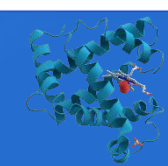


6.05.2026

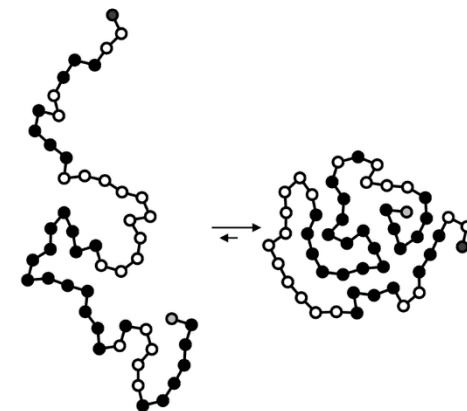
**Curs XI – Modelarea structurii
tridimensionale complete a unei proteine
pe baza secvenței de aminoacizi**

Scopul și tipuri de modelare a structurii 3D a proteinelor



Plierea proteinelor este un proces fizic,

Ținând cont de **principiul lui Anfinsen** (structura primară a catenei polipeptice, prin tipul și succesiunea aminoacizilor, este cea care conține toată informația necesară plieri proteinei în forma nativă) postulează că împachetarea lanțului de aminoacizi ai unei proteine în structura tridimensională caracteristică este un proces guvernat strict de reguli fizice ceea ce ar permite, cel puțin teoretic, stabilirea structurii tridimensionale a unei proteine pornind doar de la secvența sa de aminoacizi. Ideea de a calcula, pornind de la secvența de aminoacizi, structura tridimensională a unei proteine a fost considerată încă din momentul lansării sale în anii '70 un adevărat Sfânt Graal al biologiei moleculare.



DAR depinde de:

solvent (apă sau lipide)

pH

concentrația de săruri

temperatură

Chaperoni moleculari (GroEL, HSP)

Nivele suplimentare de complexitate

predicția structurii 3D

“holy grail of molecular biology”

“the second half of the genetic code”

Research News

Trying to Crack the Second Half of the Genetic Code

Inspired by practical problems in biotechnology and medicine, researchers are attempting to figure out the rules that govern protein folding

Researcher Scherer and his associates at Lilly Research Laboratories in Indianapolis tried to produce bovine growth hormone by inserting the gene into bacteria and getting the bacteria to synthesize huge quantities of the protein. But instead of getting the nice soluble protein they wanted, they got a clumpy mess. It is a problem that other biotechnology firms. The computer want to make known proteins and they want to make new hybrid proteins, such as monoclonal antibodies hooked to toxins to destroy cancer cells. Yet, says Lewis Kazan of the University of California in San Francisco, “the outcomes are not always as they expected.”

The reason these firms are having such difficulty is that molecular biologists have not yet cracked the second half of the genetic code: what are the rules that determine how a linear amino acid sequence will fold into a protein? The first half of the code—decoding how a sequence of DNA bases is translated into a sequence of peptides—was fully straightened and was reported more than two decades ago. The second half has been unsolved for so long, says Jonathan King of the Massachusetts Institute of Technology, “that for a long time people thought that it was a problem.” Now it is becoming hard to ignore.

Not only biotechnology firms but also molecular biologists are pushing to learn protein-folding rules. The difficult nature of DNA sequencing methods are far ahead of the study of protein structure. Molecular biologists are quickly getting the sequence of thousands of genes. Now they would like to know what the genes code for. For a typical protein, this means deciphering why some sequences of a polypeptide chain fold into an alpha-helical conformation, other regions form a beta-sheet, silk-like conformations, and still other sequences form turns and loops. And, in addition, it means understanding how all these structures that make up a protein pack together.

Finally, there is an increasing realization by some physicians who study genetic diseases that some inherited disorders may be

caused by defective dynamics of protein folding. The more that is known about protein-folding rules, the better these diseases will be understood.

The disease connection arises in studies of collagen disorders. Collagen, King notes, is the one protein for which the folding rules are relatively clear. The protein is a rigid rod made of three extended strands that are twisted together. The amino acid sequence consists of repeating units of three. At every first position there is a glycine, at the third position there is a proline or hydroxyproline 20% of the time, and in the second position, says King, “a great deal of variation can be

caused by defective dynamics of protein folding. The more that is known about protein-folding rules, the better these diseases will be understood.

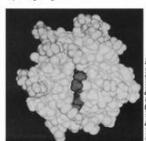
When collagen is made, the chains line up and zip together. Anything that interrupts this dynamic zipping of the chains disrupts the collagen structure. The results can be

caused by defective dynamics of protein folding. The more that is known about protein-folding rules, the better these diseases will be understood.

The disease connection arises in studies of collagen disorders. Collagen, King notes, is the one protein for which the folding rules are relatively clear. The protein is a rigid rod made of three extended strands that are twisted together. The amino acid sequence consists of repeating units of three. At every first position there is a glycine, at the third position there is a proline or hydroxyproline 20% of the time, and in the second position, says King, “a great deal of variation can be

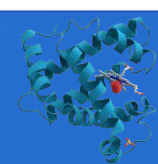
caused by defective dynamics of protein folding. The more that is known about protein-folding rules, the better these diseases will be understood.

When collagen is made, the chains line up and zip together. Anything that interrupts this dynamic zipping of the chains disrupts the collagen structure. The results can be



Protein folding patterns are hard to predict. The sequence of amino acids determines the protein's three-dimensional shape, but not for protein on an amino acid sequence is a folded protein on an amino acid.

Scopul și tipuri de modelare a structurii 3D a proteinelor



Modelarea structurii tridimensionale complete – are ca scop predicția poziției în spațiu a tuturor atomilor unei molecule proteice plecând de la secvența de aminoacizi cu o acuratețe similară cu metodele experimentale de determinare a structurii 3D.

Studiul principiilor care guvernează împachetarea proteinelor în forma lor nativă se poate realiza folosind două abordări distincte:

1. **strict pe baza legilor fizicii cuantice** – principiul lui Anfinsen este primordial, **catena polipeptidică se pliază în structura 3D pentru a atinge un nivel superior de stabilitate** (energie minimă, număr maxim de interacțiuni stabilizatoare);
2. **prin prisma teoriei evoluției** – observațiile practice au demonstrat că **structurile 3D sunt mai înalt conservate în evoluție comparativ cu secvența primară**. Această observație explică de ce:
 - a) proteinele cu secvențe primare similare se pliază în structuri 3D identice;
 - b) proteine cu secvențe primare diferite dar înrudite se pliază în structuri 3D asemănătoare.

