Repliement des protéines

avec la révolution d'AlphaFold2

Dirk Stratmann (dirk.stratmann@upmc.fr)

http://www.impmc.upmc.fr/~stratmann/

IMPMC, Sorbonne Université

octobre 2022

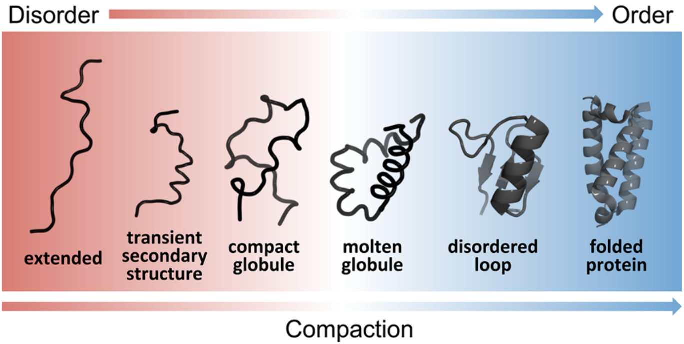
Plan

1. Processus du repliement
2. Techniques expérimentales
3. Repliement par dynamique moléculaire
4. Repliement in vivo
5. Protein design
6. Alpha Fold
7. Bibliographie

1. Processus du repliement
   - Processus du repliement
   - État déplié

Processus du repliement

Repliement



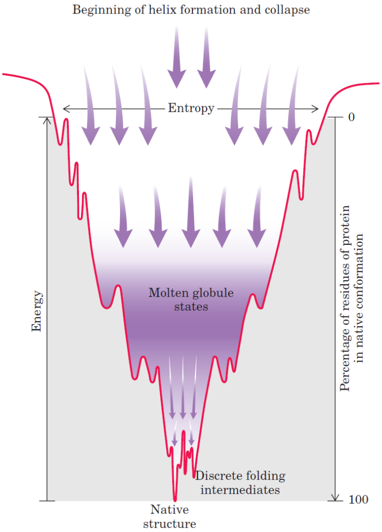
- Molten globule: interactions hydrophobes
- Structure 3D finale: liaisons hydrogènes

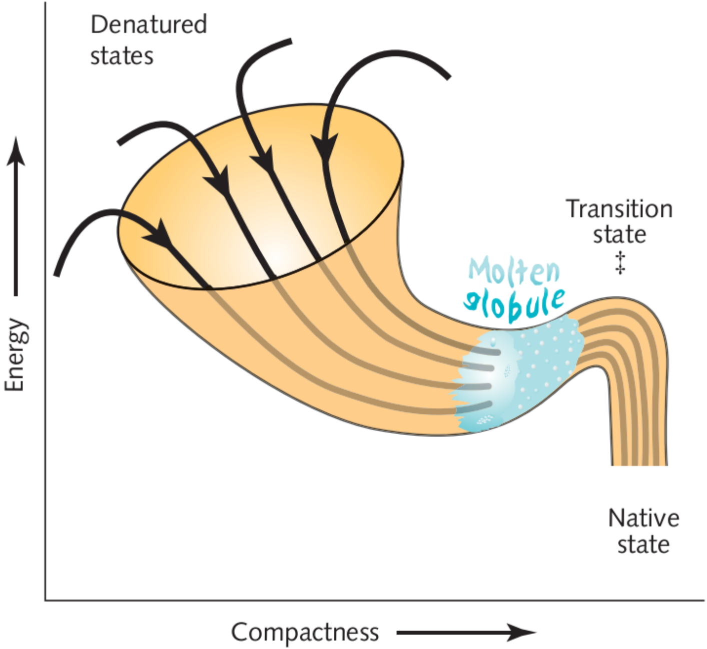
Pourquoi est le repliement des protéines si rapide ?

Paradoxe de Levinthal (1968):

- chaque résidu n'a que deux possibilité de conformation
- une protéine de 100 résidu aurait $2^{100} \approx 10^{30}$ conformations possibles.
- Conclusion: une protéine ne peut pas se replier par une recherche au hasard de la conformation native
- Elle doit suivre un chemin de repliement (*folding pathway*) plus efficace.

Entonnoir du repliement


Beginning of helix formation and collapse
Entropy
0
Energy
Molten globule states
Percentage of residues of protein in native conformation
Discrete folding intermediates
100
Native structure


Denatured states
Transition state
‡
Molten globule
Energy
Native state
Compactness

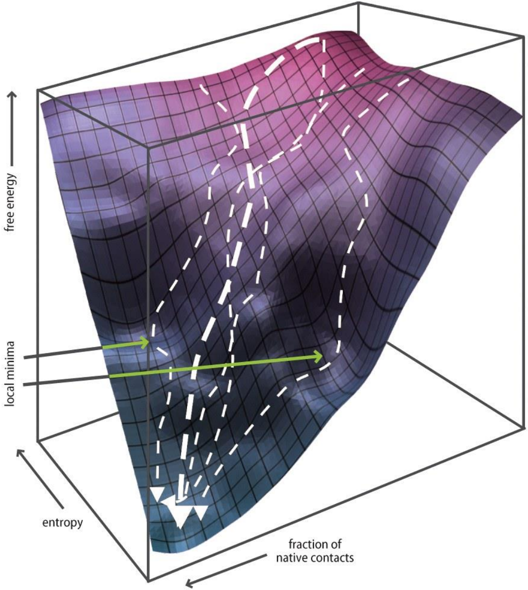
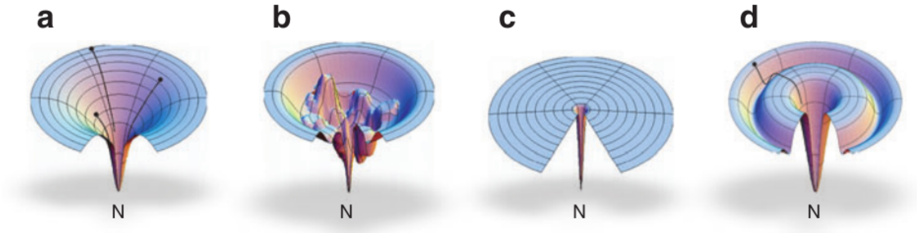


Figure 6.15 How Proteins Work (©2012 Garland Science)

Entonnoir du repliement



a: fast folder

b: kinetic traps

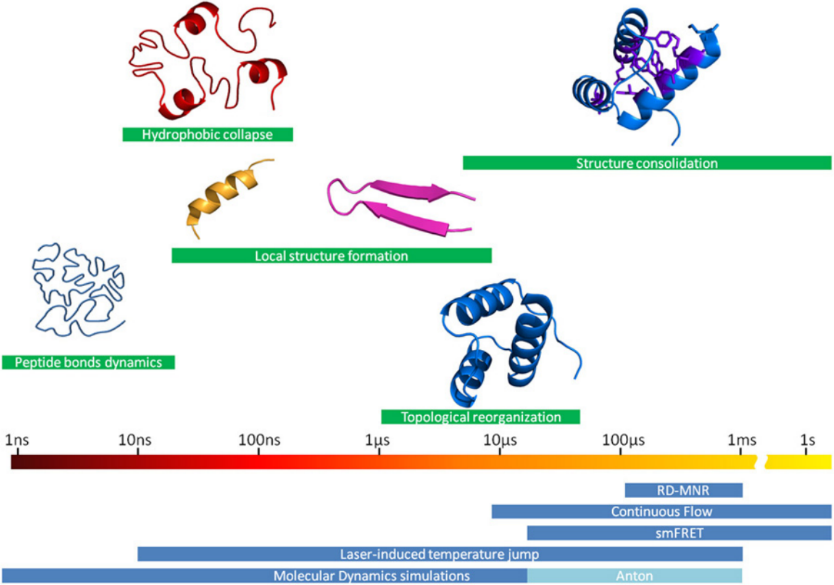
c: diffusional conformational search

d: obligatory intermediate

Ken A. Dill et al. (June 2008). en. In: *Annual Review of Biophysics* 37.1

Vitesses du repliement

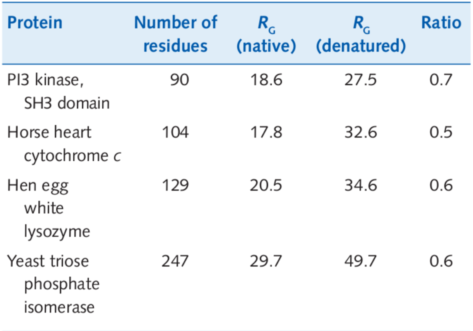
- Hélices se forment plus rapidement que des feuillet
- Hélices <=> contacts proches dans la séquence
- Feuillets <=> contacts distants dans la séquence
- Hélices: 0.1 - 1 μ*s*, β-hairpins: 1 - 10 μ*s*
- La vitesse 'limite' du repliement: N/50 μ*s*, N: nombre de résidus



Victor Muñoz and Michele Cerminara (Sept. 2016). In: *Biochem J* 473.17

État déplié

Rayon de giration



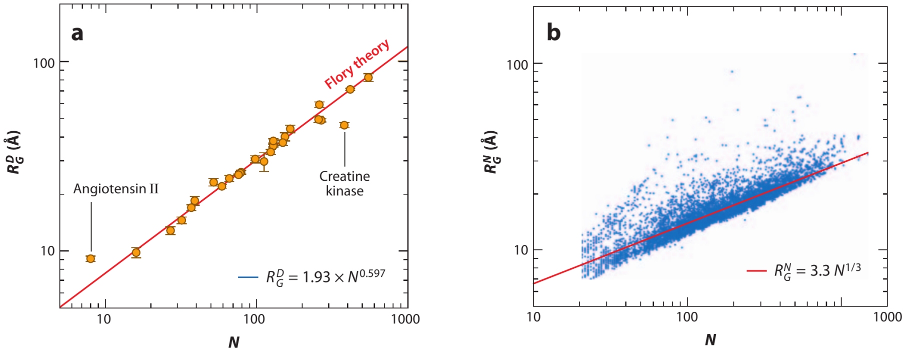
Protein	Number of residues	R_G (native)	R_G (denatured)	Ratio
PI3 kinase, SH3 domain	90	18.6	27.5	0.7
Horse heart cytochrome *c*	104	17.8	32.6	0.5
Hen egg white lysozyme	129	20.5	34.6	0.6
Yeast triose phosphate isomerase	247	29.7	49.7	0.6

- Si centre de gravité à l'origine, alors:

$$R_G = \sqrt{\frac{\sum_i m_i(x_i^2 + y_i^2 + z_i^2)}{\sum_i m_i}}$$

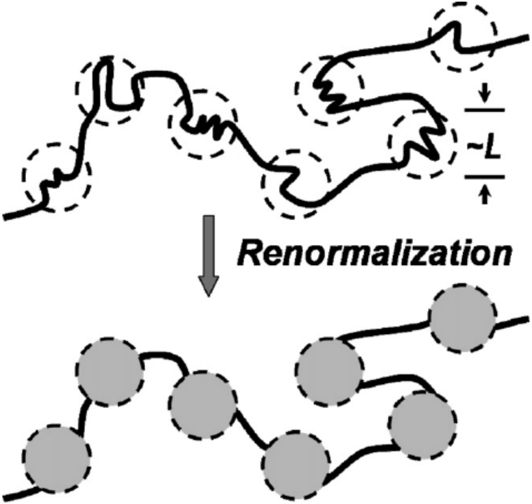
- Random coil: $R_G \propto N^{0.6}$ avec N nombre de résidus
- Globule compacte: $R_G \propto N^{0.33}$

Rayon de giration



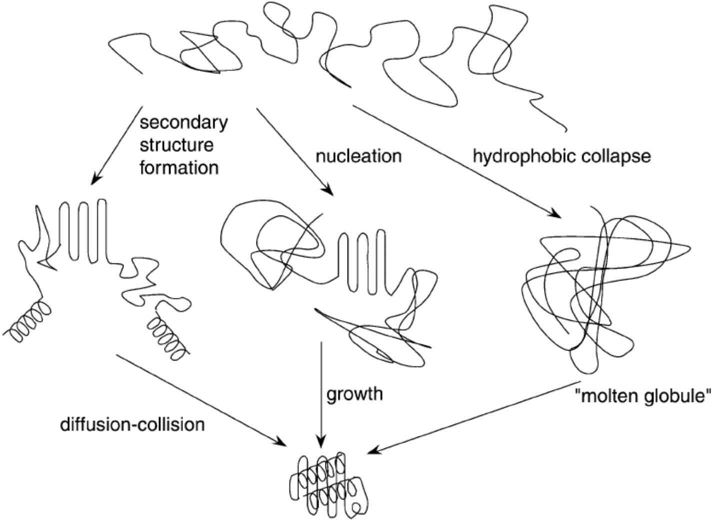
D. Thirumalai et al. (2010). In: *Annual Review of Biophysics* 39.1

Protéine dépliée = polymer sans interaction



Yiwen Chen et al. (Jan. 2008). In: *Archives of Biochemistry and Biophysics*. Highlight Issue: Protein Folding 469.1

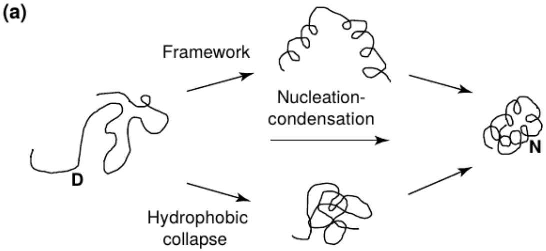
Modèles classiques pour le repliement



à gauche: "framework" ou "hierarchical",

Alan R. Fersht and Valerie Daggett (2002). In: *Cell*

Modèles classiques pour le repliement



(b)

Framework	engrailed homeodomain (α) protein A (α)	Barnase (α/β) Hen egg white lysozyme (α/β)
Nucleation-condensation	CI2 (α/β) cMyb (a) tenascin (β) FKBP12 (α/β)	
Hydrophobic collapse	?	

