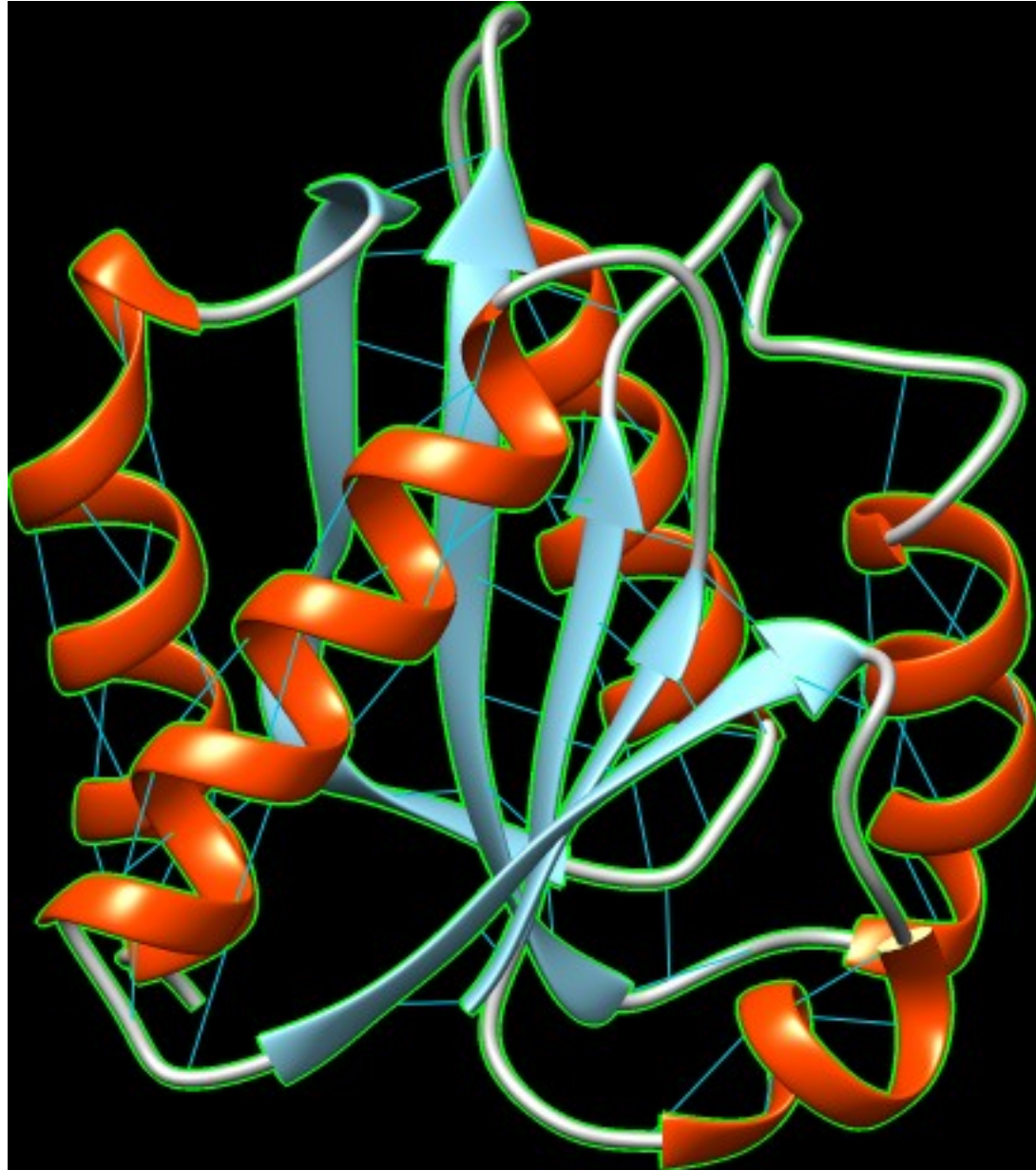
Ramachandran plot

A két torziós szöget X és Y-koordinátaként használva a kapott koordinátarendszerben ábrázolhatjuk az egyes aminosavaknak megfelelő szögpárokat.



Általában a glicin és prolin aminosavak nem szerepelnek a térképen, ezek konformációs preferenciái ugyanis egyediek. Mint az várható, az α -hélixeknek és a β -szálaknak megfelelő régiókba esik a legtöbb aminosav, de a hurkokban és kanyarokban lévők többsége is a megengedett tartományokon belül van.

A másodlagos szerkezeti elemeket
hidrogén hidak stabilizálják



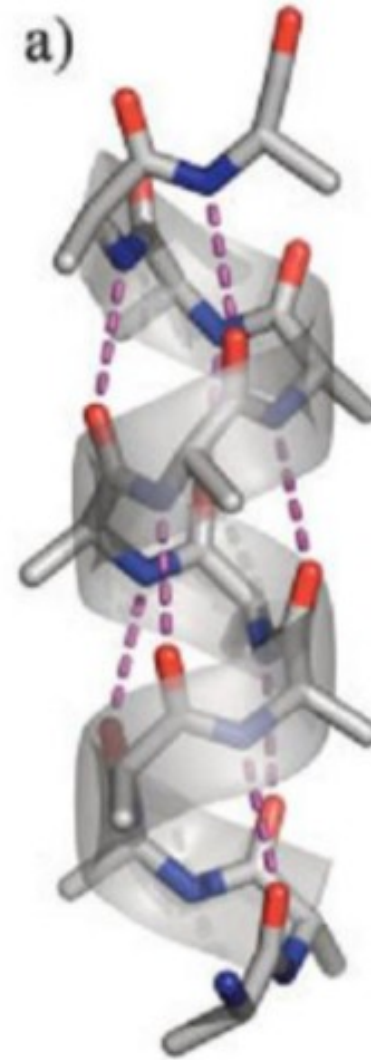
Az α hélix (jobb kezes)

A globuláris fehérjék kb 30%-a

5-40 aminosav hosszúak
(átlagosan 10)

Bár az egyes H-kötések
önmagukban viszonylag gyengék,
a hélix stabilitásához jelentősen
hozzájárulnak.

Egy adott peptid α -hélix képző
hajlama a szekvenciától függ



α helix

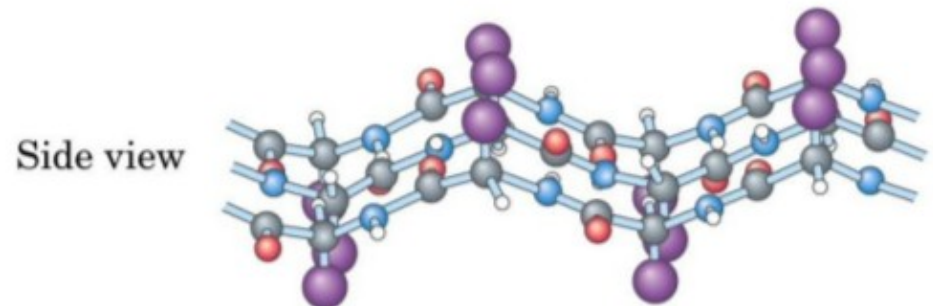
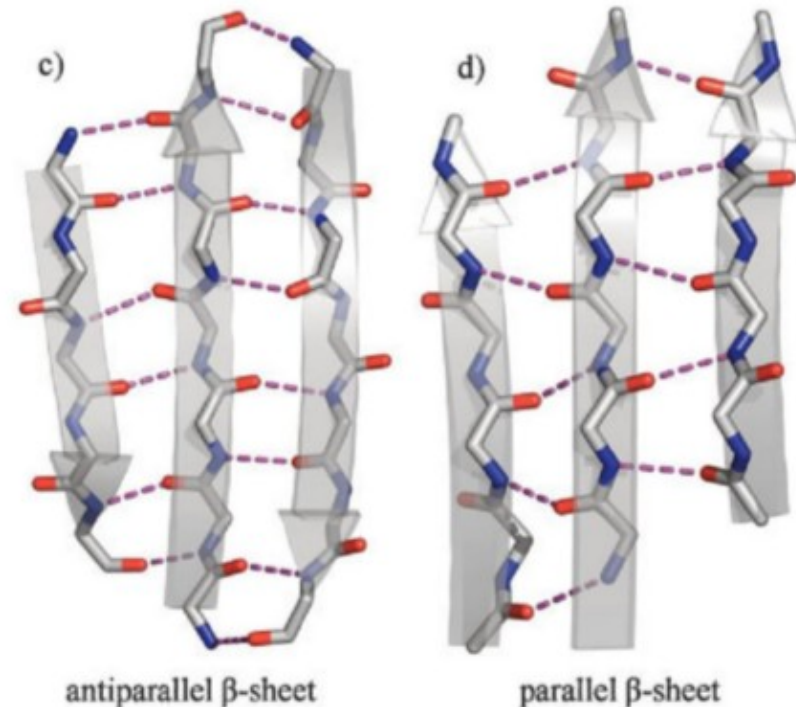
A β lemez konformáció

A globuláris fehérjék kb 30%-a

A szálak egymás mellett helyezkednek el, 5-10 aminosavból állnak

A szálak H-kötéssel kapcsolódnak egymásnak.