Folding as a purely physics process
laws of physics

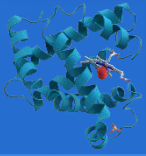
ab-initio methods

Similar sequences will have similar folds
theory and laws of protein evolution

homology modeling

protein threading

Scopul și tipuri de modelare a structurii 3D a proteinelor



Funcție de abordarea utilizată predominant, metodele de modelare a structurii 3D a proteinelor se clasifică în 3 clase mari:

A. Modelarea pe bază de omologie (homology modelling) sau **modelarea comparativă (MC, comparative modelling)** – metode ce au la bază compararea secvenței de interes cu proteine înrudite din punct de vedere evolutiv a căror structură este deja cunoscută;

B. Modelarea 'ab-initio' – metode ce calculează *de-novo* (de la 0) structura 3D a proteinei de interes pe baza interacțiunilor posibile dintre atomi, fără a ține cont de relațiile evolutive cu alte secvențe sau structuri;

C. Modelarea bazată pe **recunoașterea conformațiilor terțiare** (fold-recognition/threading) – folosesc modelarea comparativă pentru a identifica structurile secundare ale unor secțiuni mici dintr-o secvență proteică și apoi metode *ab-initio* pentru a amplasa secțiunile în spațiu pe baza interacțiunilor posibile dintre acestea.

Indiferent de tipul de metodă, structura calculată folosind metode *in-silico* pentru o secvență de interes se numește **model**. **Gradul de suprapunere a structurii model cu structura tridimensională reală a secvenței de interes** (determinată experimental prin una din metodele descrise în cursul 4) indică **acuratețea** metodei utilizate.

Arch. Biol. Sci., Belgrade, 62 (4), 857-871, 2010

DOI:10.2298/ABS1004857M

BASIC PROTEIN STRUCTURE PREDICTION FOR THE BIOLOGIST: A REVIEW

M. MIHĂȘAN

"Alexandru Ioan Cuza" University, Faculty of Biology, Department of Molecular and Experimental Biology,
6600, Iași, Romania

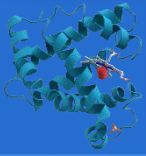
Abstract - As the field of protein structure prediction continues to expand at an exponential rate, the bench-biologist might feel overwhelmed by the sheer range of available applications. This review presents the three main approaches in computational structure prediction from a non-bioinformatician's point of view and makes a selection of tools and servers freely available. These tools are evaluated from several aspects, such as number of citations, ease of usage and quality of the results. Finally, the applications of models generated by computational structure prediction are discussed.

Keywords: Protein structure prediction, protein mode application

UDC 57.08.088.6

Nici una dintre cele trei clase de metode nu s-a impus pe plan științific în detrimentul alteia, fiecare având avantaje și dezavantaje specifice. Experimentatorul trebuie să aleagă între o metodă sau alta în funcție de secvența de interes și de rezultatul preconizat.

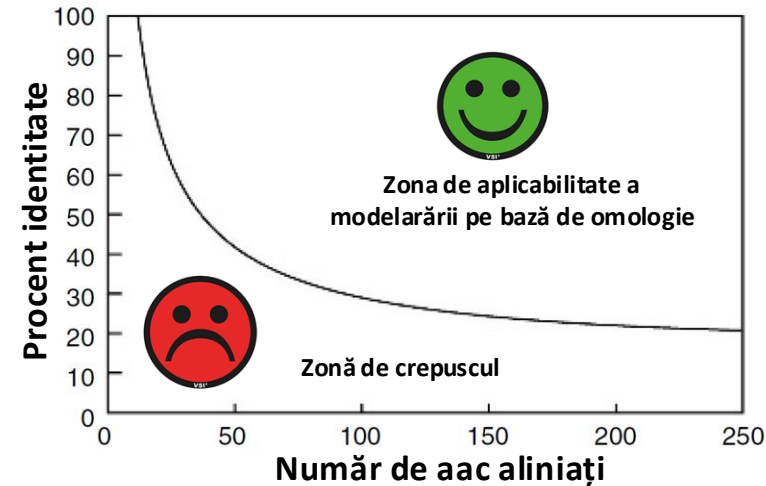
Modelarea pe bază de omologie a structurii 3D



- folosește programul BLAST pentru a identifica într-o bază de date cu structuri acele proteine ce sunt omoloage cu secvența de interes. Structurile omoloage identificate sunt folosite ca **șablon (template)** pentru a se calcula structura tridimensională a proteinei de interes și a se genera **un model** al acesteia.

Parametrul principal de care depinde acuratețea metodelor ce folosesc homologia este **procentul de identitate** (numărul de aminoacizi identici dintre cele 2 secvențe raportat la numărul total de aminoacizi) dintre secvența model și secvențele șablon. Calitatea modelului rezultat și implicit gradul de aplicabilitate a acestor metode este dependent strict de **procentul de identitate** dintre secvența șablon și cea de interes (model) astfel:

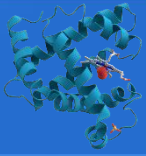
- **procentul de identitate mai mic de 25%** - identificarea de structuri omoloage este dificilă. Chiar dacă se identifică posibile structuri șablon și se pot genera modele ale structurii tridimensionale a proteinei țintă, utilitatea lor este îndoielnică, de unde și denumirea de "**zona de crepuscul**" a alinierilor secvențelor proteice;
- **procentul de identitate între 25% - 50%** - se identifică structuri omoloage, însă utilitatea lor pentru obținerea modelului lor este redusă, procesul de modelare necesită intervenție umană;
- **procentul de identitate între 55% - 75%** - structurile omoloage pot fi utilizate pentru generarea modelului; modelul poate fi generat



automat cu o acuratețe medie; în general atomii catenei peptidice sunt amplasați corect, atomii catenelor laterale nu;

- procentul de identitate **este mai mare de 75%** - mai mult de 90% dintre atomii proteinei de interes pot fi modelați automat, fără intervenție umană, cu o eroare medie de aproximativ 1 Å. Modelele realizate sunt de calitate foarte bună, concurând în acuratețe cu structurile determinate experimental prin cristalografie cu raze X;

Etapele modelării pe bază de omologie



Modelarea structurii unei proteine țintă necunoscute folosind **MC** cuprinde **șapte etape distincte**:

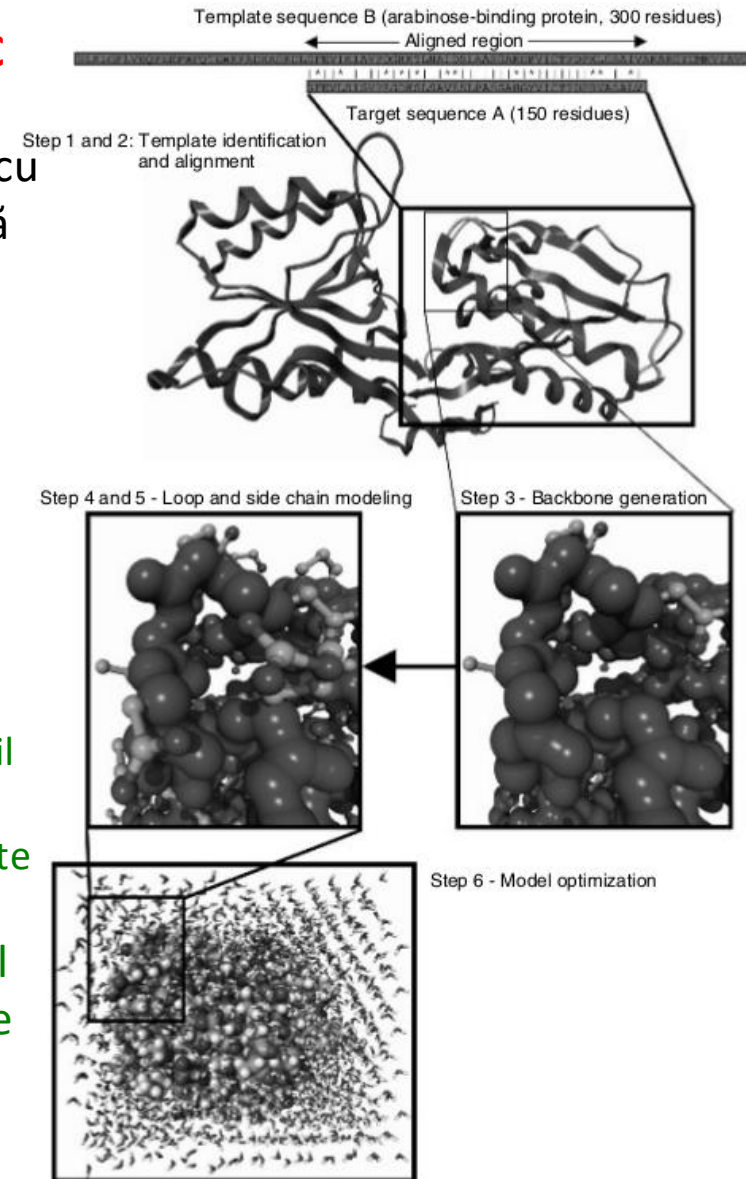
1. identificarea proteinelor înrudite din punct de vedere evolutiv cu proteina de interes investigată și care au structura tridimensională determinată experimental ce vor fi folosite ca **șablon** pentru calcularea structurii proteinei de interes;

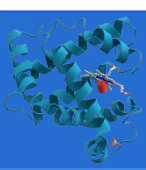
-cel mai frecvent această etapă este realizată prin folosirea programului **BLAST** și **PDB** ca bază de date;

-proteina cu nivelul **procentul de identitate cel mai ridicat** este în mod automat folosită ca **șablon**, însă trebuie ținut cont și de detalii experimentale legate de structura determinată experimental precum:

- conformația specifică a proteinei – **enzimă activă vs inactivă**;
- prezența cofactorilor sau a altor liganzi – **dacă se cunoaște că proteina țintă conține un anumit cofactor/metal, se va alege ca șablon pe cât posibil acea structură ce conține cofactorul/metalul respectiv**;
- structura cuaternară, dacă există – **dacă se cunoaște că proteina țintă este tetrameră, se va alege ca șablon pe cât posibil o structură tetrameră**;

Ca regulă generală se alege ca șablon structura care respectă cel mai bine condițiile experimentale dorite și datele experimentale existente despre proteina de țintă.





2. alinierea pe bază de omologie și poziție a secvenței proteinei țintă cu secvența proteinei șablon în scopul identificării **zonelor cu structură înalt conservată**;

-mai multe structuri șablon pot fi de asemenea alinate pentru a corela mai bine zonele cu secvențe înalt conservate cu structurile tridimensionale specifice ale acestor zone – se **obțin matrici de substituție dependente de poziția aminoacidului** într-o structură dată;

-se încearcă îmbunătățirea alinierii secvență țintă-șablon obținută în etapa anterioară prin utilizarea informațiilor structurale din structura șablon sub forma unor matrici de substituție dependente de poziție; în final se obține o **alinie structurală**;

Ex: Substituția Glu(Q)-Ala(A) – conform matricei BLOSUM32 folosită de BLAST, această substituție primește punctajul -1. Într-o structură 3D, această substituție apare frecvent și este normală pentru resturile de la suprafața moleculei și este extrem de puțin probabilă dacă A este în centrul hidrofob. Punctajul alocat pentru această substituție trebuie așadar să țină cont de poziția sa în structura 3D.

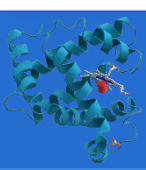
3. modelarea zonelor cu structură înalt conservată și generarea structurii 3D a scheletului polipeptidic pe baza structurii șablon;

-constă în copierea propriu-zisă a coordonatelor carteziene ale atomilor ce alcătuiesc scheletul peptidic (N, C α , C și O) pentru resturile de aminoacizi din alinierea obținută anterior de la structura șablon;

-în cazul aminoacizilor identici, se copie inclusiv coordonatele atomilor din catena laterală;

-în cazul aminoacizilor diferiți din aliniere, se pot copia și coordonatele C β ;

Unele programe (**Robetta**, **TASSER**) pot combina în această etapă mai multe structuri șablon pentru a genera modelul – din fiecare structură șablon se copie acea porțiune ce are scorul de aliniere cel mai ridicat.



4. modelarea poziției spațiale a **zonelor de buclă**;

În majoritatea cazurilor, alinierea dintre secvența țintă și cea șablon conține **GAPS** – zone din șablon ce nu există în secvența țintă (**deleții**), respectiv zone ce există în secvența țintă dar nu există în șablon (**insertii**); în cazul **delețiilor**, coordonatele carteziene ale aminoaciziilor lipsă **nu sunt transferați** către model; în cazul **insertiilor**, aminoacizi **sunt introduși în model** și coordonatele lor carteziene **trebuie calculate**. Cel mai frecvent aceste modificări nu apar în structurile secundare predominante – α -helix și β -pliat, ci în **zonele de buclă**.

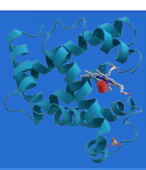
În CM conformația zonelor de buclă este calculată prin 3 abordări distincte:

a) pe baza **analogiei** cu alte zone de buclă cunoscute - se realizează o nouă căutare în PDB pentru a se identifica zonele de buclă ce încep și se termină cu aceeași aminoacizi precum bucla ce trebuie modelată; coordonatele buclei identificate sunt apoi transferate către model; majoritatea programelor de modelare prin omologie folosesc această abordare (3D-Jigsaw; Modeller; Swiss-Modeler; WHAT IF)

b) prin **identificarea conformației cu energie minimă** – se încearcă evaluarea tuturor interacțiunilor posibile dintre atomii aminoacizilor din buclă (similar cu metodele *ab-initio*)

c) Printr-o **abordare hibridă** – zona de buclă este divizată în fragmente ce sunt comparate separat cu fragmente cunoscute din PDB; fragmentele identificate ca pozitive sunt apoi asamblate folosind metode bazate pe minimizarea energiei (*ab-initio*).