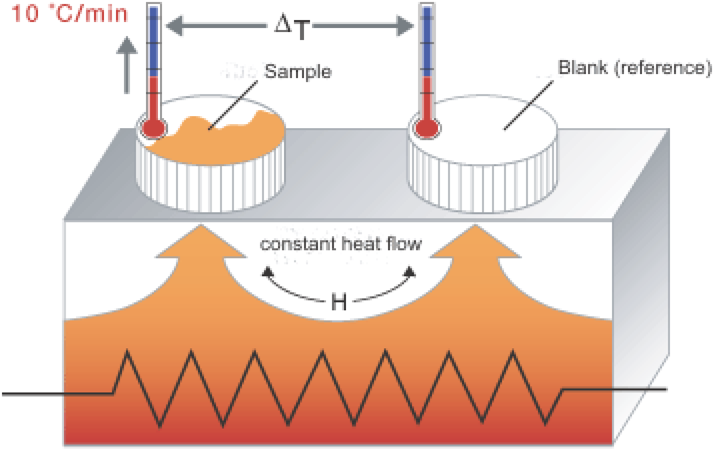
Valerie Daggett and Alan R. Fersht (2003). In: *Trends in biochemical sciences* 28.1

2 Techniques expérimentales

- Méthodes thérmodynamiques
- Techniques sur molécule unique
- RMN
- Méthodes pour cinétique "ultra rapide"

Méthodes thérmodynamiques

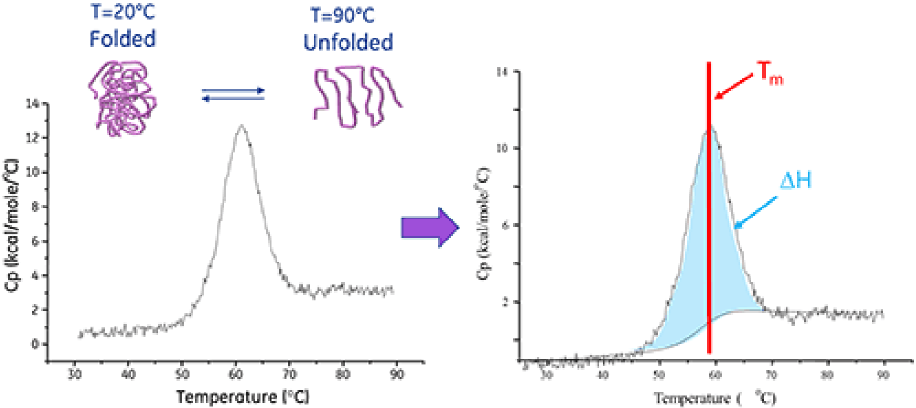
Differential scanning calorimetry (DSC)



Principe: référence et échantillion sont chauffés simultanément
Protéine absorbe de la chaleur lors du dépliement => ΔT

www.itc.tu-bs.de/Abteilungen/Makro/Methods/dsc.htm

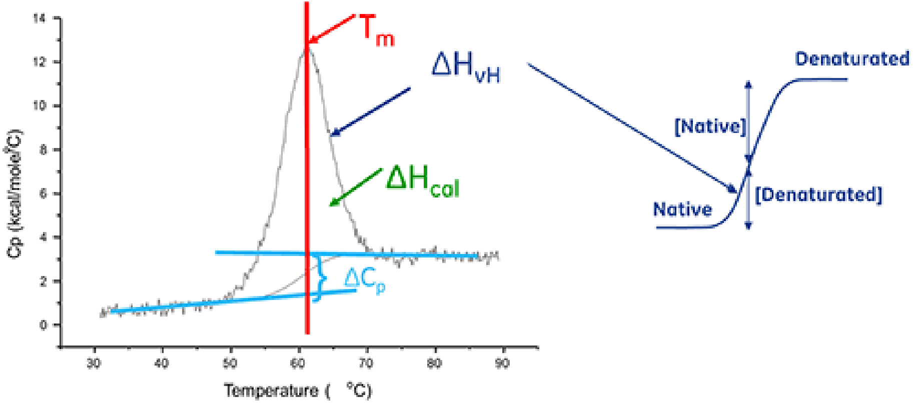
Differential scanning calorimetry (DSC)



Tm: température médiane de transistion thermique (thermal transition midpoint)
Cp: capacité calorifique (heat capacity)
H: enthalpie

www.malvern.com/fr/products/technology/differential-scanning-calorimetry

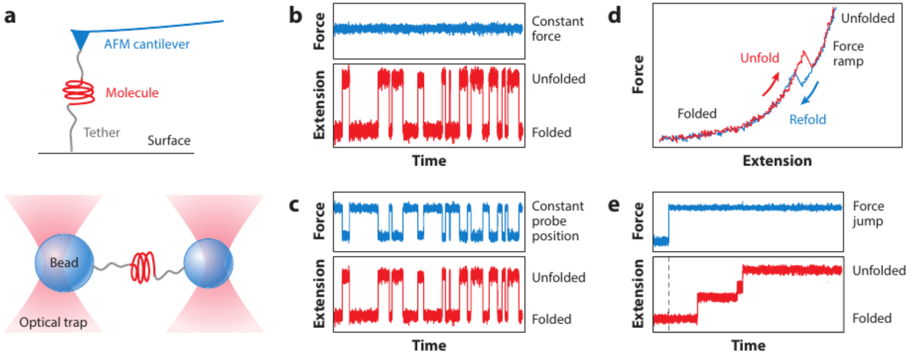
Differential scanning calorimetry (DSC)



Tm: température médiane de transistion thermique (thermal transition midpoint)
Cp: capacité calorifique (heat capacity)
H: enthalpie
www.malvern.com/fr/products/technology/differential-scanning-calorimetry

Techniques sur molécule unique

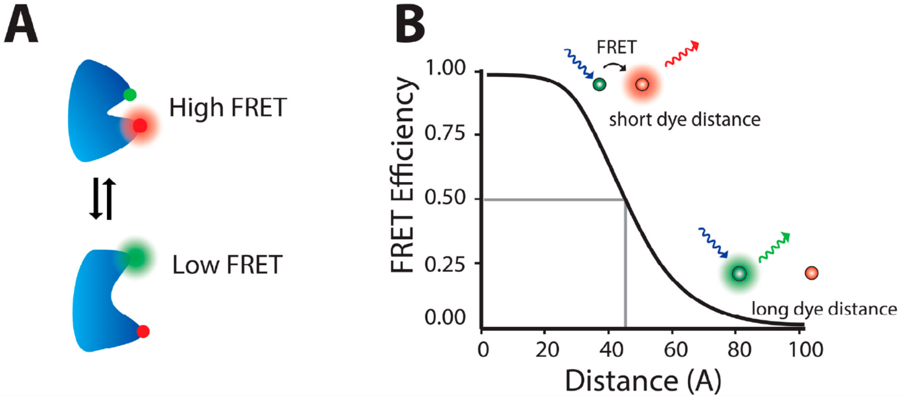
Single molecule force spectroscopy (SMFS)



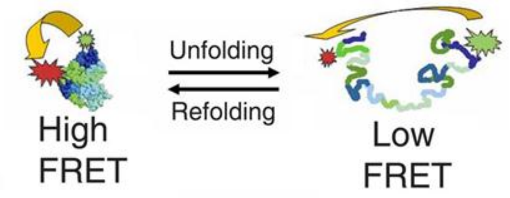
a) Atomic force microscopy (AFM) and Opticial tweezers
b) Constant force mode: extension fluctuates
c) Constant probe position: force and extension fluctuate
d) Force-ramp mode: elastic stretching is interrupted by a "rip", hysteresis indicates a nonequilibrium process
e) Force-jump mode: extension changes in steps

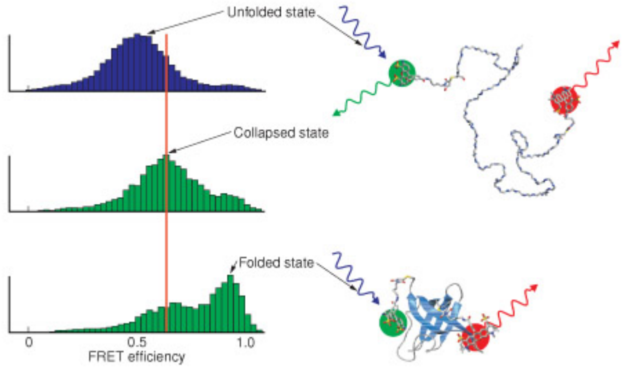
Michael T. Woodside and Steven M. Block (2014). In: *Annual Review of Biophysics* 43.1

Fluorescence resonance energy transfer (FRET)

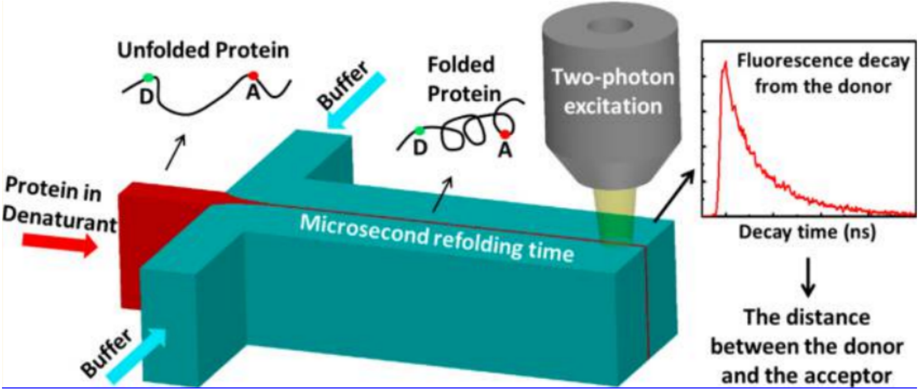


Fluorescence resonance energy transfer (FRET)

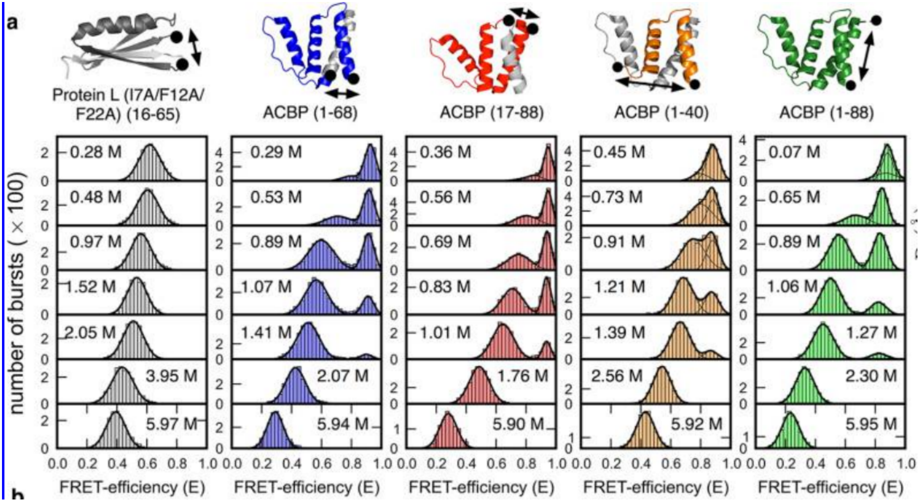




Time-resolved FRET (TR-FRET)


Unfolded Protein
D
A
Buffer
Folded Protein
D
A
Two-photon excitation
Fluorescence decay from the donor
Decay time (ns)
Protein in Denaturant
Microsecond refolding time
Buffer
The distance between the donor and the acceptor

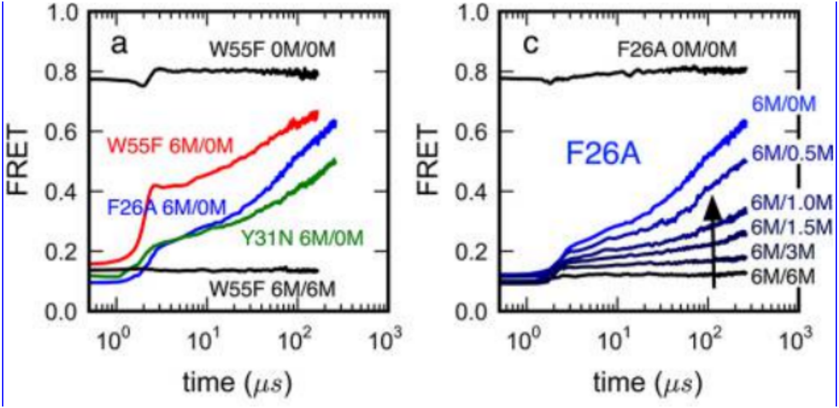
FRET sur ACBP



Vincent A. Voelz et al. (Aug. 2012). In: *J. Am. Chem. Soc.* 134.30

FRET sur ACBP - Collapse très lent

Solution dénaturant (6 M GuHCl) vers solution de repliement (0 M GuHCl):

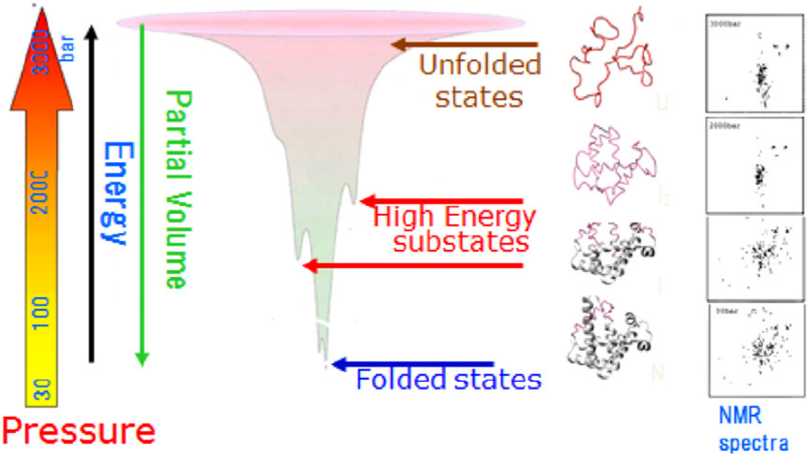


Indication pour: Formation d'une ensemble compact et hétérogènes de structures dépliées