Az egyes aminosavak eltérő β lemez-képző hajlammal rendelkeznek



Loop-ok és kanyarok

Általában hidrofíl jelleggel rendelkeznek. A fehérje külső részén fordulnak elő, a vízzel és más molekulákkal alakítanak ki H-kötéseket

Gyakran alakítanak ki kötőhelyeket, illetve aktív helyeket enzimekben illetve receptorokban.

Az egyes aminosavak eltérő loop-képző hajlammal rendelkeznek



Másodlagos szerkezet meghatározása

Alapulhat:

Hidrogén kötés mintázat

Dihedral szögek

Automatikus meghatározás algoritmusokkal

DSSP

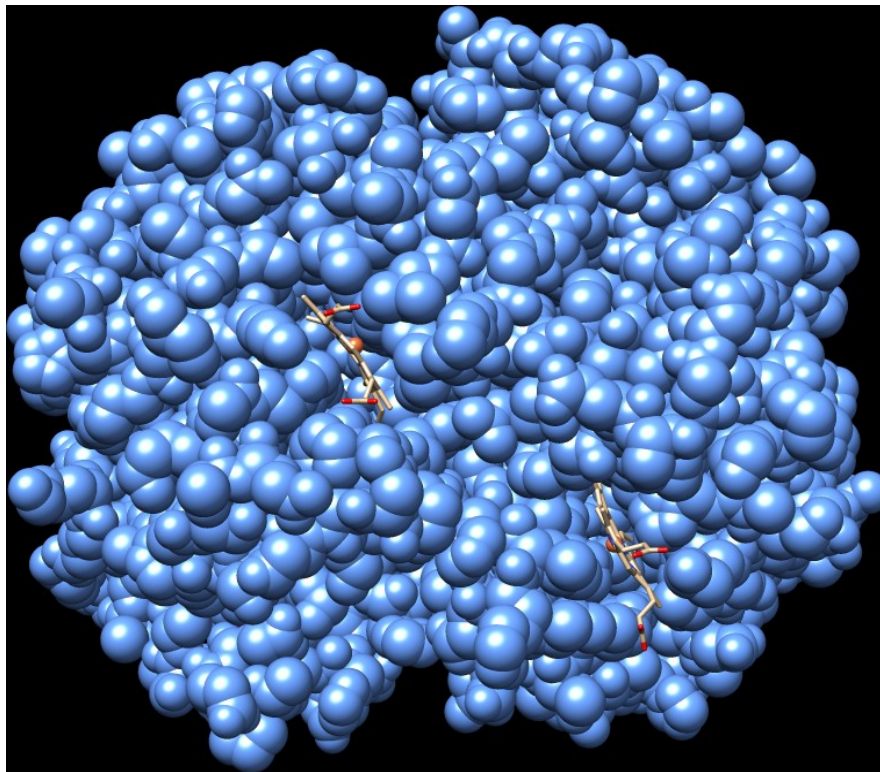
STRIDE

3 (alpha, beta, coil)

vagy több kategória (pl. turn, más típusú hélixek)

Nem egyeznek meg 100%-ban

A fehérje belseje szorosan pakolt



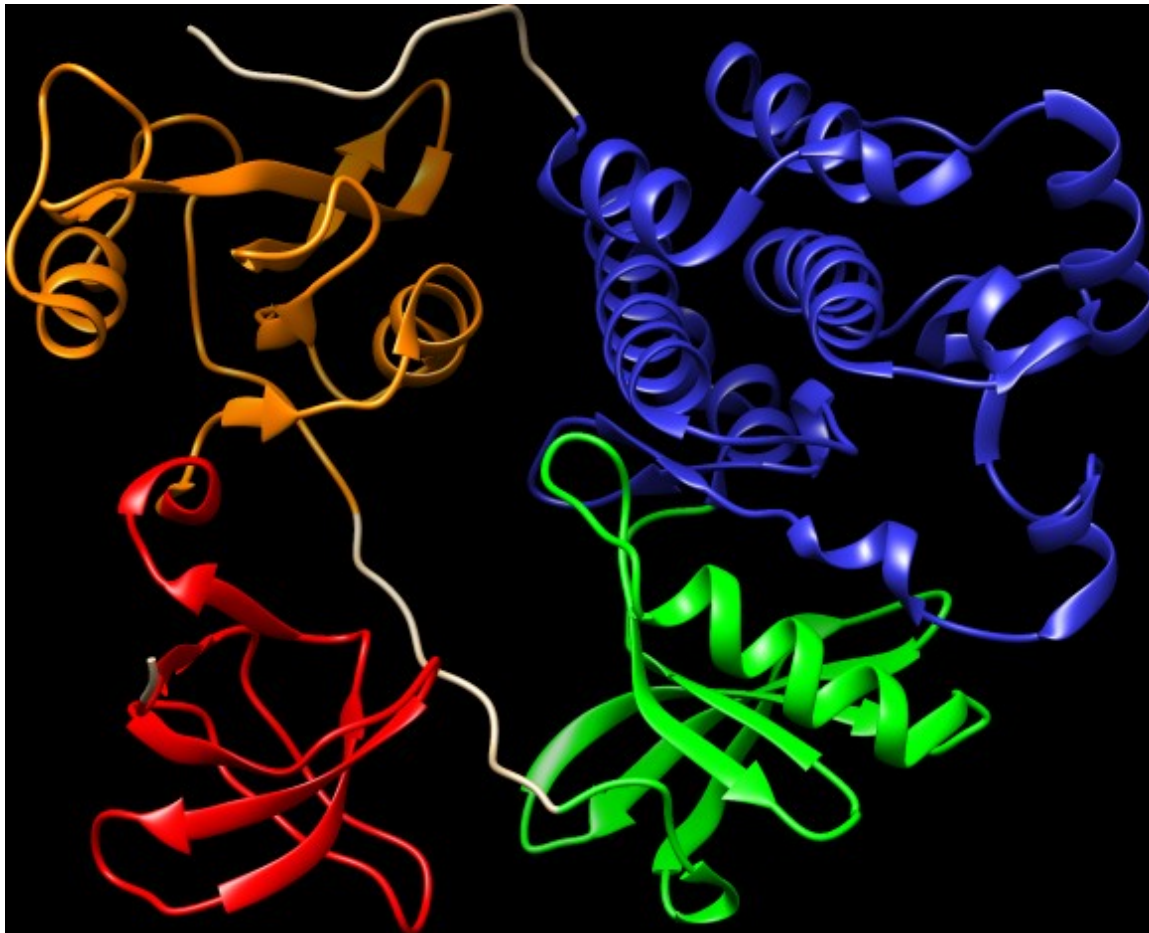
Az aminosavak elhelyezkedése

Az aminosav oldalláncok polaritásuk szerint helyezkednek el

- Az apoláros aminosavak a fehérje belsejében (Val, Leu, Ile, Met, & Phe)
- A töltött aminosavak (Arg, His, Lys, Asp, & Glu) a fehérje felszínén vannak
- A nem-töltött poláros aminosavak (Ser, Thr, Asn, Gln, és Tyr) általában a felszínen vannak, de lehetnek a fehérje belsejében is, ebben az esetben H-kötésekben vesznek részt

Domének

Számos fehérjében elkülöníthetők viszonylag kompakt szerkezeti egységek



Domének

Kompakt, önmagukban is globuláris jellegű szerkezeti egységek.