Etapele modelării pe bază de omologie



5. modelarea poziției spațiale a **catenelor laterale ale aminoacizilor**;

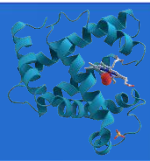
- în cazul proteinelor omoloage cu >40% identitate, conformațiile catenelor laterale sunt conservate în proporție de cel puțin 75% = > coordonatele carteziene ale atomilor din catenele laterale ale aminoacizilor identici pot fi transferate direct de la șablon către model;
- în cazul aminoacizilor ne-identici, conformația atomilor din catenele laterale este calculată pe baza unei **biblioteci de configurații comune** realizate prin înregistrarea tuturor conformațiilor posibile pentru un anumit aminoacid existente în structurile din PDB;

Poziția spațială a catenei laterale a aminoacizilor poate fi calculată cu o acuratețe de aproximativ 90% pentru aac din centrul hidrofob și de mai puțin de 50% pentru aminoacizi de la suprafața proteinei.

6. optimizarea modelului – prin parcurgerea etapelor anterioare se obține o structură tridimensională ‘grosieră’ deoarece poziția spațială a catenelor laterale a aminoacizilor influențează conformația scheletului polipeptidic și vice-versa, orice modificare a scheletului polipeptidic influențează conformația catenelor laterale; din acest motiv **etapele 3, 4 și 5 sunt repetate** până este identificată **conformația proteică cu energie minimă**;

7. validarea modelului – constă în calcularea unor indici ce descriu cât de bine modelul obținut respectă o serie de caracteristici măsurate experimental precum: lungimea legăturilor, unghiuri de torsiune, distribuția aac polari și nepolari (**care?**);

Metodele de recunoașterea a conformațiilor terțiare utilizate pentru modelarea structurii 3D a proteinelor



- au ca punct de plecare 2 observații practice:

A) secvențe de aminoacizi **similare** implică și **structuri similare**; dar

B) similaritatea la nivel de structură **NU** implică obligatoriu similaritate la nivel de secvență.

Numărul de conformații (folds) întâlnite în structura proteinelor este mult mai redus decât numărul de secvențe existente

The protein structure prediction problem could be solved using the current PDB library

Yang Zhang and Jeffrey Skolnick*

Center of Excellence in Bioinformatics, University at Buffalo, 901 Washington Street, Buffalo, NY 14203

Communicated by R. Stephen Berry, University of Chicago, Chicago, IL, September 27, 2004 (received for review November 10, 2003)

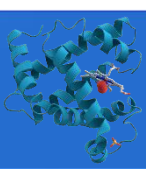
For single-domain proteins, we examine the completeness of the structures in the current Protein Data Bank (PDB) library for use in full-length model construction of unknown sequences. To address this issue, we employ a comprehensive benchmark set of 1,489

one structure (actually many) in the PDB having an rms deviation (RMSD) root to native below 4 Å with ~79% alignment coverage. This finding suggests that, at least at the level of structure alignments, the current PDB is almost a complete set of single-domain

Este astfel posibilă realizarea clasificarea conformațiilor proteice și realizarea unor biblioteci de conformații model (fold library). Metodele sunt bazate **pe identificarea porțiunilor din secvența de interes care au similarități la nivel de secvență cu un set de conformații model** și **asamblarea acestor fragmente în structura tridimensională finală ținând cont de interacțiunilor posibile dintre acestea și principiile de minimizare a energiei.**

Metodele bazate pe recunoașterea structurilor secundare sunt utilizabile atunci când nivelul de identitate dintre secvența țintă și secvențele existente în bazele de date este mai mic de **25%**.

Metodele de recunoașterea a conformațiilor terțiare utilizate pentru modelarea structurii 3D a proteinelor

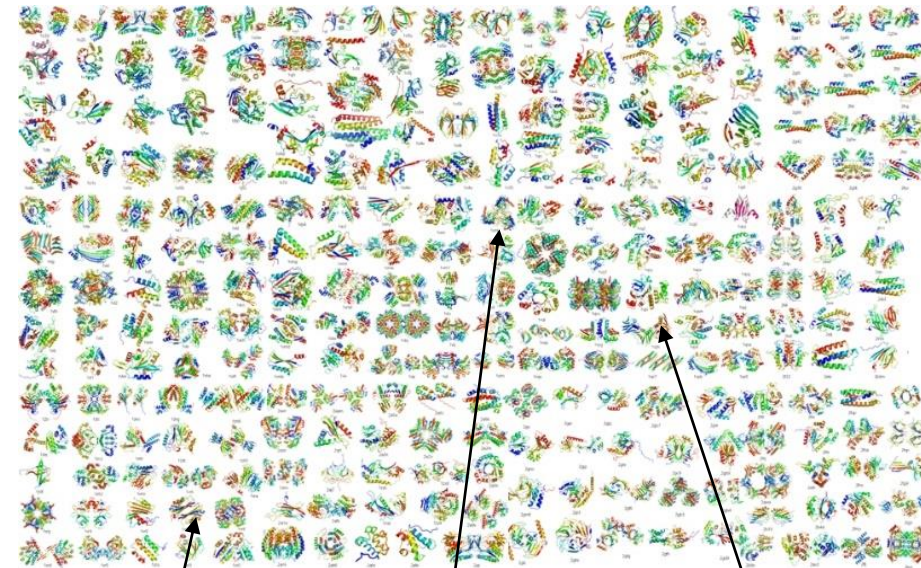


Etapele modelării unei proteine țintă sunt:

(a) Realizarea unei baze de date cu conformații model prin clasificarea și ierarhizarea pe bază de omologie a structurilor existente - într-o manieră similară cu SCOP și CATH;

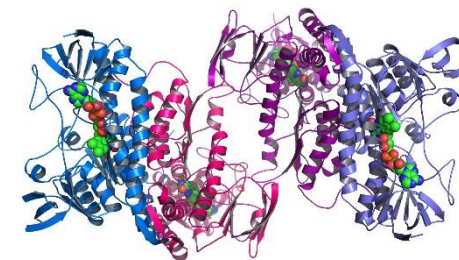
(b) Alinierea secvenței țintă cu secvențele conformațiilor terțiare model și identificarea celor existente în secvența țintă;

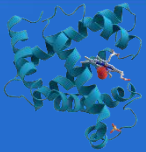
(c) Construirea modelului prin alăturarea conformațiilor terțiare identificate și calcularea interacțiunilor dintre acestea pe baza principiilor minimizării energiei.



<http://www.gnf.org/assets/001/23013.jpg>

KFINDERESYKQLTWTDTRLATGSWSLAKDFGSPAWNGKAVGGTATFWTG





Se bazează pe faptul că structura tridimensională a unei proteine corespunde unui nivel energetic minim și încearcă să identifice acest minim prin evaluarea energiei tuturor conformațiilor posibile ale respectivului lanț de aminoacizi.

- aceste metode necesită o putere de calcul enormă și nu au încă un nivel de acuratețe foarte mare, însă sunt singura alternativă pentru proteinele ce nu au omologi în bazele de date.