Vincent A. Voelz et al. (Aug. 2012). In: *J. Am. Chem. Soc.* 134.30

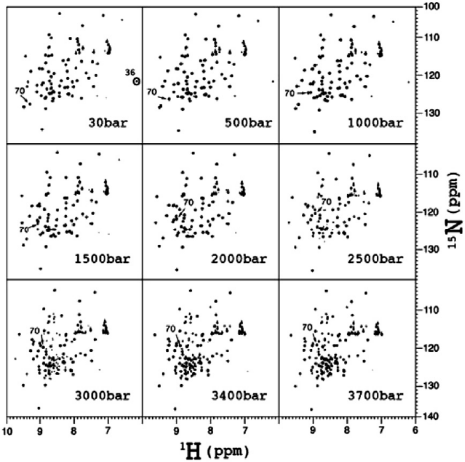
RMN

Pression et RMN



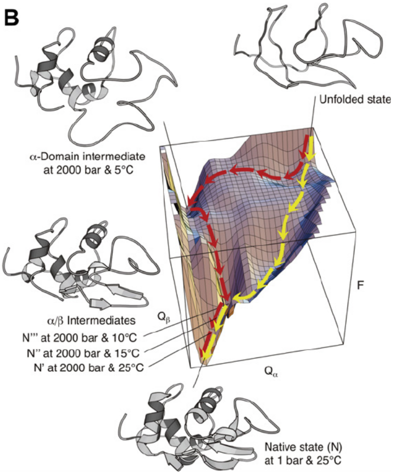
Kazuyuki Akasaka, Ryo Kitahara, and Yuji O. Kamatari (Mar. 2013). In: *Archives of Biochemistry and Biophysics*. Protein Folding and Stability 531.1–2

Pression et RMN



Kazuyuki Akasaka, Ryo Kitahara, and Yuji O. Kamatari (Mar. 2013). In: *Archives of Biochemistry and Biophysics*. Protein Folding and Stability 531.1–2

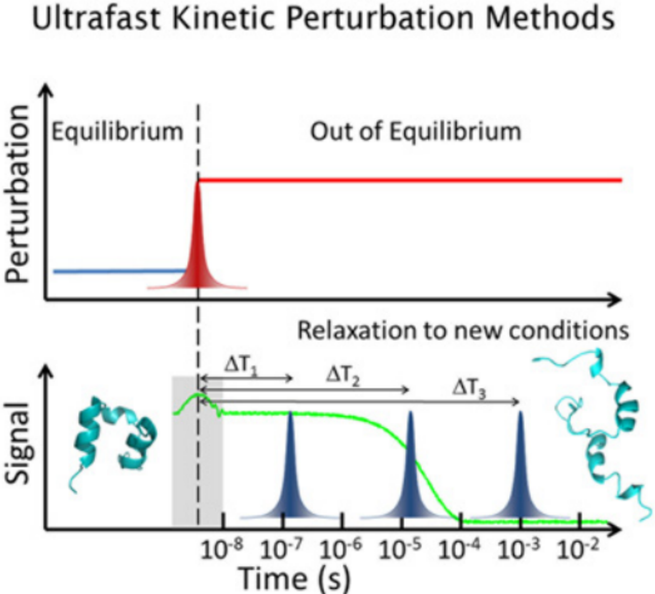
Pression et RMN



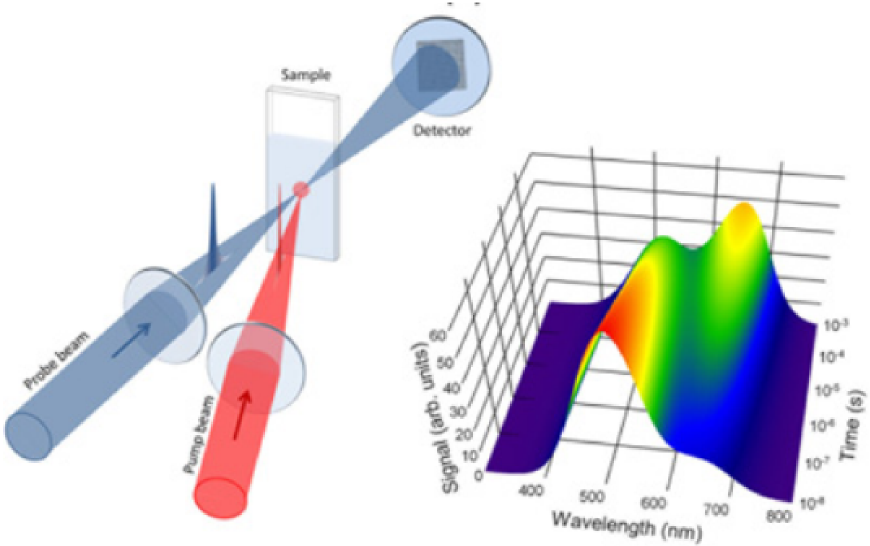
Kazuyuki Akasaka, Ryo Kitahara, and Yuji O. Kamatari (Mar. 2013). In: *Archives of Biochemistry and Biophysics*. Protein Folding and Stability 531.1–2

Méthodes pour cinétique "ultra rapide"

Laser T-jump


Ultrafast Kinetic Perturbation Methods
Perturbation
Equilibrium
Out of Equilibrium
Relaxation to new conditions
Signal
ΔT1
ΔT2
ΔT3
10-8
10-7
10-6
10-5
10-4
10-3
10-2
Time (s)

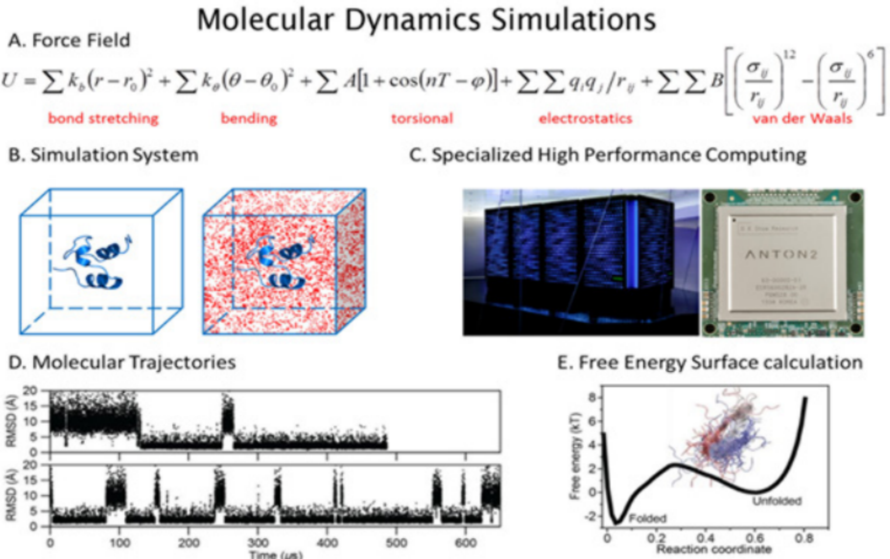
Laser T-jump


Sample
Detector
Probe beam
Pump beam
Signal (arb. units)
0
10
20
30
40
50
60
400
500
600
700
800
Wavelength (nm)
10⁻³
10⁻⁴
10⁻⁵
10⁻⁶
10⁻⁷
10⁻⁸
Time (s)

3. Repliement par dynamique moléculaire
    - Introduction
    - Temps de repliement
    - Markov State Models (MSM)
    - Etat déplié
    - Thermodynamique du repliement

Introduction

Anton



Molecular Dynamics Simulations

A. Force Field

$$U = \sum k_b(r-r_0)^2 + \sum k_\theta(\theta-\theta_0)^2 + \sum A[1+\cos(nT-\varphi)] + \sum\sum q_i q_j / r_{ij} + \sum\sum B\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

bond stretching bending torsional electrostatics van der Waals

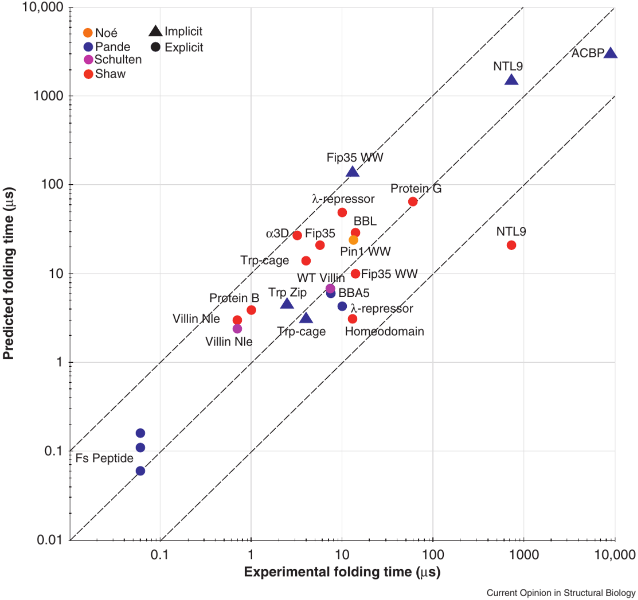
B. Simulation System

C. Specialized High Performance Computing

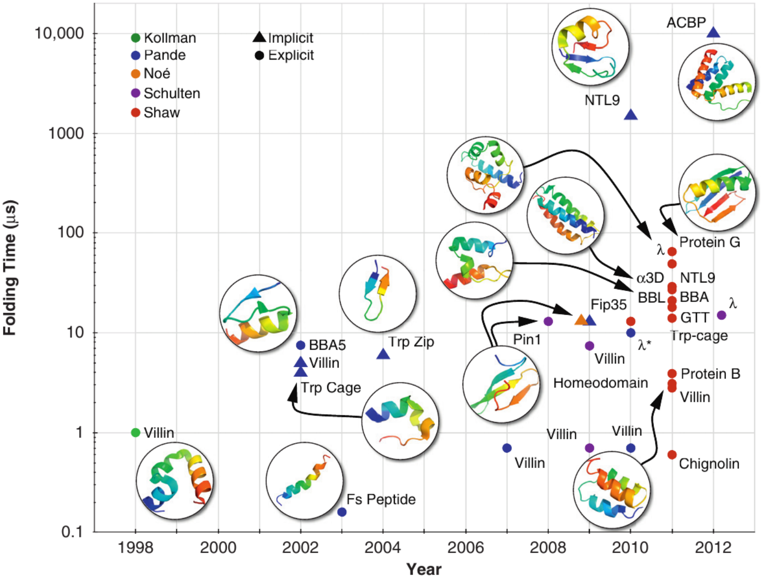
D. Molecular Trajectories

E. Free Energy Surface calculation

Temps de repliement



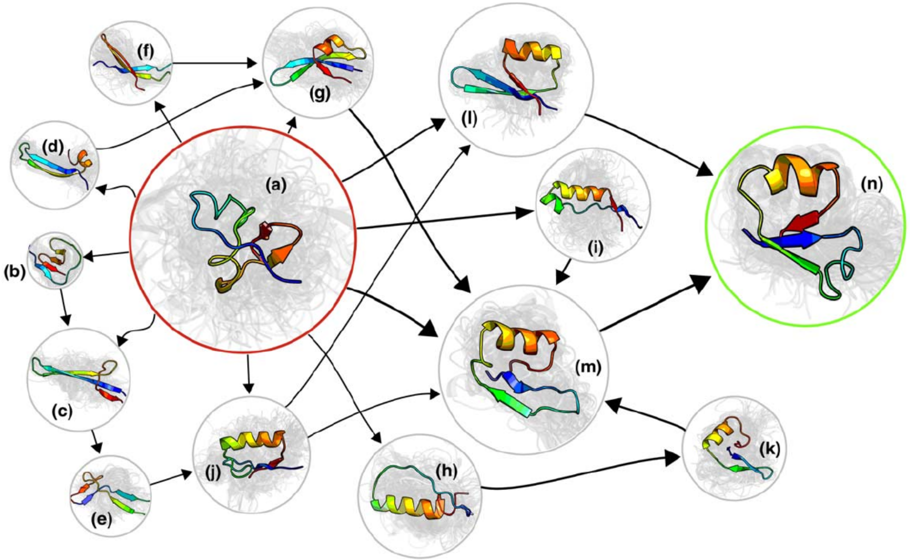
Thomas J Lane et al. (Feb. 2013). en. In: *Current Opinion in Structural Biology* 23.1



Current Opinion in Structural Biology

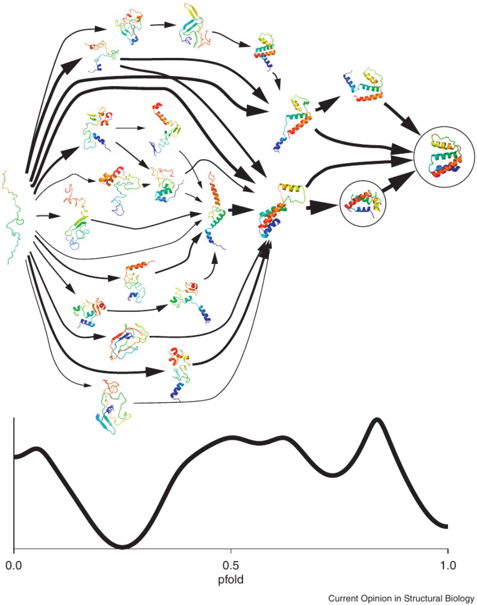
Thomas J Lane et al. (Feb. 2013). en. In: *Current Opinion in Structural Biology* 23.1

Markov State Models (MSM)