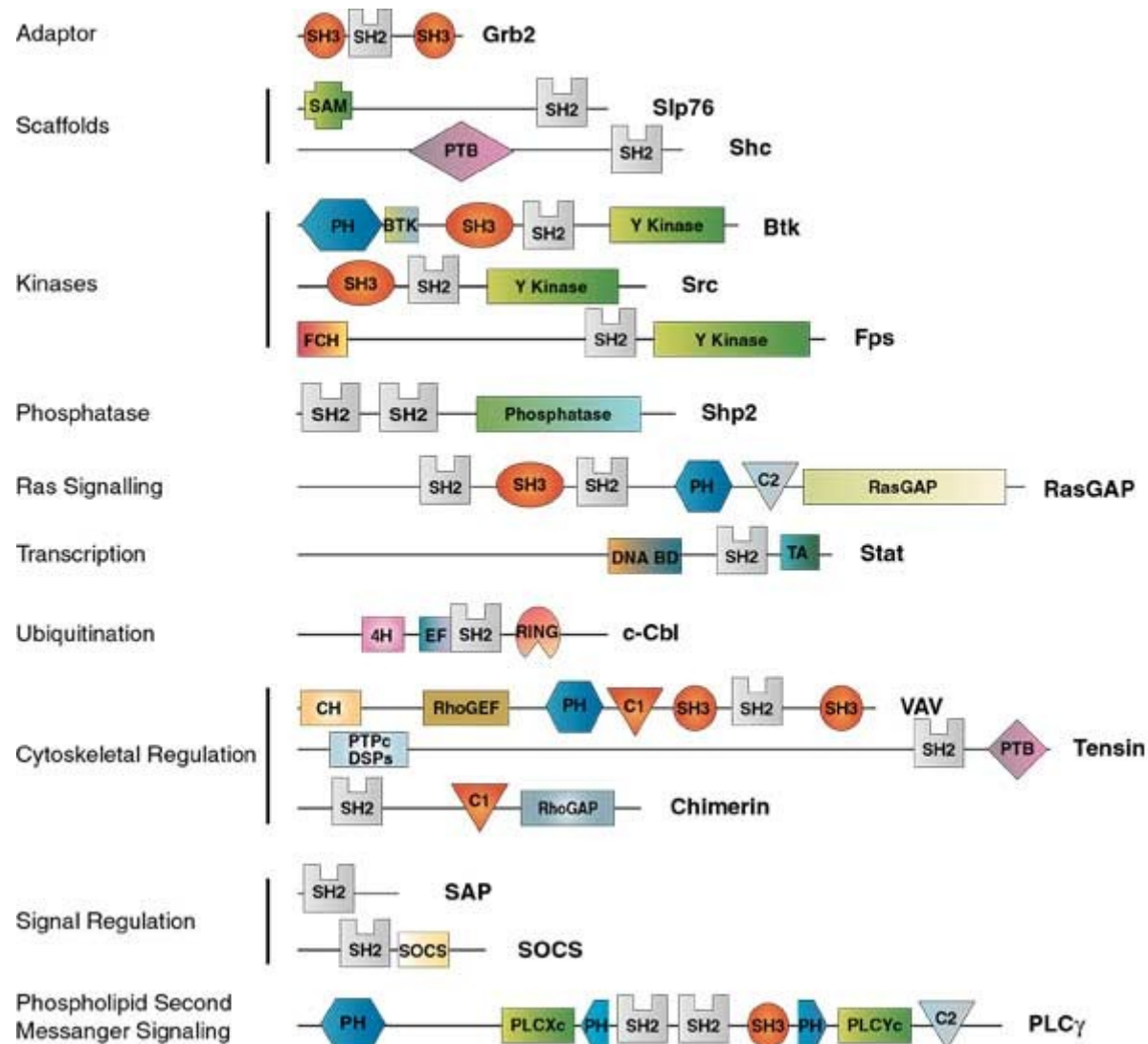
A domének a fehérjék építőkövei

Általában egy adott funkcióért felelősek

Különböző biológiai kontextusban is előfordulhatnak

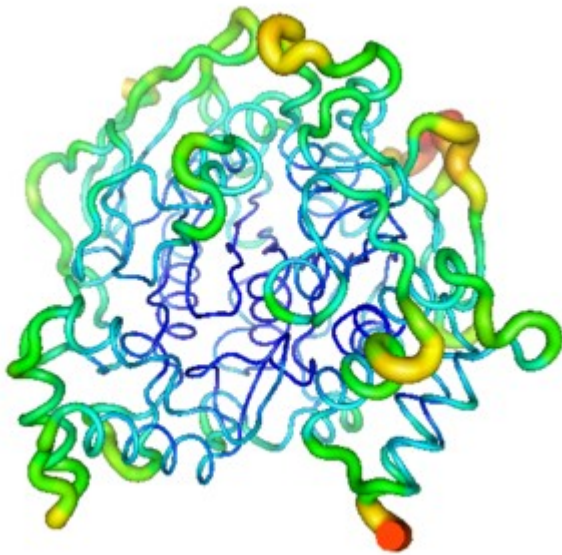


SH2 domén

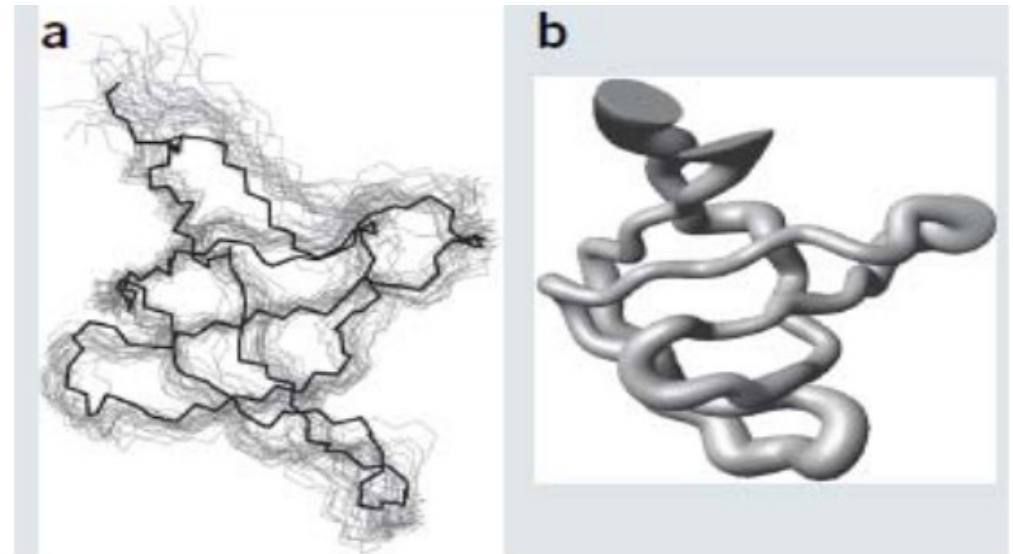


A fehérjék dinamikus molekulák

X-ray
B-faktor



NMR
Szerkezeti variabilitás



Hiányzó szerkezetek



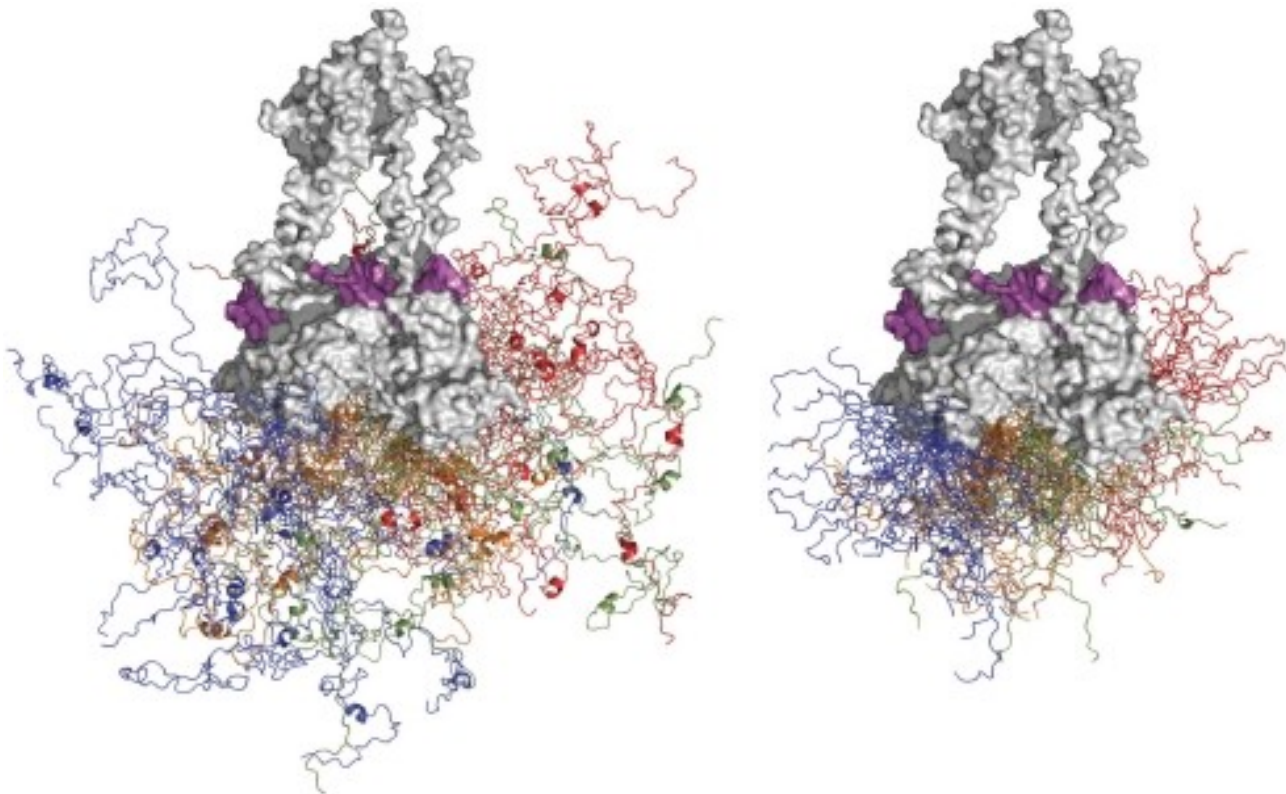
Hiányzó szegmensek a fehérje szerkezetekben



NMR structures nagy szerkezeti variabilitással

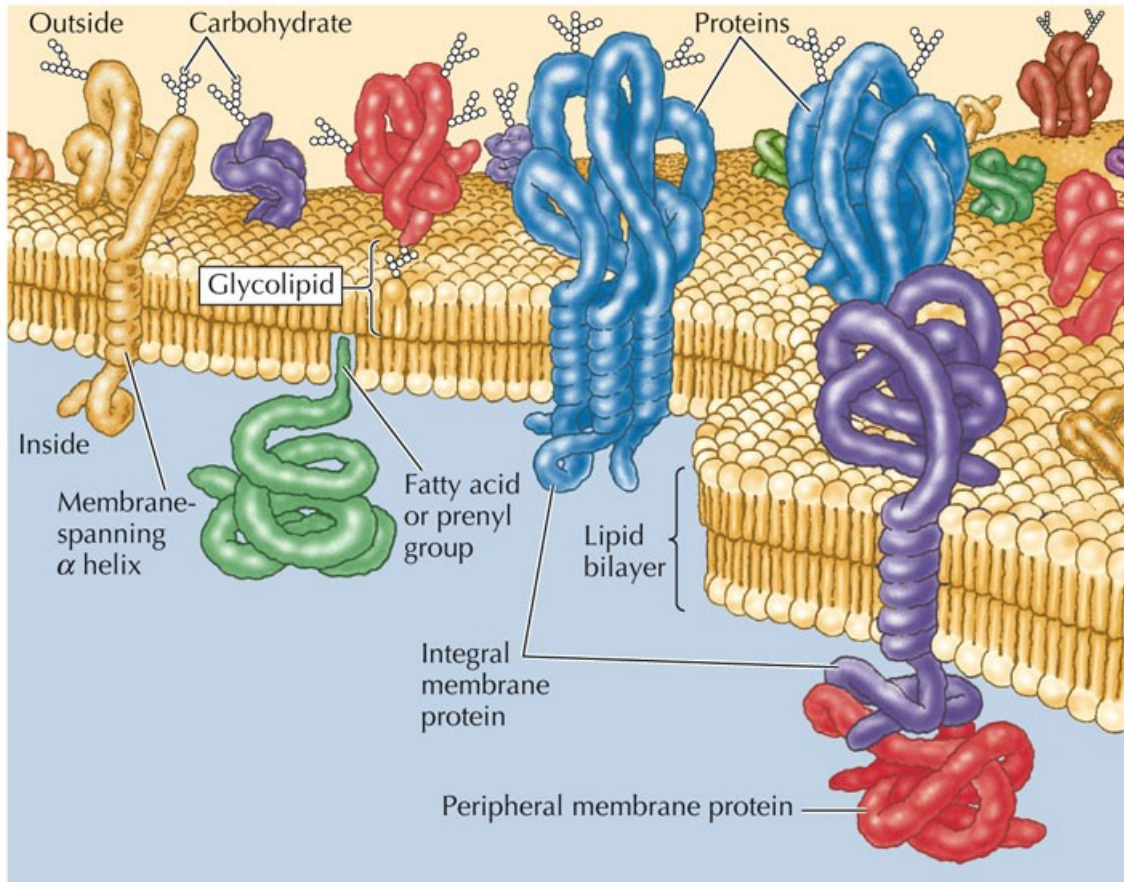