[Rosetta@home](#) And T0283 - doi ani și aproximativ 70 000 de PC-uri

The screenshot shows a news article from Nature. The main headline is "The shape of protein structures to come". The sub-headline reads "Modelling effort uses mass computing power to make breakthrough." The article is by Ewen Callaway. The text states: "By exploiting millions of hours of computing time donated by the users of 150,000 home computers, scientists have predicted the structure of a protein using just its sequence of amino acids. The project marks a significant advance in a field that's been flush with hope yet short on tangible results, experts say." There is a photo of David Baker at a computer. The article is dated 16 October 2007. The page also features navigation links like "nature news home", "news archive", "specials", "opinion", "features", "news blog", and "nature Journal". There are also sections for "Related stories" and "NatureJobs".



Aceste metode se aplică pentru:

- Proteinele ce nu au omologi de structură;
- Proteinele cu mai puțin de 100 aminoacizi.

Cum aleg metoda potrivită?



La dimensiunea actuala a bazei de date PDB (număr de structuri), probabilitatea de a identifica o proteină cu structură cunoscută similară cu o secvență randomică variază în intervalul **30-65% = orice secvență de nucleotide as avea, dacă codifică o proteină sunt 30-60% șanse să pot obține o structură tridimensională pentru acea proteină prin metode *in-silico*.**

Înainte de a alege o metodă sau un program trebuie să:

1. Verific nivelul de similaritate dintre proteina de modelat și structurile cunoscute:
 - dacă similaritatea este mai mare de 25% - **homology modeling**
 - dacă similaritatea este mai mică de 25% - **threading**
 - Nu există omologi - ***ab initio***

The screenshot shows the Protein Structure Prediction Center website. It features a navigation menu on the left with links like Home, EDSCASB Forum, FC Leads, etc. The main content area displays a list of proceedings from CASP10 down to CASP6, including titles like 'Proteins: Structure, Function, and Bioinformatics' and 'Table of Contents (OPEN ACCESS)'. There are also links for 'Selected papers' and 'Abstracts'.

Assessment of template-based protein structure predictions in CASP10

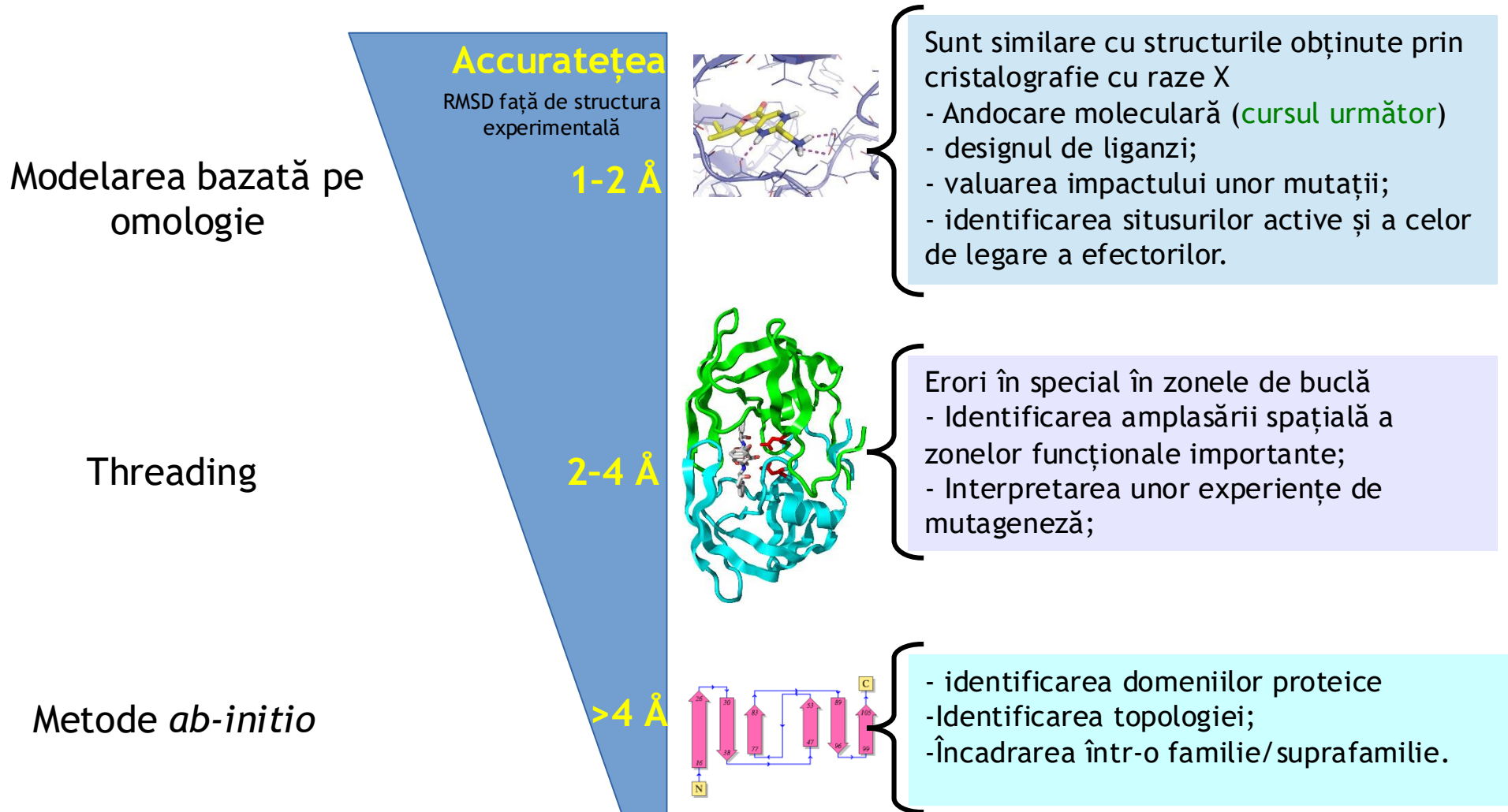
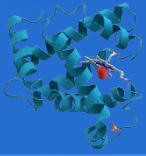
Yuanpeng J. Huang,^{1,2,3} Binchen Mao,^{1,2,3} James M. Aramini,^{1,2,3} and Gaetano T. Montelione^{1,2,3,*}

Table 1
Sum and Average Z-scores for Top 25 Performing Server Predictor Groups – 112 hAUs