Current Opinion in Structural Biology

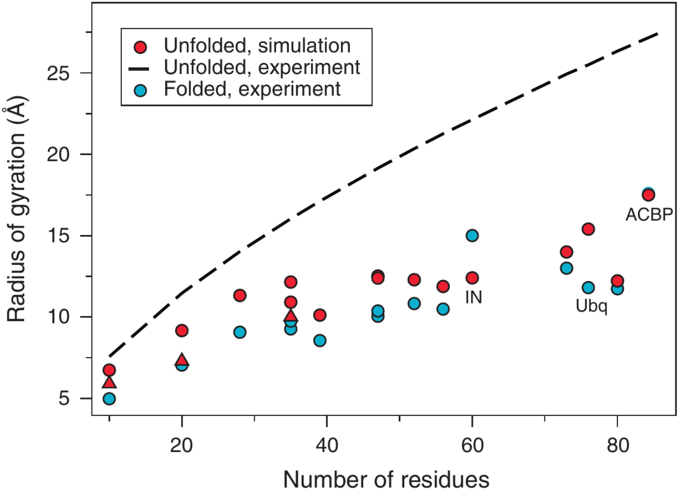
protéine NTL9, Gregory R Bowman, Vincent A Voelz, and Vijay S Pande (Feb. 2011). en. In: *Current Opinion in Structural Biology* 21.1



protéine ACBP, Thomas J Lane et al. (Feb. 2013). en. In: *Current Opinion in Structural Biology* 23.1

Etat déplié

Etat déplié: difficle en MD

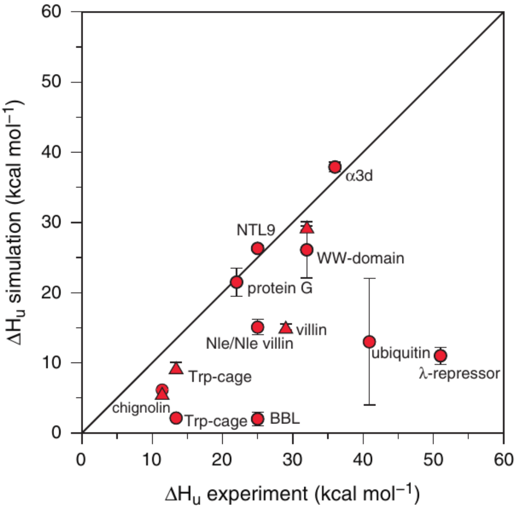


Current Opinion in Structural Biology

Stefano Piana, John L Klepeis, and David E Shaw (Feb. 2014). en. In: *Current Opinion in Structural Biology* 24

Thermodynamique du repliement

Prédiction de l’enthalpie du dépliement avec MD



Current Opinion in Structural Biology

Stefano Piana, John L Klepeis, and David E Shaw (Feb. 2014). en. In: *Current Opinion in Structural Biology* 24

Conclusions de Piana et al.

Peuvent les champs de forces en dynamique moléculaire reproduire les données expérimentales du repliement ?

Oui pour :

- Structure native, repliée
- Taux de repliement

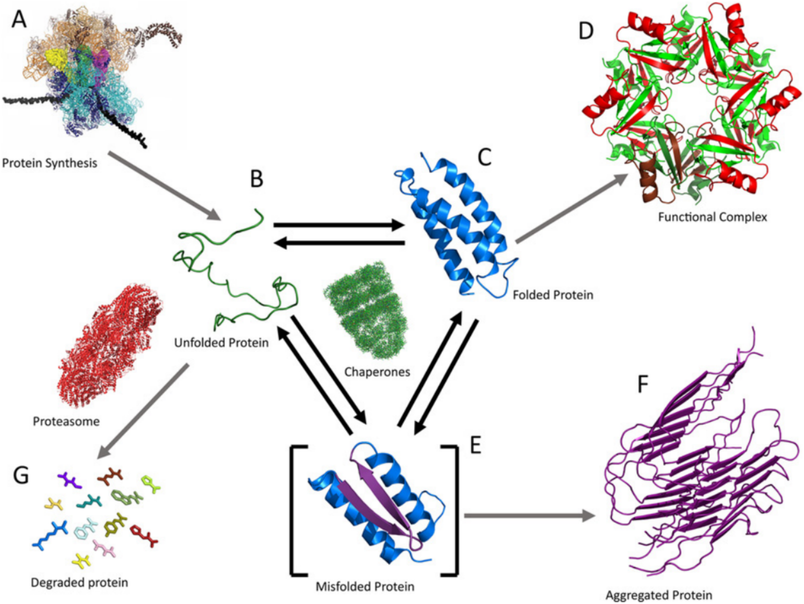
Non pour:

- Cinétique détaillée
- Structures dépliées
- Enthalpie du l'état repliée plus bas qu'en expérimental

4 Repliement in vivo

- Introduction
- Environnement cellulaire

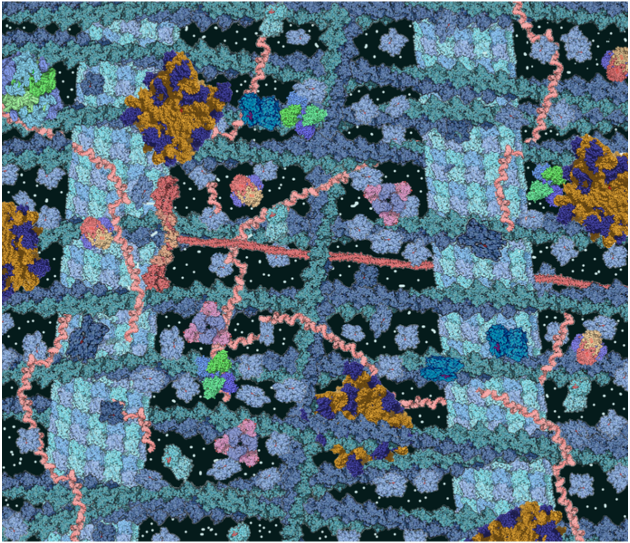
Introduction

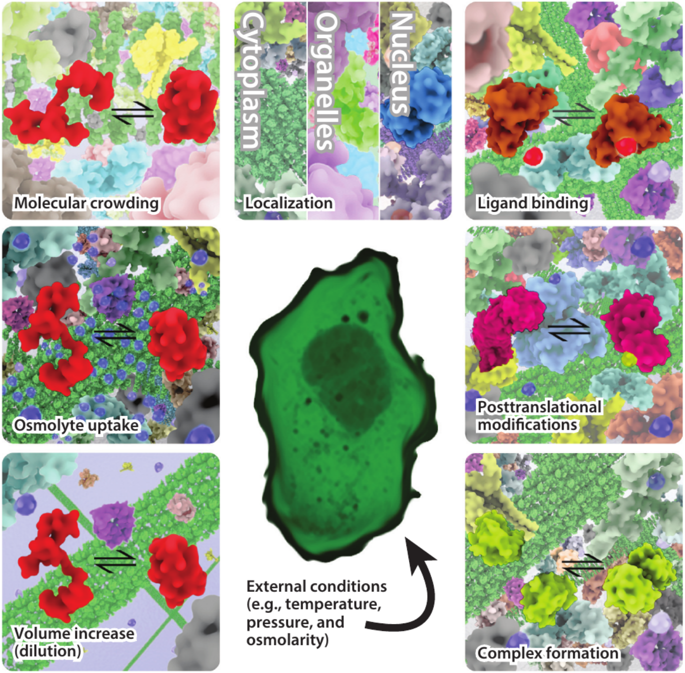


Victor Muñoz and Michele Cerminara (Sept. 2016). In: *Biochem J* 473.17

Environnement cellulaire

Molecular crowding



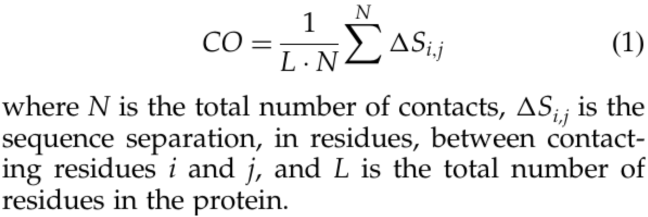

Molecular crowding
Cytoplasm
Organelles
Nucleus
Localization
Ligand binding
Osmolyte uptake
Posttranslational modifications
Volume increase (dilution)
External conditions (e.g., temperature, pressure, and osmolarity)
Complex formation

5 Protein design

- David Baker - first work
- David Baker - 15 years of protein design
- Peptides cycliques ciblant des interactions protéine-protéine

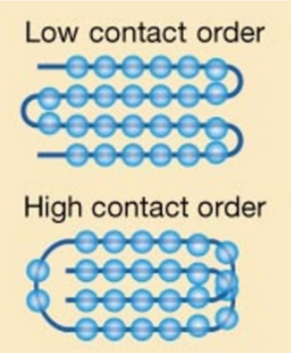
David Baker - first work

Contact order and folding kinetics



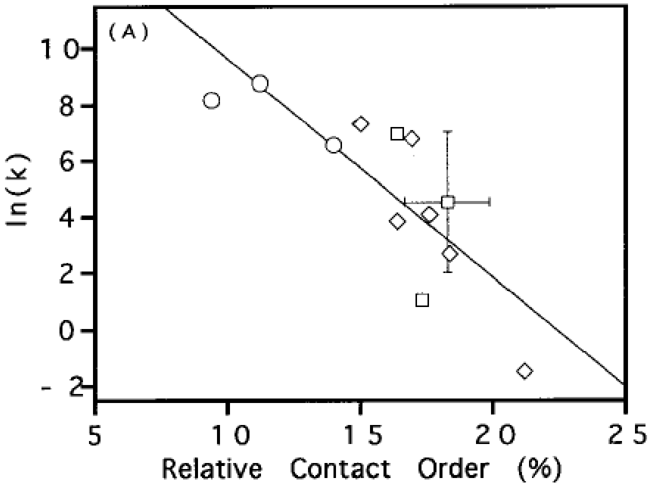
$$CO = \frac{1}{L \cdot N} \sum^{N} \Delta S_{i,j} \qquad (1)$$

where N is the total number of contacts, $\Delta S_{i,j}$ is the sequence separation, in residues, between contacting residues i and j, and L is the total number of residues in the protein.



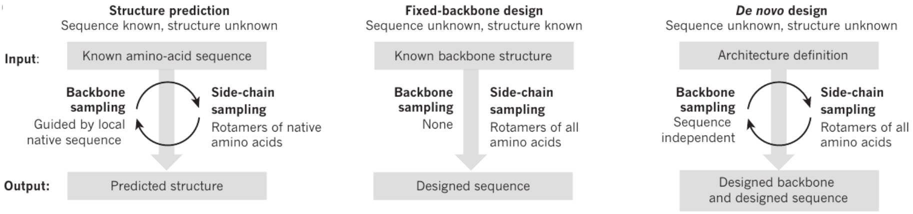
Kevin W Plaxco, Kim T Simons, and David Baker (Apr. 1998). In: *Journal of Molecular Biology* 277.4

Contact order and folding kinetics



Kevin W Plaxco, Kim T Simons, and David Baker (Apr. 1998). In: *Journal of Molecular Biology* 277.4

De novo protein design


Structure prediction
Sequence known, structure unknown
Fixed-backbone design
Sequence unknown, structure known
***De novo* design**
Sequence unknown, structure unknown
Input:
Known amino-acid sequence
Known backbone structure
Architecture definition
Backbone sampling
Guided by local native sequence
Side-chain sampling
Rotamers of native amino acids
Backbone sampling
None
Side-chain sampling
Rotamers of all amino acids
Backbone sampling
Sequence independent
Side-chain sampling
Rotamers of all amino acids
Output:
Predicted structure
Designed sequence
Designed backbone and designed sequence

Po-Ssu Huang, Scott E. Boyken, and David Baker (Sept. 2016). en. In: *Nature* 537.7620

Protein design - first sucess story

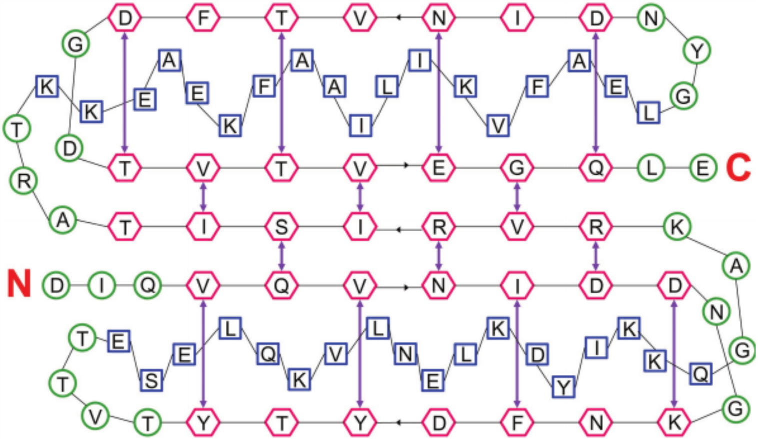
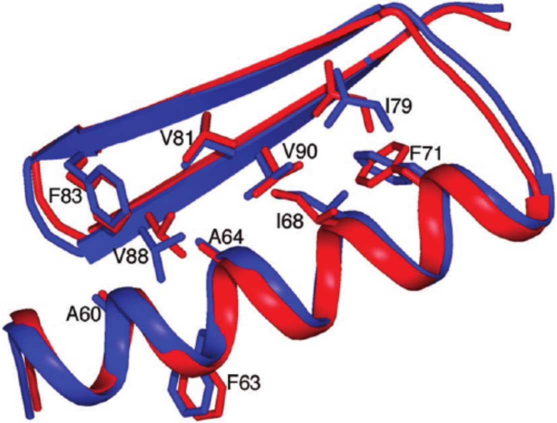


Fig. 1. A two-dimensional schematic of the target fold (hexagon, strand; square, helix; circle, other). Hydrogen bond partners are shown as purple arrows. The amino acids shown are those in the final designed (Top7) sequence.