Rendezetlen fehérjék

Önmagukban nem alakítanak ki jól definiált térszerkezetet natív(szerű) körülmények között



Humán p53

Membrán fehérjék



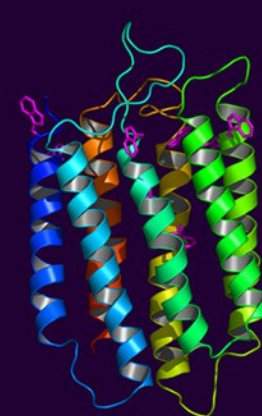
THE CELL, Fourth Edition, Figure 2.25 © 2006 ASM Press and Sinauer Associates, Inc.

Fontosak:

Energia termelés
Transzport
Sejt-sejt kapcsolat
Jelátvitel

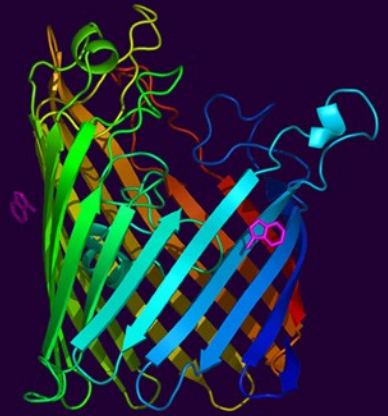
Gyógyszercélpont

The known structures of transmembrane proteins belong to two classes, based on their transmembrane secondary structure.