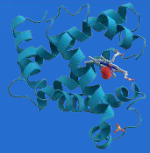
Group	Name	N_model	GDT-HA	GDC-all	RPF	LDLT	Sum	Avg-a	Avg-s	MolPro
330s	BAKER-ROSETTASERVER	112	52.03	62.49	77.60	75.18	66.83	0.60	0.60	219.08
005s	Zhang-Server	112	54.21	48.41	78.52	66.59	61.83	0.55	0.55	15.62
108s	FMS	112	37.21	49.23	72.74	74.08	58.32	0.52	0.52	2.03
114s	QUARK	111	58.68	41.91	70.40	60.60	55.90	0.50	0.50	9.40
370s	HHpred-thread	111	44.39	48.89	50.60	58.05	50.48	0.46	0.46	-144.25
122s	RaptorX-ZY	112	43.44	42.14	53.96	47.22	46.69	0.42	0.42	-70.88
430s	HHpredA	112	48.20	48.70	45.30	49.21	46.25	0.41	0.41	-137.80
223s	HHpredAQ	112	40.82	43.87	44.73	48.88	44.58	0.40	0.40	-139.75
486s	RaptorX	112	42.91	43.50	42.62	36.23	41.32	0.37	0.37	-36.43
424s	MULTICOM-NOVEL	112	31.30	34.88	45.11	47.34	39.66	0.35	0.35	27.56
125s	MULTICOM-REFINE	112	27.70	33.03	39.74	42.52	35.75	0.32	0.32	30.57
081s	MULTICOM-CLUSTER	112	24.39	30.72	39.49	41.41	34.00	0.30	0.30	30.46
335s	TASSER-VMT	112	29.13	24.93	48.74	28.75	32.89	0.29	0.29	-88.38
103s	PconsM	112	24.21	25.59	43.61	37.44	32.71	0.29	0.29	27.14
488s	chunK-TASSER	112	24.95	29.28	38.73	36.42	32.35	0.28	0.28	-26.19
292s	Pcons-net	112	19.93	20.25	32.86	28.66	23.95	0.21	0.21	66.23
286s	Mufold-MD	112	10.72	16.35	32.47	30.63	22.54	0.20	0.20	93.95
222s	MULTICOM-CONSTRUCT	112	14.40	19.42	26.07	29.11	22.25	0.20	0.20	16.48
333s	MUFOLD-Server	112	16.05	15.96	27.70	23.79	20.88	0.19	0.19	-13.69
281s	SeqS-server	112	13.31	23.50	12.71	25.40	18.73	0.17	0.17	75.22
411s	FALCON-TOPO	112	6.67	7.67	17.54	11.72	10.90	0.10	0.10	-30.77
456s	FALCON-TOPO-X	112	3.35	5.14	11.72	5.90	6.53	0.06	0.06	-36.42
124s	PconsD	111	0.89	-0.42	14.28	8.15	5.73	0.05	0.05	0.36
348s	Phyre2_A	112	7.88	5.33	2.20	6.29	5.43	0.05	0.05	-82.57
413s	ZHOU-SPARKS-X	112	-3.20	-5.76	11.41	8.48	2.73	0.02	0.02	-86.11

2. Verific CASP (Critical Assessment of protein Structure Prediction) – un forum central unde sunt evaluate și clasificate programele de predicție a structurii 3D a proteinelor
<http://predictioncenter.org/>

La ce pot folosi structurile obținute?

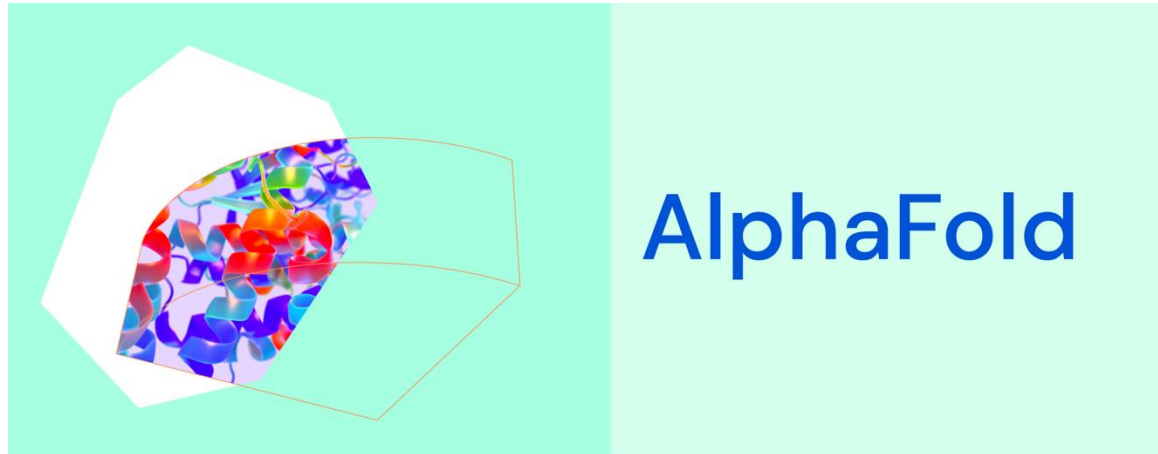
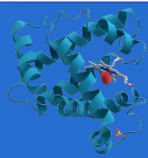