Brian Kuhlman et al. (Nov. 2003). en. In: *Science* 302.5649

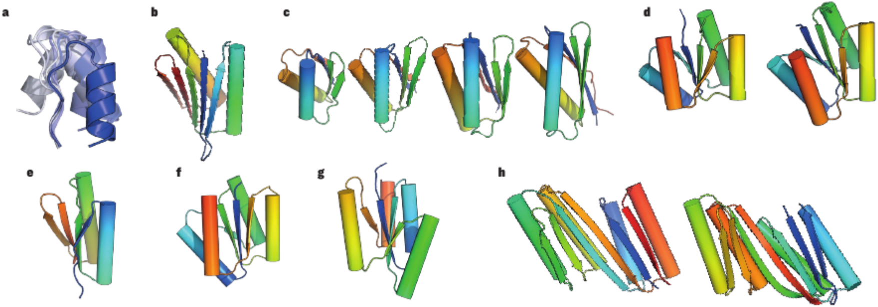
Protein design - first sucess story



Brian Kuhlman et al. (Nov. 2003). en. In: *Science* 302.5649

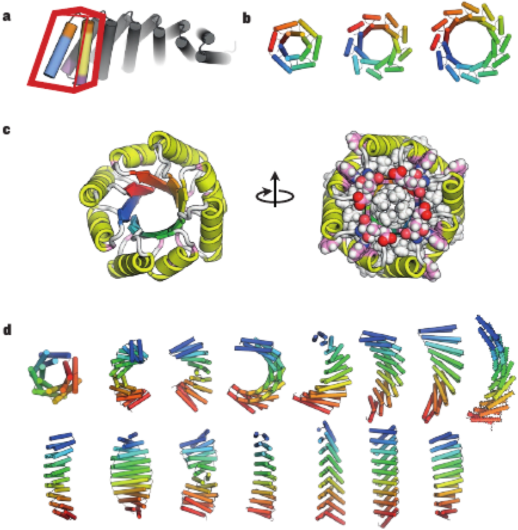
David Baker - 15 years of protein design

Protein design - examples



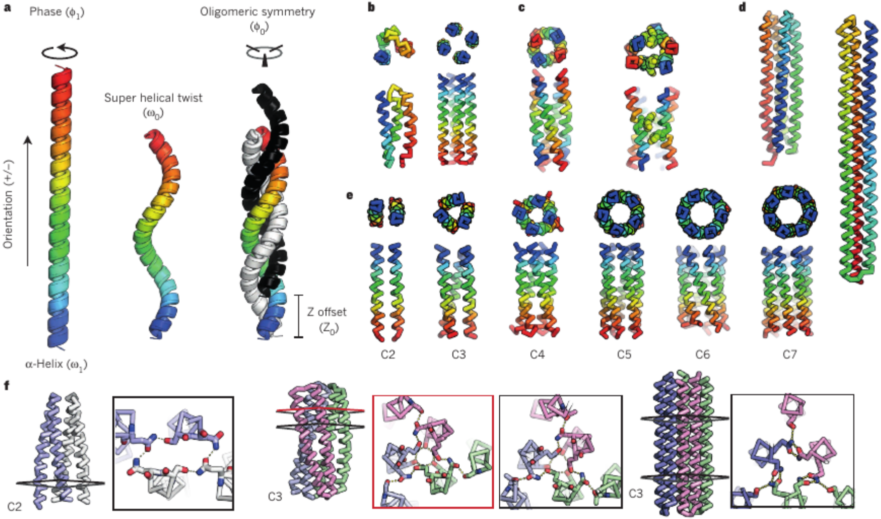
Po-Ssu Huang, Scott E. Boyken, and David Baker (Sept. 2016). en. In: *Nature* 537.7620

Protein design - examples



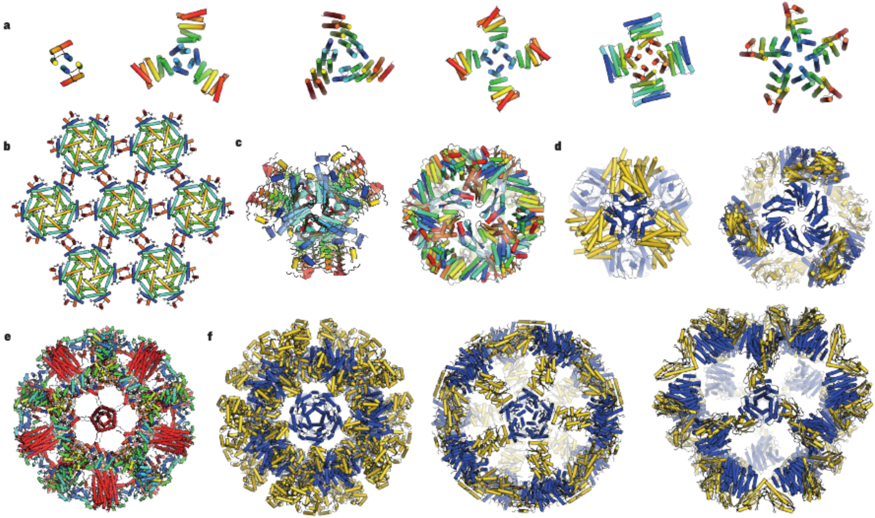
Po-Ssu Huang, Scott E. Boyken, and David Baker (Sept. 2016). en. In: *Nature* 537.7620

Protein design - examples



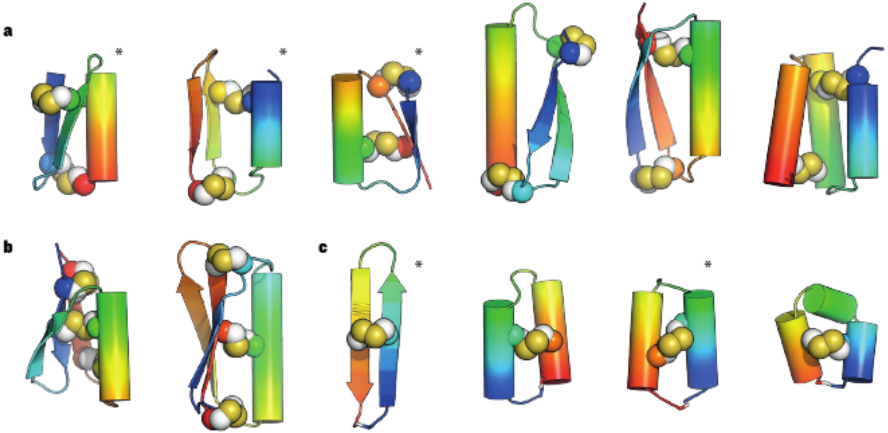
Po-Ssu Huang, Scott E. Boyken, and David Baker (Sept. 2016). en. In: *Nature* 537.7620

Protein design - examples



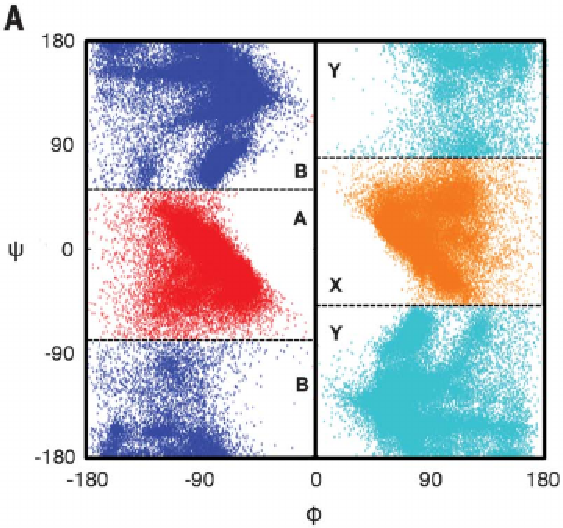
Po-Ssu Huang, Scott E. Boyken, and David Baker (Sept. 2016). en. In: *Nature* 537.7620

Protein design - examples



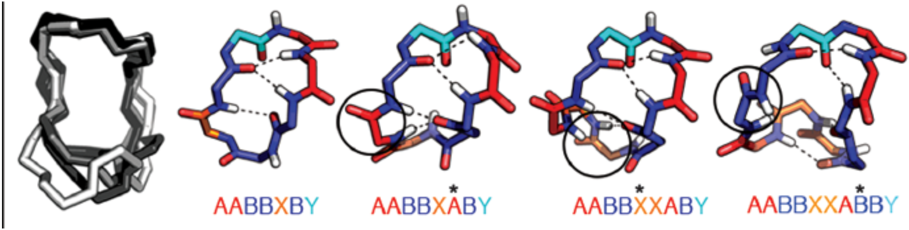
Po-Ssu Huang, Scott E. Boyken, and David Baker (Sept. 2016). en. In: *Nature* 537.7620

Protein design - macrocycles



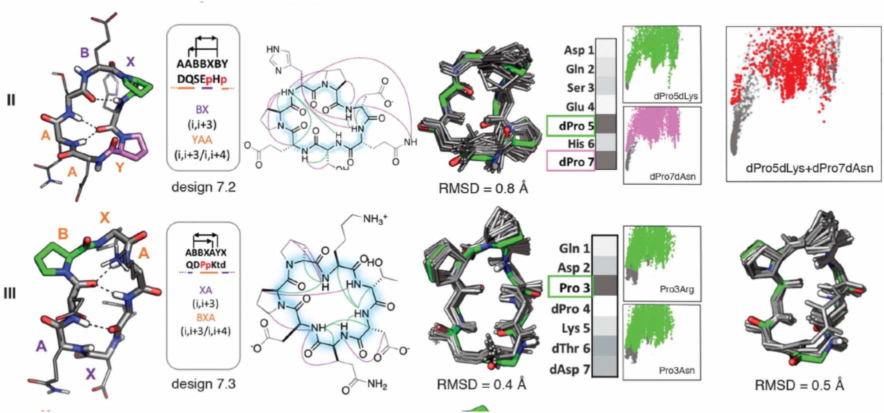
Parisa Hosseinzadeh et al. (Dec. 2017). en. In: *Science* 358.6369

Protein design - macrocycles



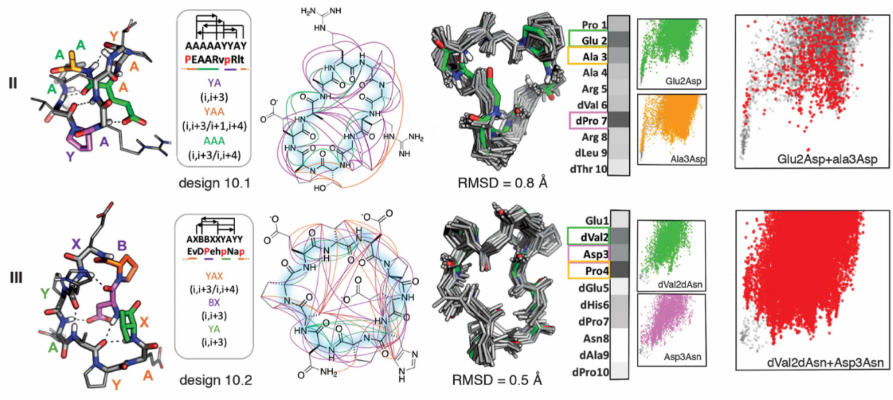
Parisa Hosseinzadeh et al. (Dec. 2017). en. In: *Science* 358.6369

Protein design - macrocycles



Parisa Hosseinzadeh et al. (Dec. 2017). en. In: *Science* 358.6369

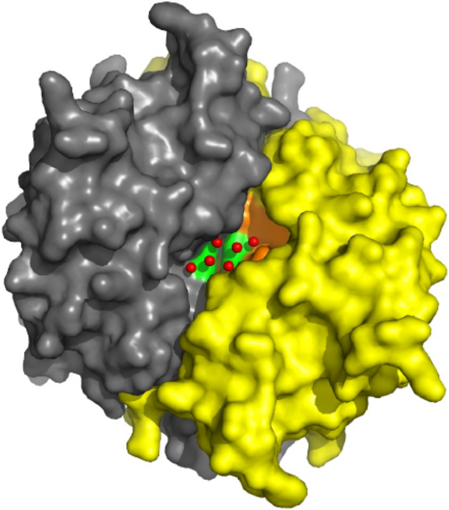
Protein design - macrocycles



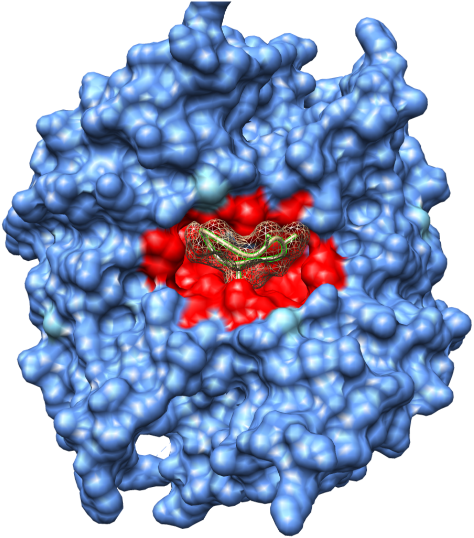
Parisa Hosseinzadeh et al. (Dec. 2017). en. In: *Science* 358.6369

Peptides cycliques ciblant des interactions protéine-protéine

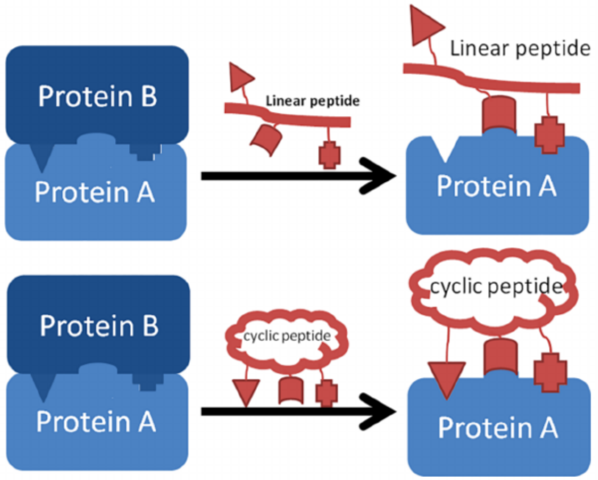
Drug design: cibler les interactions protéine-protéine



Drug = peptide cyclique



Drug = peptide cyclique



6 Alpha Fold

- Alpha Fold - La révolution
- Applications de Alpha Fold
- Autres méthodes utilisant le deep learning
- Comprendre Alpha Fold
- Mutations corrélées, MSA
- Utiliser Alpha Fold
- Pour aller plus loin: EMBL webinar
- L’avenir selon DeepMind

Alpha Fold - La révolution

Le press release de CASP du 30 nov 2020

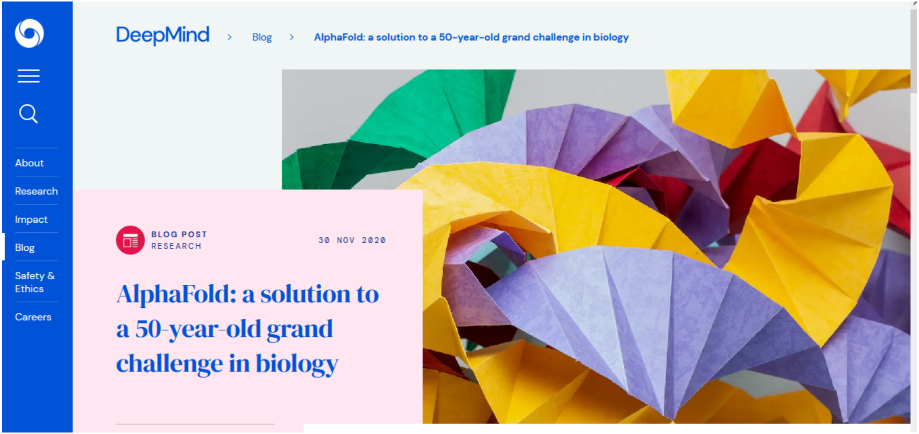
"Artificial intelligence solution to a 50-year-old science challenge could ‘revolutionise’ medical research"

"Today (Monday) researchers at the 14th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP14) will announce that an artificial intelligence (AI) solution to the challenge has been found."