α -helical Bundles

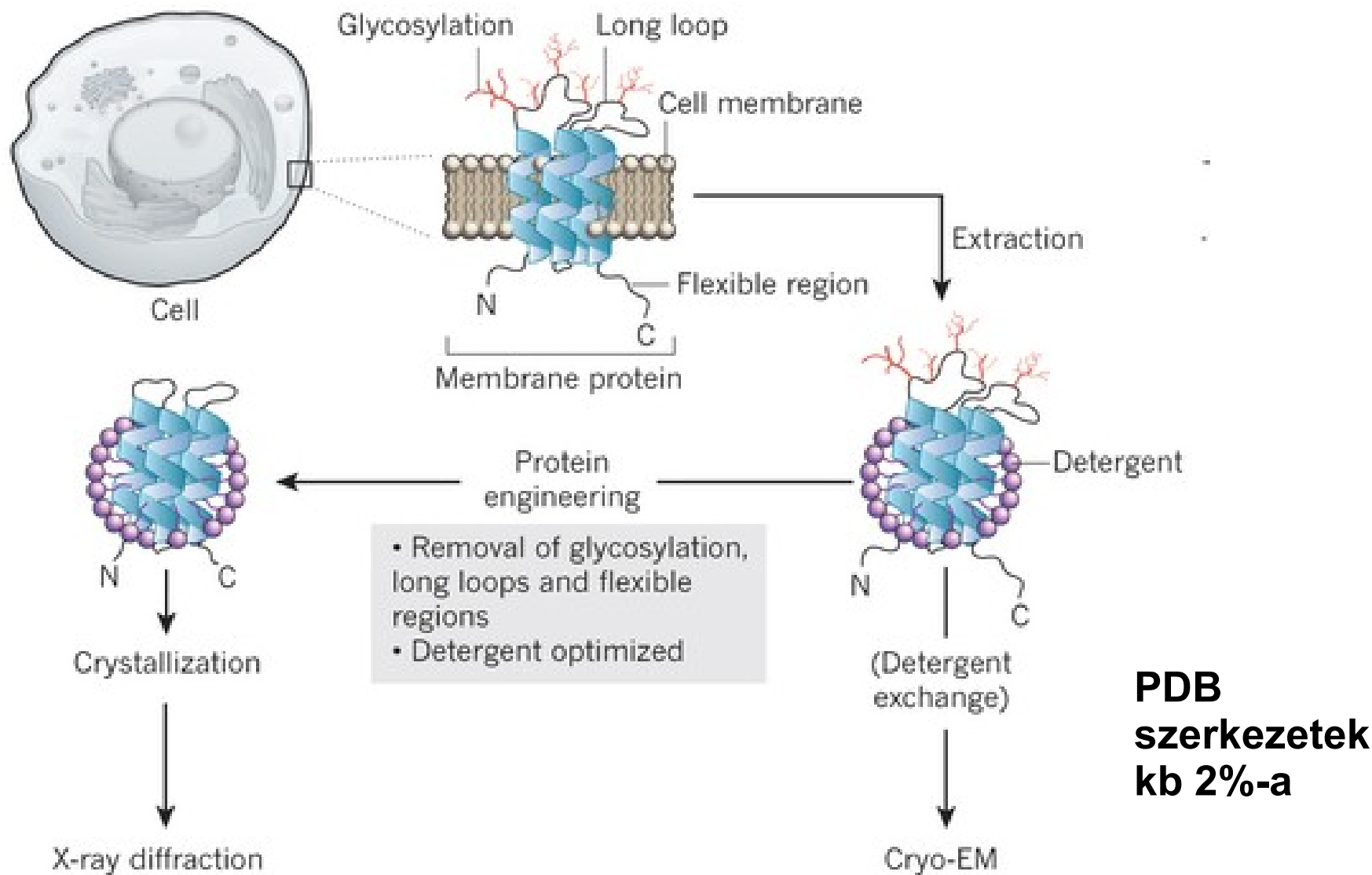
Example Bacteriorhodopsin (PDB 1AP9)



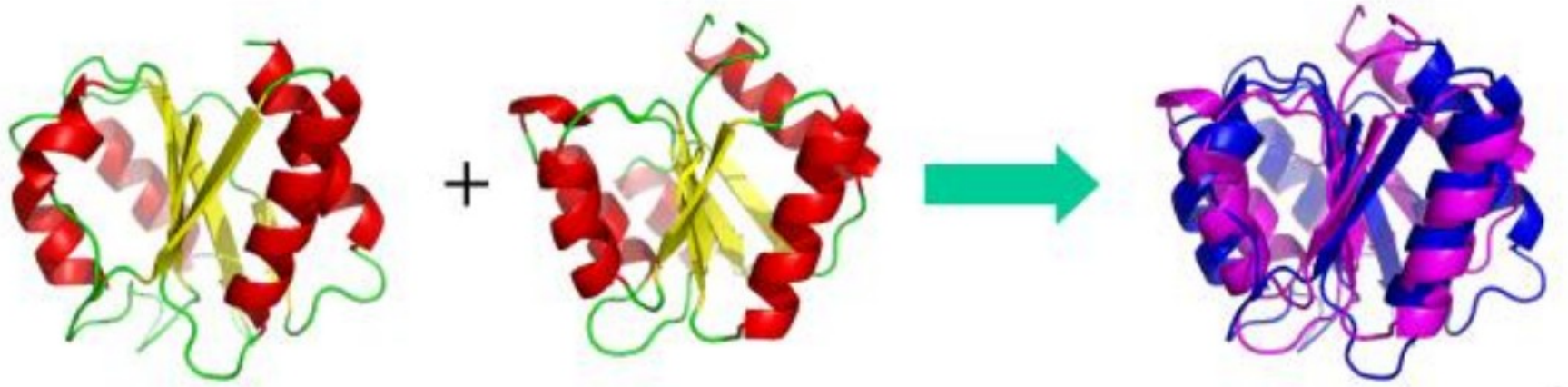
β -Barrels

Example: Matrix Porin (PDB 1OMF, Subunit)

TM fehérjék szerkezet meghatározása

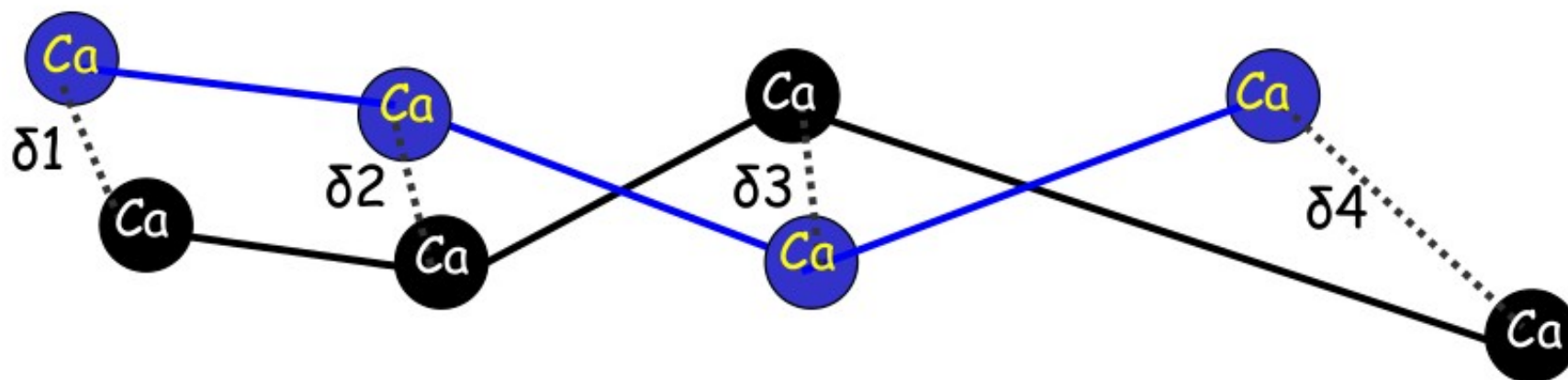


Mennyire hasonló két szerkezet



Szuperpozíció: pozíciók közötti távolság minimalizálása

RMSD



Root Mean Square Deviation (RMSD):