Programe și servere utile



Name	Comments	URL	Name	Comments	URL
	Stand-alone Homology modeling		HHPred (Sodingt et al., 2005)	homology detection and structure prediction by HMM-HMM comparison	http://toolkit.tuebingen.mpg.de/hhpred#
CABS (Kolinski, 2004)	<i>de novo</i> folding of small proteins, comparative modeling. Also accessible through the Selvita Protein Modeling Platform OS: Linux	www.biocomp.chem.uw.edu.pl/services.php	Swiss-Modell (Arnold et al., 2006a; Bordoli et al., 2009a; Kiefer et al., 2009; Peitsch, 1995)	fully automated protein structure homology-modeling server, accessible also from the program DeepView (Swiss Pdb-Viewer)	http://swissmodel.expasy.org/
Fold-X (Schymkowitz et al., 2005)	commercial program, on site registration possible for a 15 days trial	http://foldx.crg.es/		Threading/fold recognition	
Modeller (Fiser and Sali, 2003b; Sali et al., 1995; Sanchez and Sali, 1997)	command line interface, GUIs and web-servers also available OS: Windows, Mac, Linux	www.salilab.org/modeller/	3D-PSSM (Kelley et al., 2000)	Since 2004 the development of this server has been frozen	http://www.sbg.bio.ic.ac.uk/~3dpssm/index2.html
What If	commercial program	http://swift.cmbi.ru.nl/whatif/	Phyre (Kelley and Sternberg, 2009b)	The successor of 3D-PSSM	http://www.sbg.bio.ic.ac.uk/~phyre/
nest	part of the JACKAL software package, combines template-based methods with <i>ab initio</i> -like energy minimization principles OS:SGI 6.5, Intel Linux and Sun solaris.	http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:Jackal	I-TASSER (Zhang, 2007; Zhang, 2008a; Zhang, 2009a)	3D models are built based on multiple-threading alignments by LOMETS and iterative TASSER simulations; was ranked as the No 1 server for protein structure prediction in recent CASP7 and CASP8 experiments (the "Zhang-Server")	http://zhanglab.ccmb.med.umich.edu/I-TASSER/
Biskit (Gruenberg et al., 2007)	a python library for structural bioinformatics research OS:Linux, Windows	http://biskit.pasteur.fr/	LOOPP (Meller and Elber, 2001; Teodorescu et al., 2004; Tobi and Elber, 2000)	fold recognition program based on the collection of numerous signals, merging them into a single score, and generating atomic coordinates based on an alignment into a homolog template structure	http://cbsuapps.tc.cornell.edu/loopp.aspx
	Threading/fold recognition		Muster (Wu and Zhang, 2008)	it generate sequence-template alignments by combining sequence profile-profile alignment with multiple structural information	http://zhanglab.ccmb.med.umich.edu/MUSTER/
SUPERFAMILY (Gough et al., 2001; Madera et al., 2004; Wilson et al., 2007; Wilson et al., 2009)	Hidden Markov modeling	http://supfam.org/SUPERFAMILY/		Ab-initio	
	Servers Homology modeling		ModLoop (Fiser et al., 2000; Fiser and Sali, 2003a)	automated modeling of loops in protein structures, relies on the loop modeling routine in MODELLER	http://modbase.compbio.ucsf.edu/modloop/
3D-JIGSAW (Bates et al., 2001)	fully automated system which can be also run in interactive mode	www.bmm.icnet.uk/~3djigsaw/	Phyre (Kelley and Sternberg, 2009b)	The successor of 3D-PSSM	http://www.sbg.bio.ic.ac.uk/~phyre/
EsyPred3D (Lambert et al., 2002)	automated homology modeling, The final three dimensional structure is built using the modeling package MODELLER	http://www.fundp.ac.be/sciences/biologie/urbm/bioinfo/esypred/		Metaserver	
Geno3D (Combet et al., 2002)	comparative protein structure modeling by spatial restraints satisfaction, generates models containing up to 500 amino acids	http://geno3d-pbil.ibcp.fr	Lomets (Wu and Zhang, 2007)	generates 3D models by collecting high-scoring target-to-template alignments from 8 locally-installed threading programs	http://zhanglab.ccmb.med.umich.edu/LOMETS/
HHPred (Sodingt et al., 2005)	homology detection and structure prediction by HMM-HMM comparison	http://toolkit.tuebingen.mpg.de/hhpred#	GeneSilico (Kurowski and Bujnicki, 2003)	on-site registration required	www.genesilico.pl/meta2/
HHPred (Sodingt et al., 2005)	homology detection and structure prediction by HMM-HMM comparison	http://toolkit.tuebingen.mpg.de/hhpred#	Meta-PP (Eyrich and Rost, 2003; Rost, 1996)	the job can be submitted to up to 12 servers, among which 2 threading and two homology modeling servers	http://www.cs.bgu.ac.il/~dfischer/predictproteins/submit_meta.html
HHPred (Sodingt et al., 2005)	homology detection and structure prediction by HMM-HMM comparison	http://toolkit.tuebingen.mpg.de/hhpred#	3D-JURY (Ginalski et al., 2003)	uses about 10 different servers for a prediction, among which 3D-PSSM and	http://meta.bioinfo.pl/submit_wizard.pl
			Robetta (Chivian and Baker, 2006; Chivian et al., 2003; Kim et al., 2004)	homology modeling, ab initio structure prediction, and structure prediction using NMR constraints	http://robetta.bakerlab.org/

AlphaFold - o revoluție în predicția structurii tridimensionale a proteinelor



Article | [Open Access](#) | [Published: 15 July 2021](#)

Highly accurate protein structure prediction with AlphaFold

[John Jumper](#) , [Richard Evans](#), [Alexander Pritzel](#), [Tim Green](#), [Michael Figurnov](#), [Olaf Ronneberger](#), [Kathryn Tunyasuvunakool](#), [Russ Bates](#), [Augustin Židek](#), [Anna Potapenko](#), [Alex Bridgland](#), [Clemens Meyer](#), [Simon A. A. Kohl](#), [Andrew J. Ballard](#), [Andrew Cowie](#), [Bernardino Romera-Paredes](#), [Stanislav Nikolov](#), [Rishub Jain](#), [Jonas Adler](#), [Trevor Back](#), [Stig Petersen](#), [David Reiman](#), [Ellen Clancy](#), [Michal](#)

[Zielinski](#), ... [Demis Hassabis](#)  [+ Show authors](#)

[Nature](#) **596**, 583–589 (2021) | [Cite this article](#)

817k Accesses | **3625** Citations | **3272** Altmetric | [Metrics](#)

nature

[Explore content](#)  [About the journal](#)  [Publish with us](#)  [Subscribe](#)

[nature](#) > [news](#) > [article](#)

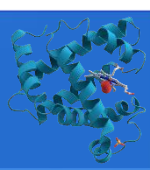
NEWS | 30 November 2020

‘It will change everything’: DeepMind’s AI makes gigantic leap in solving protein structures

Google’s deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

[Ewen Callaway](#)

AlphaFold - o revoluție în predicția structurii tridimensionale a proteinelor



AlphaFold - un program de inteligență artificială (AI) bazat pe algoritmi de învățare profundă:

- Programul a primit un set de antrenament format din peste 170.000 de proteine provenite dintr-un depozit public de secvențe și structuri proteice.
- Este bazat pe rețele neuronale artificiale și a învățat relația dintre structură și funcție folosind acest set de antrenament.
- În cazul unei secvențe necunoscute, algoritmul de inteligență artificială identifică fragmente de structură, pe care le assemblează ulterior pentru a obține structura tridimensională completă a secvenței analizate.

AlphaFold - a ocupat primul loc în clasamentul general al celei de-a 13-a ediții a competiției CASP (2018) (https://predictioncenter.org/casp13/zscores_final.cgi, Team A7D)

- A realizat predicții extrem de precise ale structurii pentru proteine considerate printre cele mai dificile, în absența unor șabloane structurale deja cunoscute;

AlphaFold 2 - a ocupat primul loc în clasamentul general al celei de-a 14-a ediții a competiției CASP (2020)

(https://predictioncenter.org/casp14/zscores_final.cgi);

- a atins un nivel de precizie considerat comparabil cu cel obținut prin metode experimentale, cum ar fi cristalografia cu raze X”

AlphaFold-multimer - cea mai utilizată abordare în cea de-a 15-a ediții CASP, 2022

- poate calcula cu acuratețe structura structurii cuaternare

AlphaFold 3 - predicția interacțiunilor dintre molecule

nature

Explore content ▾ About the journal ▾ Publish with us ▾

nature > articles > article

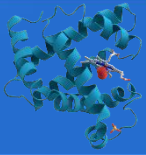
Article | [Open access](#) | Published: 08 May 2024

Accurate structure prediction of biomolecular interactions with AlphaFold 3

[Josh Abramson](#), [Jonas Adler](#), [Jack Dunger](#), [Richard Evans](#), [Tim Green](#), [Alexander Pritzel](#), [Olaf Ronneberger](#), [Lindsay Willmore](#), [Andrew J. Ballard](#), [Joshua Bambrick](#), [Sebastian W. Bodenstein](#), [David A. Evans](#), [Chia-Chun Hung](#), [Michael O'Neill](#), [David Reiman](#), [Kathryn Tunyasuvunakool](#), [Zachary Wu](#), [Akvilė Žemgulytė](#), [Eirini Arvaniti](#), [Charles Beattie](#), [Ottavia Bertolli](#), [Alex Bridgland](#), [Alexey Cherepanov](#), [Miles Congreve](#), ... [John M. Jumper](#)  + Show authors