"Nearly 50 years ago, Christian Anfinsen was awarded a Nobel Prize for showing that it should be possible to determine the shape of proteins based on their sequence of amino acids – the individual building blocks that make up proteins. That’s why our community of scientists have been working on the biennial CASP challenge."

`https://predictioncenter.org/casp14/doc/CASP14_press_release.html`

DeepMind de google basé à Londres



DeepMind › Blog › AlphaFold: a solution to a 50-year-old grand challenge in biology

BLOG POST
RESEARCH

30 NOV 2020

AlphaFold: a solution to a 50-year-old grand challenge in biology

Le relais immédiat dans Nature de l’annonce de CASP

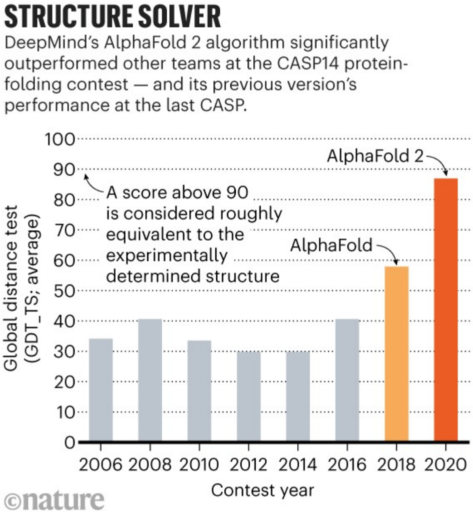
30 nov 2020, Nature, "’It will change everything’: DeepMind’s AI makes gigantic leap in solving protein structures"

`https://www.nature.com/articles/d41586-020-03348-4`

La presse

30 nov 2020, The New York Times, London AI claims breakthrough that could accelerate drug discovery
1 déc 2020, France Culture, Le "problème de repliement des protéines" résolu par une intelligence artificielle
11 déc 2020, Les Échos, DeepMind met les chercheurs du monde entier au tapis

CASP 14 et AlphaFold 2

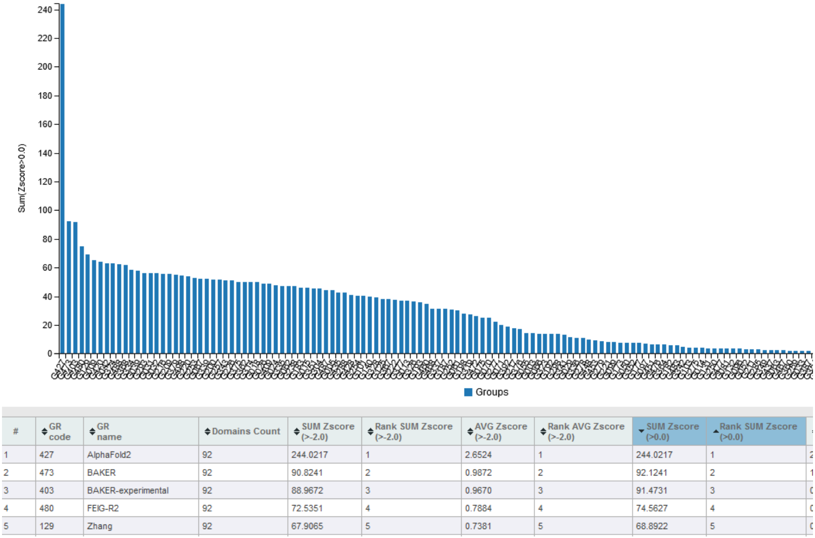


STRUCTURE SOLVER

DeepMind's AlphaFold 2 algorithm significantly outperformed other teams at the CASP14 protein-folding contest — and its previous version's performance at the last CASP.

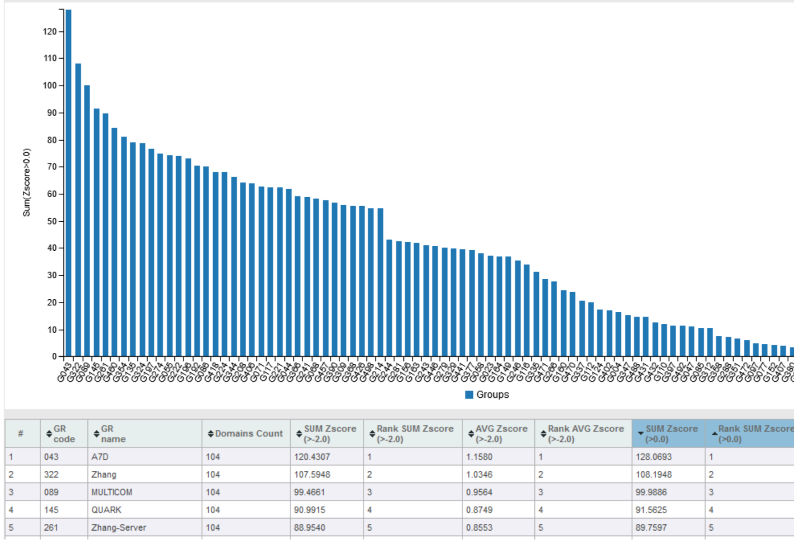
©nature

CASP 14 et AlphaFold 2



#	GR code	GR name	Domains Count	SUM Zscore (>-2.0)	Rank SUM Zscore (>-2.0)	AVG Zscore (>-2.0)	Rank AVG Zscore (>-2.0)	SUM Zscore (>0.0)	Rank SUM Zscore (>0.0)	
1	427	AlphaFold2	92	244.0217	1	2.6524	1	244.0217	1	2
2	473	BAKER	92	90.8241	2	0.9872	2	92.1241	2	1
3	403	BAKER-experimental	92	88.9672	3	0.9670	3	91.4731	3	0
4	480	FEIG-R2	92	72.5351	4	0.7884	4	74.5627	4	0
5	129	Zhang	92	67.9065	5	0.7381	5	68.8922	5	0

CASP 13 et AlphaFold 1


Sum(Zscore>0.0)
0
10
20
30
40
50
60
70
80
90
100
110
120
G043 G322 G089 G145 G261 G460 G354 G135 G324 G197 G274 G055 G222 G196 G192 G086 G418 G224 G344 G208 G406 G071 G117 G221 G044 G366 G241 G068 G457 G390 G309 G368 G426 G498 G214 G244 G281 G156 G163 G243 G446 G279 G329 G441 G377 G058 G023 G164 G149 G246 G118 G335 G471 G266 G160 G470 G337 G112 G124 G402 G004 G347 G488 G431 G432 G110 G397 G492 G047 G085 G312 G358 G288 G351 G472 G097 G077 G152 G407 G380
Groups

#	GR code	GR name	Domains Count	SUM Zscore (>-2.0)	Rank SUM Zscore (>-2.0)	AVG Zscore (>-2.0)	Rank AVG Zscore (>-2.0)	SUM Zscore (>0.0)	Rank SUM Zscore (>0.0)
1	043	A7D	104	120.4307	1	1.1580	1	128.0693	1
2	322	Zhang	104	107.5948	2	1.0346	2	108.1948	2
3	089	MULTICOM	104	99.4661	3	0.9564	3	99.9886	3
4	145	QUARK	104	90.9915	4	0.8749	4	91.5625	4
5	261	Zhang-Server	104	88.9540	5	0.8553	5	89.7597	5

La presse en 2021

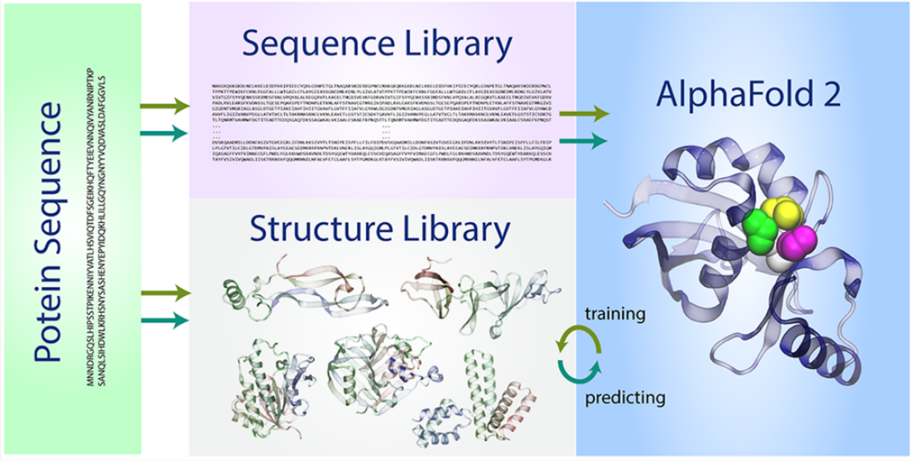
3 oct 2021, Forbes: "AlphaFold is the most important achievement in AI - Ever"

`https://www.forbes.com/sites/robtoews/2021/10/03/alphafold-is-the-most-important-achievement-in-ai-ever?sh=2857ff656e0a`

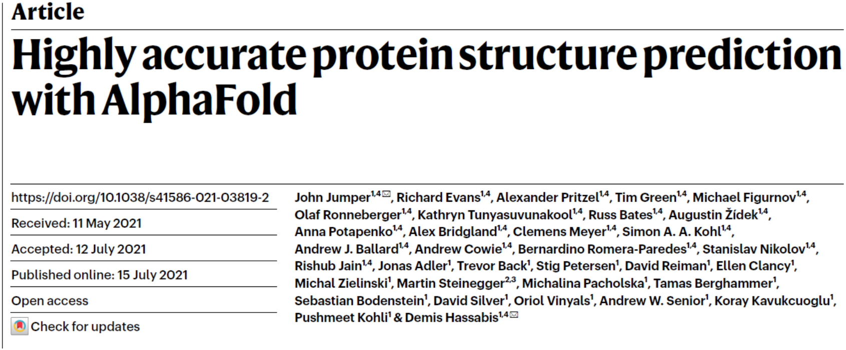
22 juillet 2021, Fortune: "In giant leap for biology, DeepMind’s A.I. reveals secret building blocks of human life"

`https://fortune.com/2021/07/22/deepmind-alphafold-human-proteome-database-proteins/`

Principe de Alpha Fold 2


Potein Sequence
MNNDRGQSLHIPSSTPIKENNIYVATLHSVIQTDFSGEIKHQFTYEIEVNNQIVYANRNIPTKP
SANQLSIHDWLKRHSNYSASHENYEPYIDQKHLILLGQYNGNYYVQDVASLDAFGGVLS
Sequence Library
Structure Library
AlphaFold 2
training
predicting

15 juillet 2021: Détails de la méthode et code source



Article

Highly accurate protein structure prediction with AlphaFold

https://doi.org/10.1038/s41586-021-03819-2

Received: 11 May 2021

Accepted: 12 July 2021

Published online: 15 July 2021

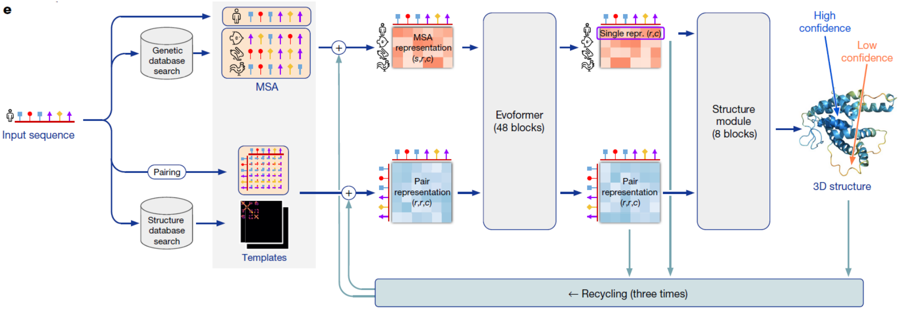
Open access

Check for updates

John Jumper[1,4]✉, Richard Evans[1,4], Alexander Pritzel[1,4], Tim Green[1,4], Michael Figurnov[1,4], Olaf Ronneberger[1,4], Kathryn Tunyasuvunakool[1,4], Russ Bates[1,4], Augustin Žídek[1,4], Anna Potapenko[1,4], Alex Bridgland[1,4], Clemens Meyer[1,4], Simon A. A. Kohl[1,4], Andrew J. Ballard[1,4], Andrew Cowie[1,4], Bernardino Romera-Paredes[1,4], Stanislav Nikolov[1,4], Rishub Jain[1,4], Jonas Adler[1], Trevor Back[1], Stig Petersen[1], David Reiman[1], Ellen Clancy[1], Michal Zielinski[1], Martin Steinegger[2,3], Michalina Pacholska[1], Tamas Berghammer[1], Sebastian Bodenstein[1], David Silver[1], Oriol Vinyals[1], Andrew W. Senior[1], Koray Kavukcuoglu[1], Pushmeet Kohli[1] & Demis Hassabis[1,4]✉

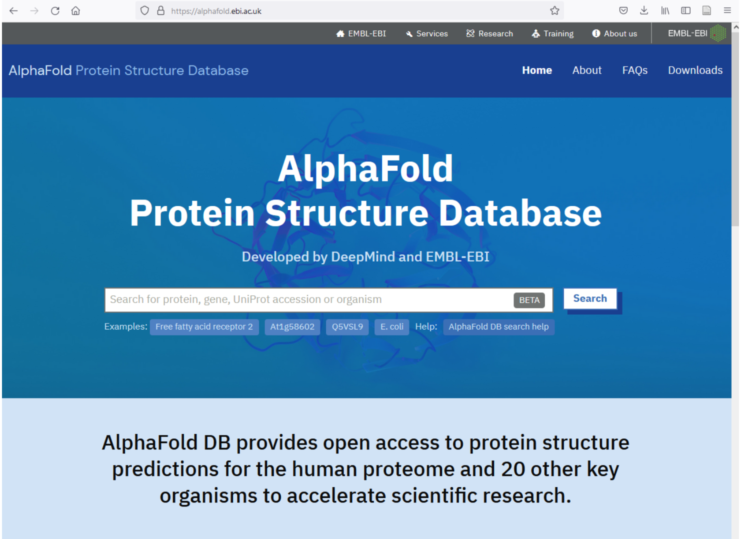
+5000 citations sur google scholar à ce jour (6 oct 2022) !