A legáltalánosabban használt függvény a szerkezeti hasonlóság mérésére

RMSD az átlagos távolság szuperimponált szerkezetek ekvivalens atomjai között

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

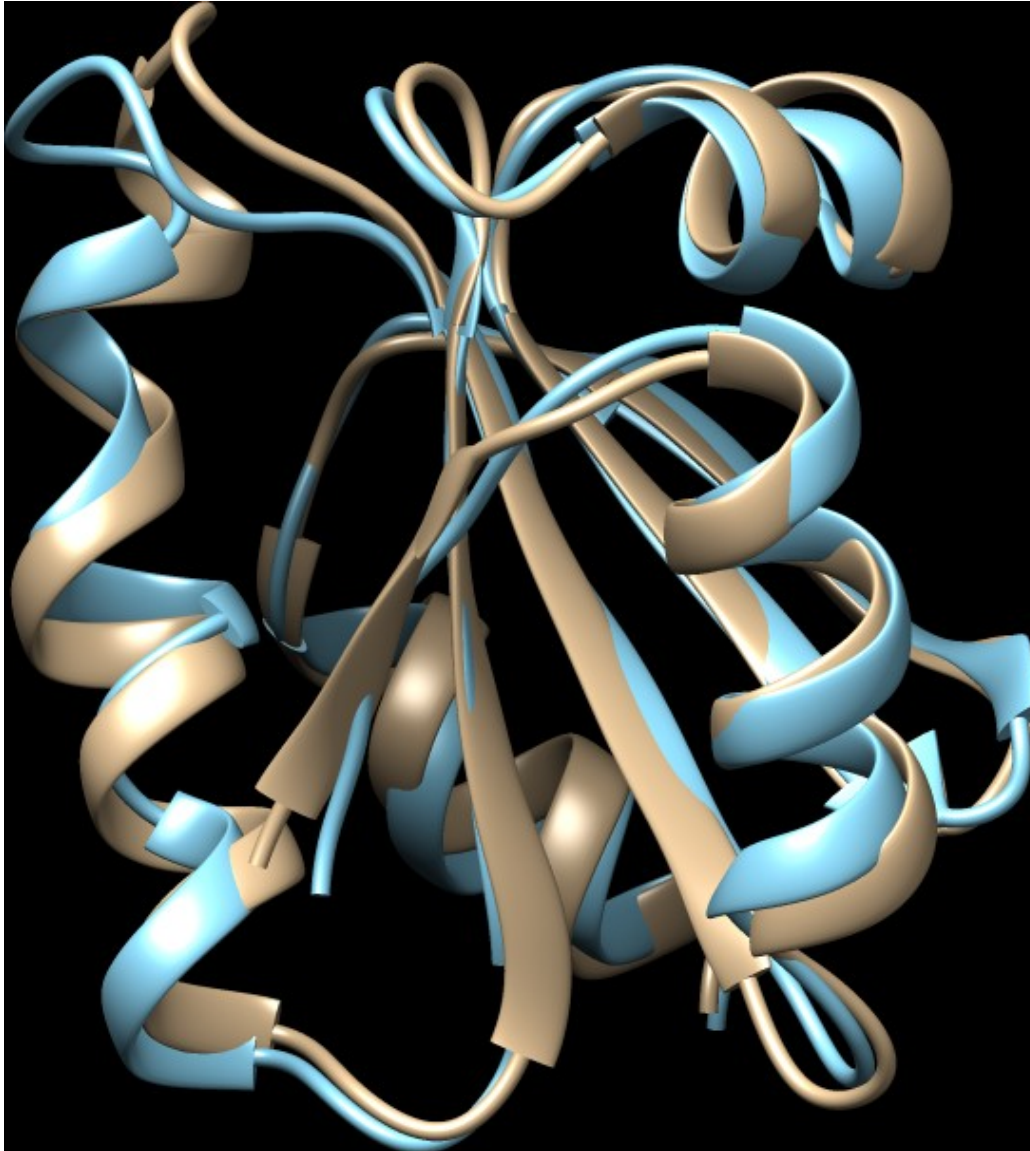


Kristály kontaktusok



RMSD: 0.955A

NMR vs X-ray

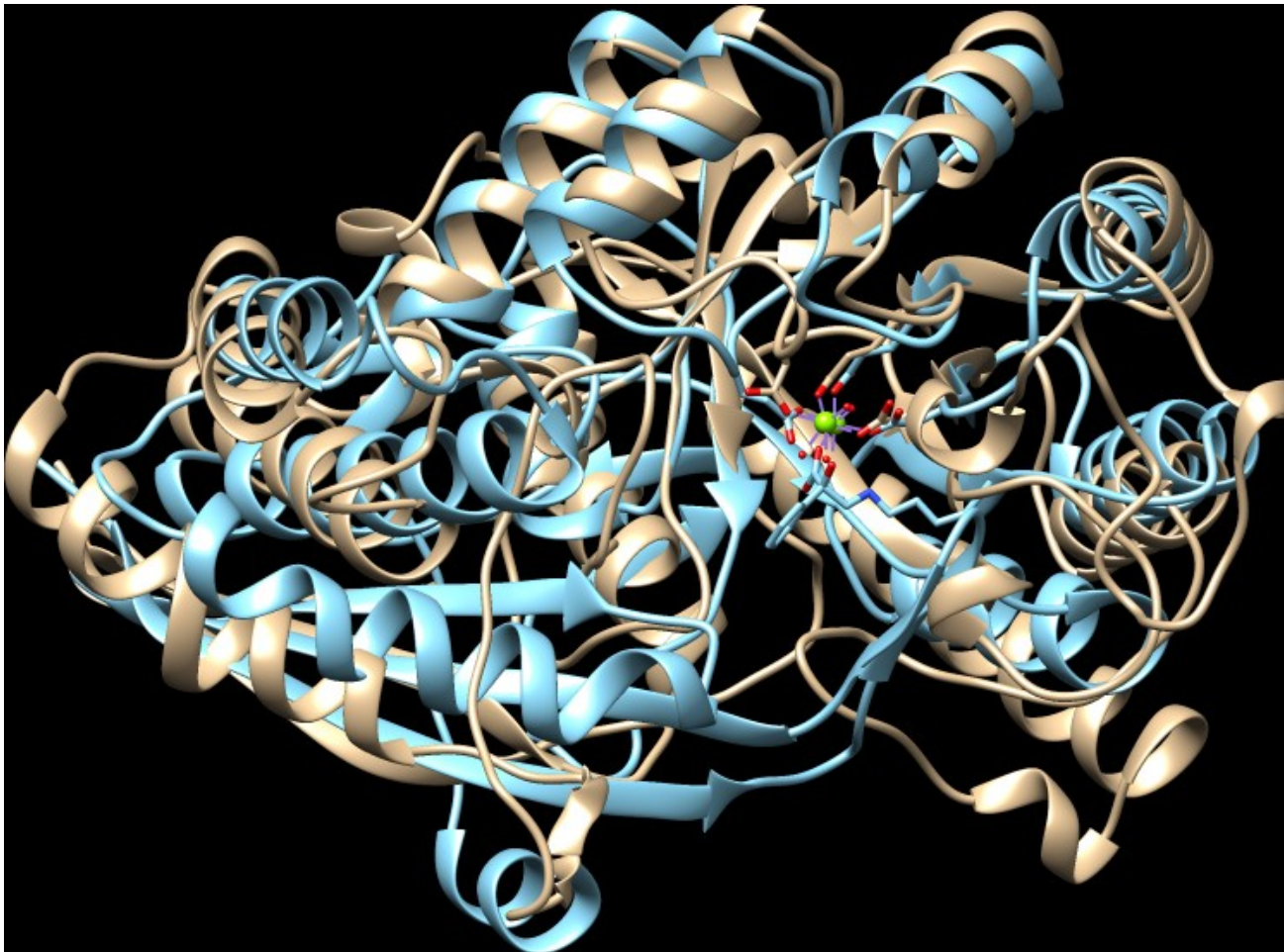


THIOREDOXIN

X-ray: 1ert
NMR : 3trx

RMSD 0.877A

Evolúciósan rokon fehérjék szerkezete általában hasonló



1ebhA: enolase

1mns : mandelate racemase

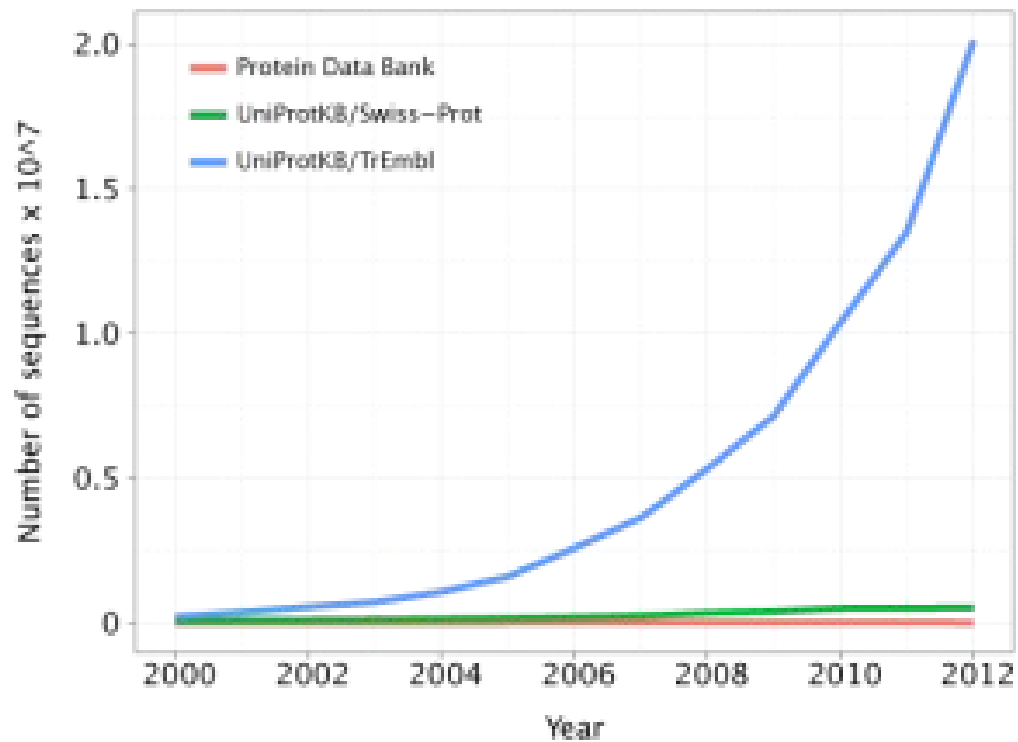
Szekvencia azonosság: 25%

Aktív centrum nagyon hasonló

Hasonló kémiai reakció

Eltérő szubsztát

Szekvencia szakadék

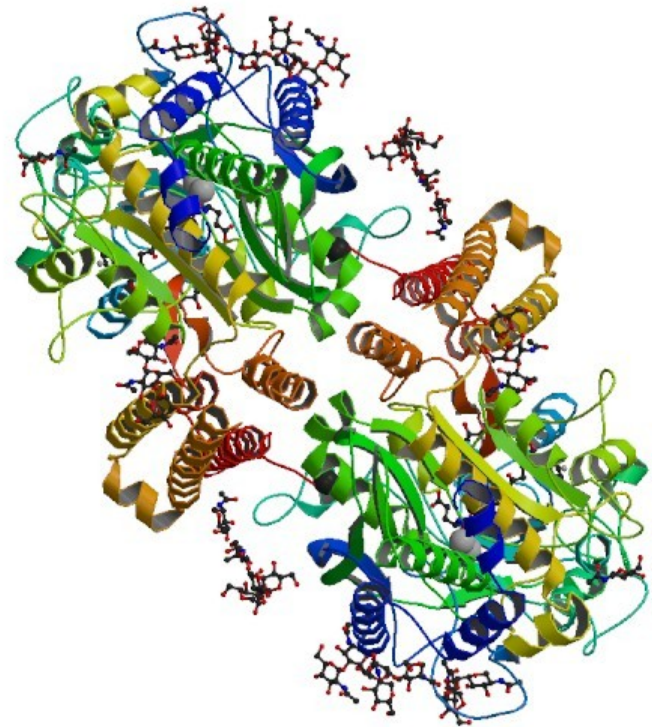


	2014
Szekvenciák	50 000 000
Szerkezetek	100 000

Térszerkezeti predikciók

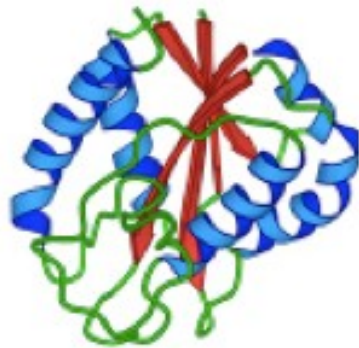
>Protein

```
RSKSSNEATNITPKHNMKAFLDELKAENIKKFLYNFTQIPHLAGTEQNFQLAKQIQSQWKEFGLDSVELAHYDVLLSYPN  