Nature 630, 493–500 (2024) | [Cite this article](#)

AlphaFold - cum se poate utiliza?



deepmind / alphafold Public

<> Code Issues 111 Pull requests 10 Actions Projects Security Insights

main 1 branch 10 tags Go to file Code

Htomlinson14 and Copybara-Service Bump version to 2.2.4 ... 5cb2f8c 21 days ago 77 commits

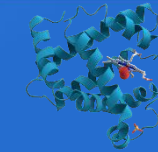
afdb	Internal change	3 months ago
alphafold	Fix incorrect length_scale argument	26 days ago
docker	Reduce size of docker image by cleaning apt, conda and pip caches	26 days ago
imgs	Initial release of AlphaFold.	15 months ago
notebooks	Make using GPU in relaxation optional, with default not to use.	last month
scripts	Update multimer models to v2.	7 months ago
.dockerignore	Update .dockerignore	9 months ago
CONTRIBUTING.md	Initial release of AlphaFold.	15 months ago

AlphaFold 2 disponibil pe GitHub și poate fi rulat:

1. **Local** - pe un calculator cu sistemul de operare Linux ce are la dispoziție GPUs și cel puțin 2.2TB spațiu. Programul se controlează din linia de comandă;
2. În **cloud** folosind Google Colab:

<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>

AlphaFold - exemple de aplicații



Input

MTDKRETLMSMLSKAYANPTIKAEPALRALIETNAKK
VDEGDDEKAYVTAVTQLSHDISKYLIHHAVPEELVAV
FNYIKKDVPAADIDAARYRAQALAAGLVAIPIVWGH

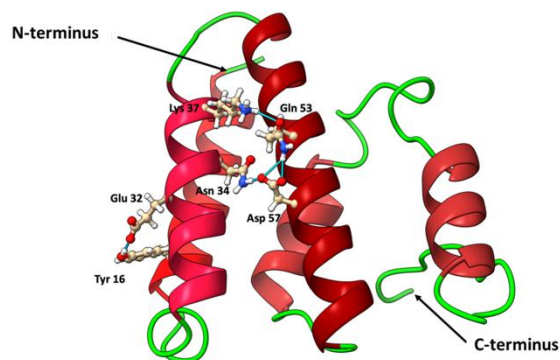
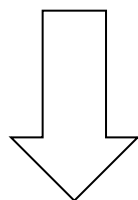
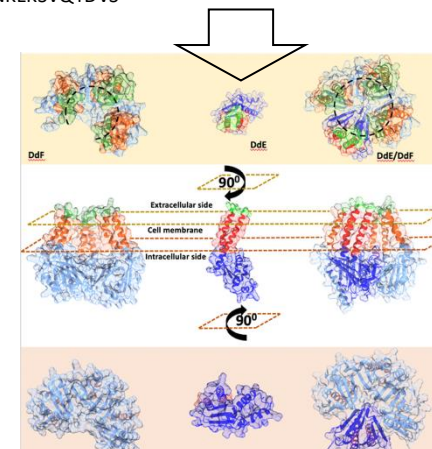


Figure 2. AlphaFold2 predicted structure of lacticaseicin 30 showing five helices (various shades of red) connected by coil regions (green). Five predicted H-bonds (blue) stabilize the position of the three helices.

Input

Protein DdC :
MENYFSNFSSDQNF MIDFLLSEGN ISKM CCKGYSYSKVKKLQY INEKIGKARYSQDSLKEYLDILVSEDI LFPEIAKLIYK
KHKEML
Protein DdD :
MSLNV IVLLALLAVQTWLIPLRNN KFLLLIVPSIFVLSIYIENLSLLIIGLMFGFFIYMAGLTQWDRINKERQKKYLKEK
FFNEK
Protein DdE :
MKNEFVMHHNYLLFSIVKNIIGSILVIGIAGLVQTIIDQKFFSIVLIVVIIIMPFDIFYEIYVKKYLIKTDYKNIYICEGLIIRKV
VIPKAVYTVIETEVYVFKKYWKKIELKTIVKDYKLLGIGLNDVRRRIKEEVIR
Protein DdF :
MKKLNKQIFLTNAVSIIRSLIIPFSVFLIKKKLTVLLCVYIYIYEVFKKVTYFYSVRYGFFETYFVFEKGMFNKRILRIDRQEL
KYVDSIEFTQSFLYKLLKQYKLVILNLKEEEDKIVIRCIDETEKYRFSSFFEVDTPVKLKENVLDLVFKPMKDDIIMSFFSANIFF
VLLSSLDYIDNFSYLNIPGKSITMNIYLVFLTIVFYSIILVQLRYGNYSVESTSEYLYKKN GILDYNSKELCIENVNTMIMYQ
NVFMRFWKIFNIYCLTTKNDNSKNEFKNM LFPYIRYDTFKNIVIEHSQFKFEILEKQLFKYRCLFCLLNTMCFIFLGLVLY
LGYFYGYRYWFLLLIFVCKRYFDSFFISILVKEKKICLIVIPGILFNRKVLILLNGIETIKYSNVFIFKKYKLIINSN NLP SKKYLR
GVTETQVSKVFKELNNKEKSVQTDVS



Article

Structure of Lacticaseicin 30 and Its Engineered Variants Revealed an Interplay between the N-Terminal and C-Terminal Regions in the Activity against Gram-Negative Bacteria

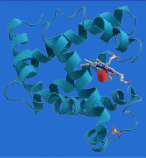
Désiré Madi-Moussa ¹, Barbara Deracinois ¹, Radja Teiar ¹, Yanyan Li ², Marius Mihasan ³,
Christophe Flahaut ¹, Sylvie Rebuffat ², Françoise Coucheny ^{1,*} and Djamel Drider ^{1,*}

Article


Advances in Characterizing the Transport Systems of and Resistance to EntDD14, A Leaderless Two-Peptide Bacteriocin with Potent Inhibitory Activity

Adrián Pérez-Ramos ¹, Rabia Ladjouzi ¹, Marius Mihasan ², Radja Teiar ¹, Abdellah Benachour ³
and Djamel Drider ^{1,*}

Strcuturi calculate cu AlphaFold disponibileavailable



<https://alphafold.ebi.ac.uk/>



The screenshot shows the AlphaFold Protein Structure Database website. The main heading is "AlphaFold Protein Structure Database" with the subtext "Developed by DeepMind and EMBL-EBI". A search bar is present with a "BETA" label and a "Search" button. Below the search bar, there are examples of search terms: "Free fatty acid receptor 2", "A1g58602", "Q5VSL9", and "E. coli". The website also features a navigation menu with options like "Home", "About", "FAQs", "Downloads", and "API". At the bottom, there is a section for "October Molecule of the Month" featuring a 3D structure of RSV Fusion Glycoprotein.

<https://www.rcsb.org/>