Principe de Alpha Fold 2


e
Input sequence
Genetic database search
Pairing
Structure database search
MSA
Templates
MSA representation (s,r,c)
Pair representation (r,r,c)
Evoformer (48 blocks)
Single repr. (r,c)
Pair representation (r,r,c)
Structure module (8 blocks)
High confidence
Low confidence
3D structure
← Recycling (three times)

Applications de Alpha Fold

Alpha Fold DB



AlphaFold Protein Structure Database

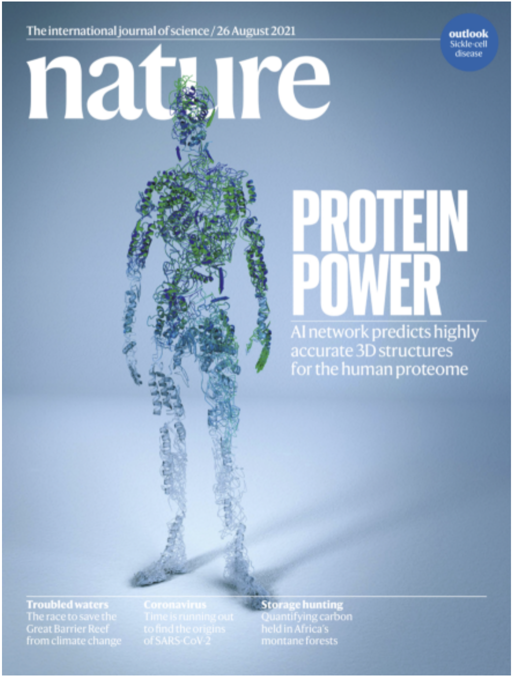
AlphaFold
Protein Structure Database

Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism

Examples: Free fatty acid receptor 2 At1g58602 Q5VSL9 E. coli Help: AlphaFold DB search help

AlphaFold DB provides open access to protein structure predictions for the human proteome and 20 other key organisms to accelerate scientific research.



The international journal of science / 26 August 2021

outlook
Sickle-cell
disease

nature

PROTEIN
POWER

AI network predicts highly
accurate 3D structures
for the human proteome

Troubled waters
The race to save the
Great Barrier Reef
from climate change

Coronavirus
Time is running out
to find the origins
of SARS-CoV-2

Storage hunting
Quantifying carbon
held in Africa's
montane forests

Autres méthodes utilisant le deep learning

Le pionnier: RaptorX de Jinbo Xu et al.

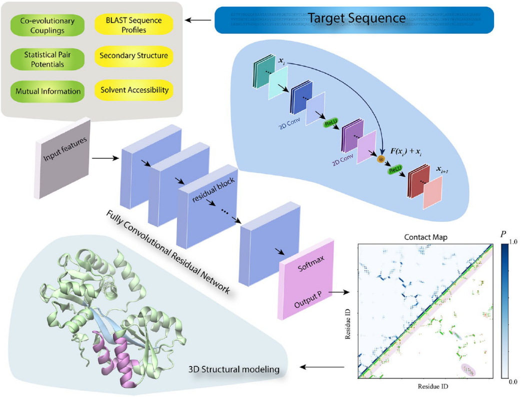
Serveur RaptorX:
`http://raptorx.uchicago.edu/`
CASP 14: score 38 vs 244 pour AlphaFold 2
"This server was officially ranked 1st in contact prediction in both CASP12 and CASP13 and initiated the revolution of protein structure prediction by deep learning."

DESTINI de Jeffrey Skolnick et al.

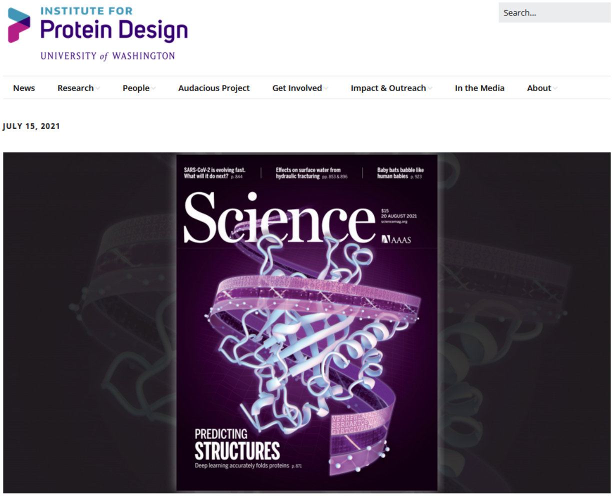
Serveur DESTINI:

`https://sites.gatech.edu/cssb/destini/`

CASP 14: score 29 vs 244 pour AlphaFold 2



Le rattrapage de David Baker



INSTITUTE FOR Protein Design

UNIVERSITY of WASHINGTON

Search...

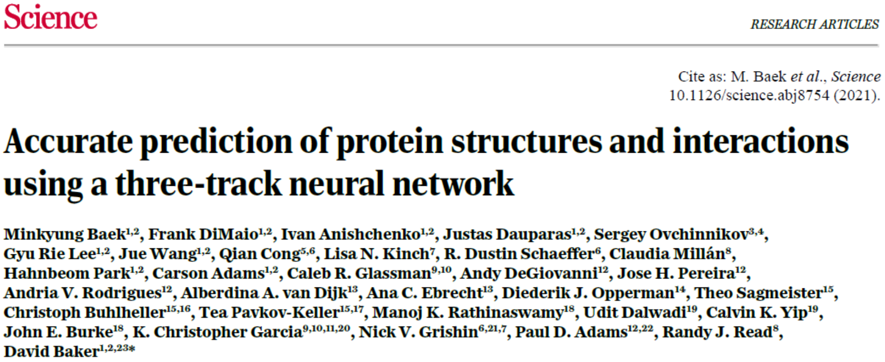
News | Research | People | Audacious Project | Get Involved | Impact & Outreach | In the Media | About

JULY 15, 2021

SARS-CoV-2 is evolving fast. What will it do next? p. 844
Effects on surface water from hydraulic fracturing pp. 853 & 896
Baby bats babble like human babies p. 923
Science
$15
20 AUGUST 2021
sciencemag.org
AAAS
PREDICTING STRUCTURES
Deep learning accurately folds proteins p. 871

RoseTTAFold: Accurate protein structure prediction accessible to all

Le rattrapage de David Baker



Science

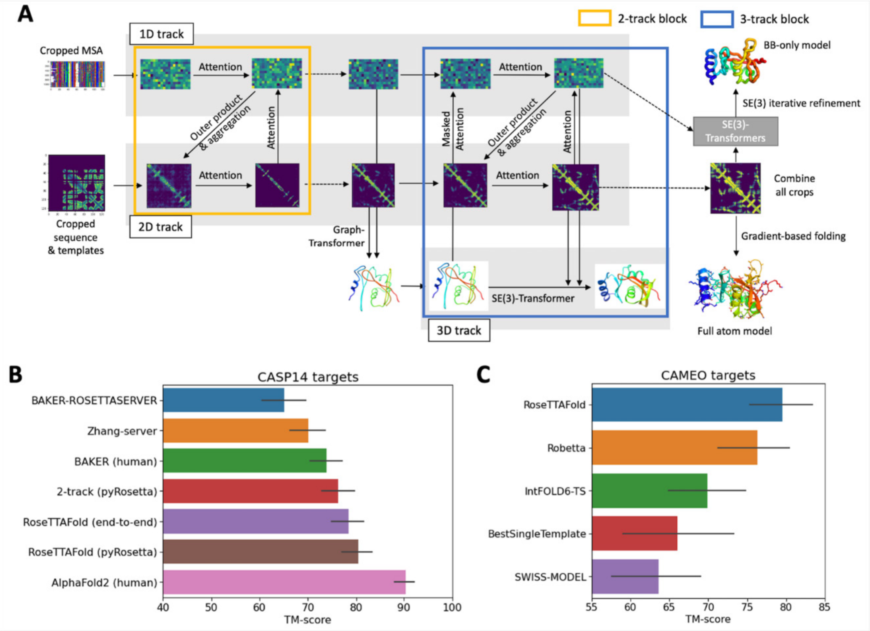
RESEARCH ARTICLES

Cite as: M. Baek *et al.*, *Science*
10.1126/science.abj8754 (2021).

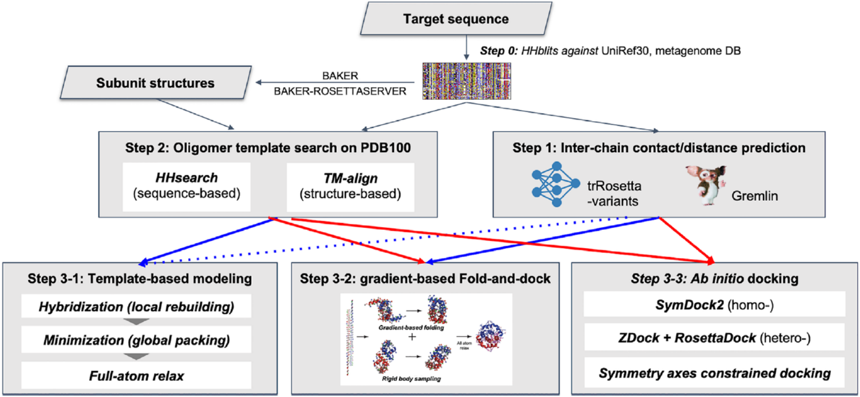
Accurate prediction of protein structures and interactions using a three-track neural network

Minkyung Baek[1,2], Frank DiMaio[1,2], Ivan Anishchenko[1,2], Justas Dauparas[1,2], Sergey Ovchinnikov[3,4], Gyu Rie Lee[1,2], Jue Wang[1,2], Qian Cong[5,6], Lisa N. Kinch[7], R. Dustin Schaeffer[6], Claudia Millán[8], Hahnbeom Park[1,2], Carson Adams[1,2], Caleb R. Glassman[9,10], Andy DeGiovanni[12], Jose H. Pereira[12], Andria V. Rodrigues[12], Alberdina A. van Dijk[13], Ana C. Ebrecht[13], Diederik J. Opperman[14], Theo Sagmeister[15], Christoph Buhlheller[15,16], Tea Pavkov-Keller[15,17], Manoj K. Rathinaswamy[18], Udit Dalwadi[19], Calvin K. Yip[19], John E. Burke[18], K. Christopher Garcia[9,10,11,20], Nick V. Grishin[6,21,7], Paul D. Adams[12,22], Randy J. Read[8], David Baker[1,2,23]*

publié le même jour que le Nature sur AlphaFold2 (15 juillet 2021) !
mais "que" 1120 citations sur google scholar à ce jour (6 oct 2022)


A
2-track block
3-track block
1D track
Cropped MSA
Attention
Outer product & aggregation
Attention
Attention
Masked Attention
Outer product & aggregation
Attention
Attention
BB-only model
SE(3) iterative refinement
SE(3)-Transformers
Combine all crops
Cropped sequence & templates
2D track
Graph-Transformer
SE(3)-Transformer
Gradient-based folding
3D track
Full atom model
B
CASP14 targets
BAKER-ROSETTASERVER
Zhang-server
BAKER (human)
2-track (pyRosetta)
RoseTTAFold (end-to-end)
RoseTTAFold (pyRosetta)
AlphaFold2 (human)
40
50
60
70
80
90
100
TM-score
C
CAMEO targets
RoseTTAFold
Robetta
IntFOLD6-TS
BestSingleTemplate
SWISS-MODEL
55
60
65
70
75
80
85
TM-score

Next frontier: Oligomères



Target sequence

Step 0: *HHblits against* UniRef30, metagenome DB

Subunit structures

BAKER

BAKER-ROSETTASERVER

Step 2: Oligomer template search on PDB100

HHsearch (sequence-based)

TM-align (structure-based)

Step 1: Inter-chain contact/distance prediction

trRosetta -variants

Gremlin

Step 3-1: Template-based modeling

Hybridization (local rebuilding)

Minimization (global packing)

Full-atom relax

Step 3-2: gradient-based Fold-and-dock

Step 3-3: Ab initio docking

SymDock2 (homo-)

ZDock + RosettaDock (hetero-)

Symmetry axes constrained docking

FIGURE 1 The oligomer structure modeling procedure used by the BAKER-experimental group

RosettaFold: github et serveur

`https://github.com/RosettaCommons/RoseTTAFold`

`https://robetta.bakerlab.org/`

voir aussi ici:

`https://www.rosettacommons.org/`

Comprendre Alpha Fold

Le blog de "Oxford Protein Informatics Group"

https://www.blopig.com/blog/2021/07/
alphafold-2-is-here-whats-behind-the-structure-predict

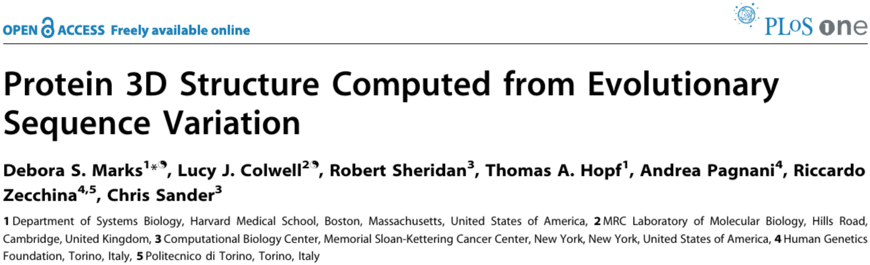
"... we have many new questions. What is the secret sauce before the news splash, and why is it so effective? Is it a piece of code that the average user can actually run? What are AlphaFold 2's shortcomings? And, most important of all, what will it mean for computational biology? And for all of us?"