KTHPNYISIIINEDGNEIFNTSLFEPPPPGYENVSDIVPPPSAFSPQGMPEGLVYVNYARTEDFFKLERDMKINCSGKIV  
IARYGKVFRGNKVKNALAGAKGVILYSDPADYFAPGVKSYPDGWNLPGGGVQRGNILNLNGAGDPLTPGYPANEYAYRR  
GIAEAVGLPSIPVHPIGYYDAQKLLKMGGSAPPDSSWRGSLKVPYNVGPFTGNFSTQKVKMHIHSTNEVTRIYNVIGT  
LRGAVEPDRYVILGGHRDSWVFGGIDPQSGAAVVHEIVRSFGTLKKEGWRPRRTILFASWDAEEFGLLGSTEWAEENSRL  
LQERGVAYINADSSIEGNYTLRVDCPLMYSLVHNLTKELKSPDEGFEGKSLYESWTKKSPSPSEFSGMPRISKLGSGNDF  
EVFFQRLGIASGRARYTKNWETNKFSGYPLYHSVYETYELVEKFYDPMFKYHLTVAQVRGGMVFELANSIVLPFD CRDYA  
VVLRYADKIYISIMKHPQEMKTYSVSFDLSLSAVKNFTEIASKFSERLQDFDKSNPIVLRMMNDQLMFLERAFIDPLGL  
PDRPFYRHVIYAPSSH NKYAGESFPGIYDALFDIESKVDPSKAWGEVKRQIYVAAFTVQAAAETLSEVA
```



Fehérje folding

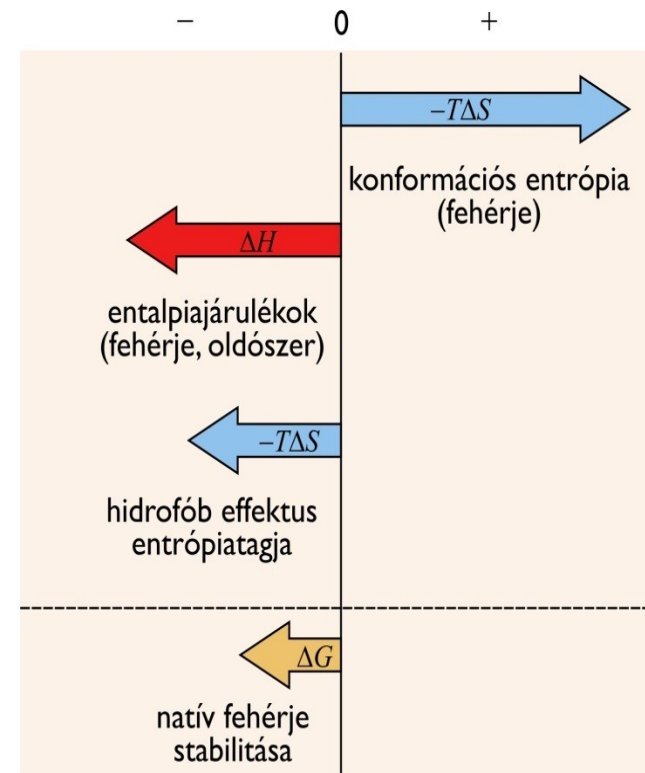
GFCHIKAYTRLIMVG...



Folding

(physics)

$$\Delta G = \Delta H - T\Delta S$$



Térszerkezet meghatározása fizikai alapon

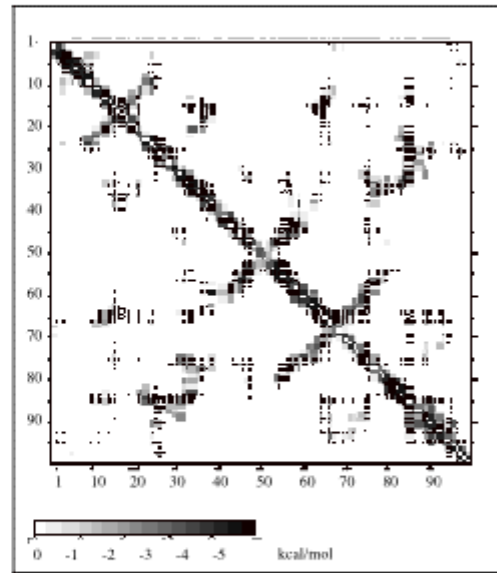
- nagyon sok konformáció, hatalmas konformációs tér
- fizikai energiafüggvény nem ismert pontosan

Predikciók típusai

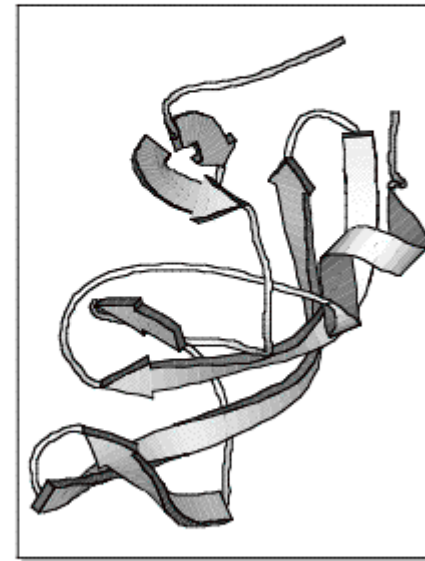
Notation: protein structure 1D, 2D, 3D

P	PP	P	128	110
Q	QQQY		175	97
I	FFQVI		70	E 60
T	SSIVR		77	E 69
L	LLSTL		120	E 14
W	WNQED		238	E 81
Q	RKQAK		169	E 97
R	RRRPQ		200	62
P	PPPPP		24	48
L	VVTKE	E	71	E 59
V	VVLII	E	14	E 0
T	TTKEK	E	74	E 69
I	AALIV	E	0	E 0
K	HYKKF	E	90	E 73
I	IILVI		4	E 0
G	EENG		46	41
G	GGGTG		62	53
Q	QKRR		68	71
L	PPLWW	E	118	E 59
K	VVFKV	E	31	E 73
E	EESKK	E	124	E 95
A	VVGLG	E	1	E 0
L	LLILL	E	29	E 0
L	LLLVV	E	24	E 0
D	DDDDD		49	E 58
T	TTTTT		72	51
G	GGGGG		62	30
A	AAAAA		17	0
D	DDDDD		102	79
D	DDAKE		69	58
T	SSTTV		1	69
V	IIVIV	E	14	E 0
L	VVIVL	E	0	E 0

1D

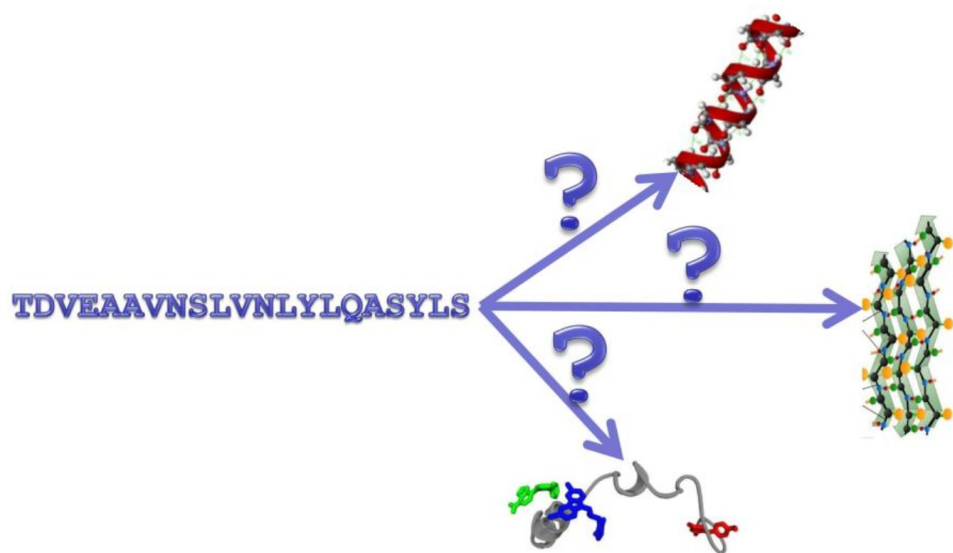


2D



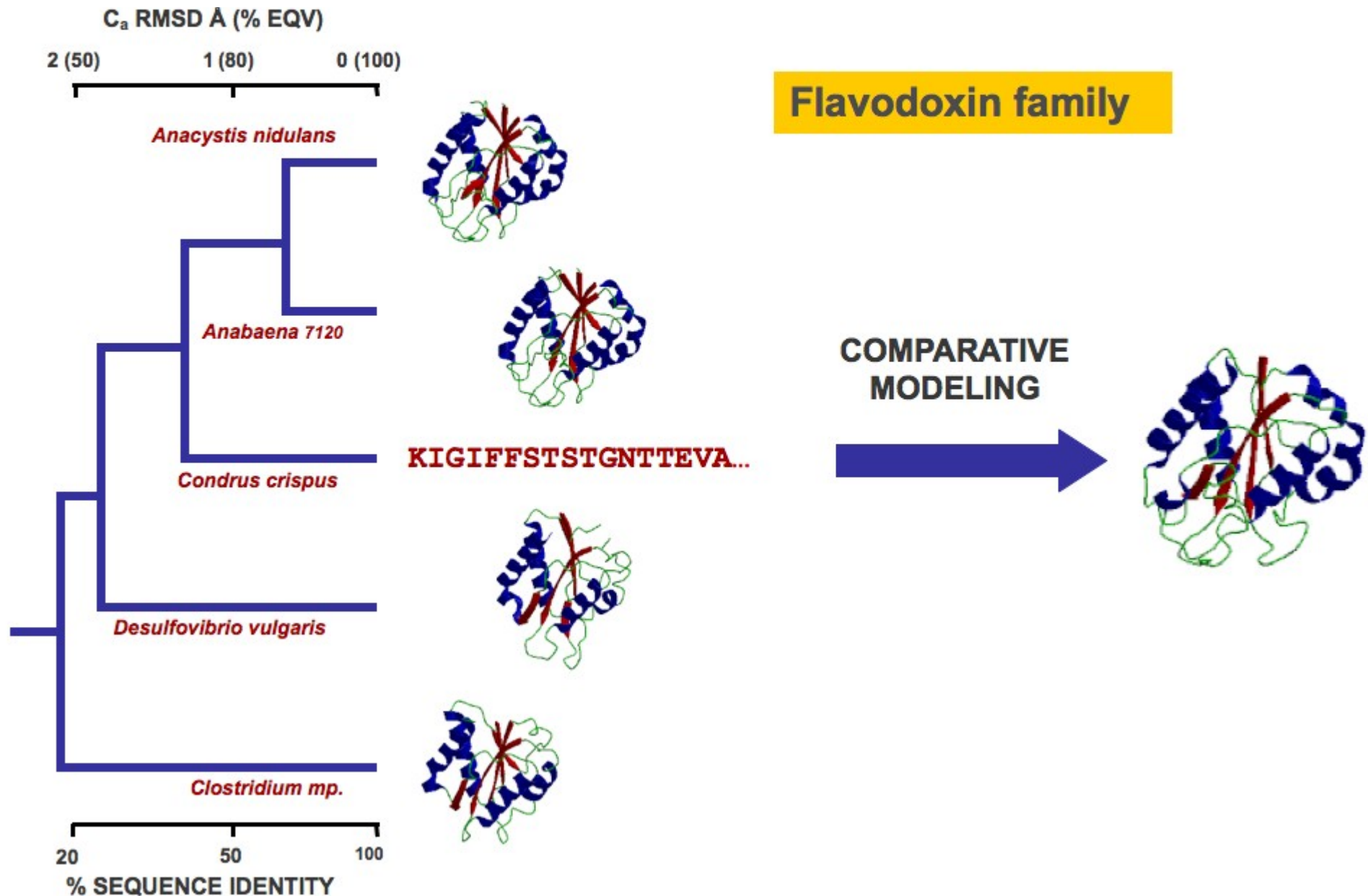
3D

Másodlagos szerkezet predikciója



	α -hélix	β -lánc szerkezetek	β -kanyarok
Glu			
Met			
Ala			
Leu			
Lys			
Phe			
Gln			
Trp			
Ile			
Val			
Asp			
His			
Arg			
Thr			
Ser			
Cys			
Asn			
Tyr			
Pro			
Gly			

Összehasonlító szerkezet modellezés





The Nobel Prize in Chemistry 2013

Martin Karplus, Michael Levitt, Arieh Warshel

The Nobel Prize in Chemistry 2013



© Harvard University

Martin Karplus



Photo: © S. Fisch

Michael Levitt



Photo: Wikimedia Commons

Arieh Warshel

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"*.

I.L.:

's sort of nice in more general terms to see that computational science, computational biology is being recognized." He added, "It's become a very large field and it's always in some ways been the poor sister, or the ugly sister, to experimental biology."