Le blog de Mohammed AlQuraishi

https://moalquraishi.wordpress.com/2020/12/08/
alphafold2-casp14-it-feels-like-ones-child-has-left-ho
amp/

Mutations corrélées, MSA

Mutations corrélées, MSA



OPEN ACCESS Freely available online

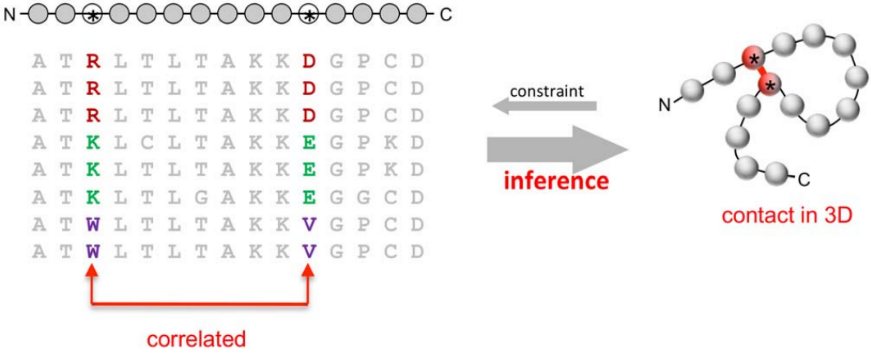
PLoS one

Protein 3D Structure Computed from Evolutionary Sequence Variation

Debora S. Marks[1*¶], Lucy J. Colwell[2¶], Robert Sheridan[3], Thomas A. Hopf[1], Andrea Pagnani[4], Riccardo Zecchina[4,5], Chris Sander[3]

1 Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, United States of America, **2** MRC Laboratory of Molecular Biology, Hills Road, Cambridge, United Kingdom, **3** Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, New York, United States of America, **4** Human Genetics Foundation, Torino, Italy, **5** Politecnico di Torino, Torino, Italy

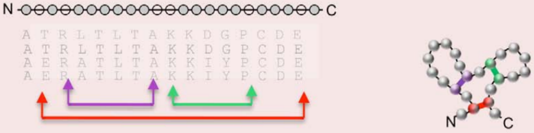
Principe



```
A T R L T L T A K K D G P C D
A T R L T L T A K K D G P C D
A T R L T L T A K K D G P C D
A T K L C L T A K K E G P K D
A T K L T L T A K K E G P K D
A T K L T L G A K K E G G C D
A T W L T L T A K K V G P C D
A T W L T L T A K K V G P C D
```

Protocole de 2011 sans deep-learning

Align evolutionary diverged sequences



A T R L T L T A K K D G P C D E
A T R L T L T A K K D G P C D E
A E R A T L T A K K I Y P C D E
A E R A T L T A K K I Y P C D E

Calculate covariance matrix for each pair of sequence positions for all pairs of amino acids (A,B)

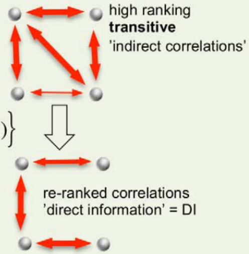
$$C_{ij}(A,B) = f_{ij}(A,B) - f_i(A)P_j(B)$$

$$C^{-1}_{ij}(A,B) = -e_{ij}(A,B)_{i\neq j}$$

Identify maximally informative pair couplings using **statistical model** of entire protein to infer residue-residue co-evolution

$$P^{Dir}_{ij}(A,B) = \frac{1}{Z}\exp\left\{e_{ij}(A,B) + \tilde{h}_i(A) + \tilde{h}_j(B)\right\}$$

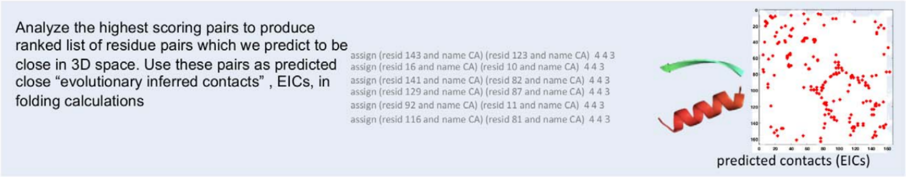
$$DI_{ij} = \sum_{A,B=1}^{q} P^{Dir}_{ij}(A,B)\ln\frac{P^{Dir}_{ij}(A,B)}{f_i(A)f_j(B)}$$



high ranking **transitive** 'indirect correlations'

re-ranked correlations 'direct information' = DI

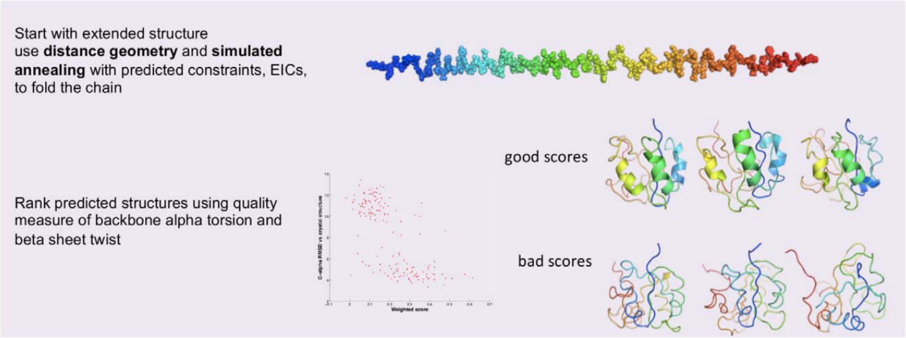
Protocole de 2011 sans deep-learning



Analyze the highest scoring pairs to produce ranked list of residue pairs which we predict to be close in 3D space. Use these pairs as predicted close "evolutionary inferred contacts" , EICs, in folding calculations

```
assign (resid 143 and name CA) (resid 123 and name CA)  4 4 3
assign (resid 16 and name CA) (resid 10 and name CA)  4 4 3
assign (resid 141 and name CA) (resid 82 and name CA)  4 4 3
assign (resid 129 and name CA) (resid 87 and name CA)  4 4 3
assign (resid 92 and name CA) (resid 11 and name CA)  4 4 3
assign (resid 116 and name CA) (resid 81 and name CA)  4 4 3
```

predicted contacts (EICs)



Start with extended structure use **distance geometry** and **simulated annealing** with predicted constraints, EICs, to fold the chain

Rank predicted structures using quality measure of backbone alpha torsion and beta sheet twist

good scores

bad scores

Utiliser Alpha Fold

google colab

https://colab.research.google.com/notebooks/intro.ipynb?utm_source=scs-index

https://www.youtube.com/watch?v=inN8seMm7UI

Exemple pour faire de la MD ou aussi AlphaFold2 + MD:

https://github.com/pablo-arantes/Making-it-rain

Alpha Fold avec google colab

une version amélioré d'alphaFold2:

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

ou l'original:

`https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb`

Alpha Fold en local

On peux l’installer aussi sur une machine en local, mais il faut qu’elle soit très puissante:

`https://github.com/deepmind/alphafold`

" The simplest way to run AlphaFold is using the provided Docker script. This was tested on Google Cloud with a machine using the nvidia-gpu-cloud-image with 12 vCPUs, 85 GB of RAM, a 100 GB boot disk, the databases on an additional 3 TB disk, and an A100 GPU."

Rien que la carte graphique A100 coûte le prix d’une petite voiture:

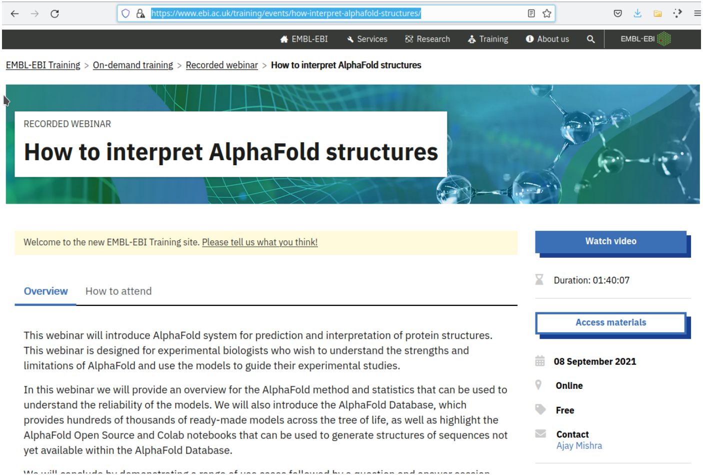
`https://fr.aliexpress.com/item/1005002408111365.html`

Pour aller plus loin: EMBL webinar

EMBL webinar "How to interpret AlphaFold structures"

8 sept 2021

`https://www.ebi.ac.uk/training/events/how-interpret-alphafold-structures/`


https://www.ebi.ac.uk/training/events/how-interpret-alphafold-structures/
EMBL-EBI
Services
Research
Training
About us
EMBL-EBI

EMBL-EBI Training > On-demand training > Recorded webinar > How to interpret AlphaFold structures

RECORDED WEBINAR

How to interpret AlphaFold structures

Welcome to the new EMBL-EBI Training site. Please tell us what you think!

Duration: 01:40:07

This webinar will introduce AlphaFold system for prediction and interpretation of protein structures. This webinar is designed for experimental biologists who wish to understand the strengths and limitations of AlphaFold and use the models to guide their experimental studies.

In this webinar we will provide an overview for the AlphaFold method and statistics that can be used to understand the reliability of the models. We will also introduce the AlphaFold Database, which provides hundreds of thousands of ready-made models across the tree of life, as well as highlight the AlphaFold Open Source and Colab notebooks that can be used to generate structures of sequences not yet available within the AlphaFold Database.

We will conclude by demonstrating a range of use cases followed by a question and answer session

08 September 2021

Online

Free

Contact
Ajay Mishra

DeepMind

Introduction to AlphaFold

Presenter: Kathryn Tunyasuvunakool
Research Scientist at DeepMind

How does it work? (the short version)

Article

Highly accurate protein structure prediction with AlphaFold

https://doi.org/10.1038/s41586-021-03819-2

Received: 11 May 2021

Accepted: 12 July 2021

Published online: 15 July 2021

Open access

Check for updates

John Jumper[1,4✉], Richard Evans[1,4], Alexander Pritzel[1,4], Tim Green[1,4], Michael Figurnov[1,4], Olaf Ronneberger[1,4], Kathryn Tunyasuvunakool[1,4], Russ Bates[1,4], Augustin Žídek[1,4], Anna Potapenko[1,4], Alex Bridgland[1,4], Clemens Meyer[1,4], Simon A. A. Kohl[1,4], Andrew J. Ballard[1,4], Andrew Cowie[1,4], Bernardino Romera-Paredes[1,4], Stanislav Nikolov[1,4], Rishub Jain[1,4], Jonas Adler[1], Trevor Back[1], Stig Petersen[1], David Reiman[1], Ellen Clancy[1], Michal Zielinski[1], Martin Steinegger[2,3], Michalina Pacholska[1], Tamas Berghammer[1], Sebastian Bodenstein[1], David Silver[1], Oriol Vinyals[1], Andrew W. Senior[1], Koray Kavukcuoglu[1], Pushmeet Kohli[1] & Demis Hassabis[1,4✉]

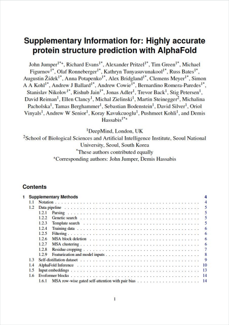
Proteins are essential to life, and understanding their structure can facilitate a mechanistic understanding of their function. Through an enormous experimental effort[1–4], the structures of around 100,000 unique proteins have been determined[5], but this represents a small fraction of the billions of known protein sequences[6,7]. Structural coverage is bottlenecked by the months to years of painstaking effort required to determine a single protein structure. Accurate computational approaches are needed to address this gap and to enable large-scale structural bioinformatics. Predicting the three-dimensional structure that a protein will adopt based solely on its amino acid sequence—the structure prediction component of the 'protein folding problem'[8]—has been an important open research problem for more than 50 years[9]. Despite recent progress[10–14], existing methods fall far short of atomic accuracy, especially when no homologous structure is available. Here we provide the first computational method that can regularly predict protein structures with atomic accuracy even in cases in which no similar structure is known. We validated an entirely redesigned version of our neural network-based model, AlphaFold, in the challenging 14th Critical Assessment of protein Structure Prediction (CASP14)[15], demonstrating accuracy competitive with experimental structures in a majority of cases and greatly outperforming other methods. Underpinning the latest version of AlphaFold is a novel machine learning approach that incorporates physical and biological knowledge about protein structure, leveraging multi-sequence alignments, into the design of the deep learning algorithm.

The development of computational methods to predict three-dimensional (3D) protein structures from the protein sequence has proceeded along two complementary paths that focus on either the physical interactions or the evolutionary history. The physical interaction programme heavily integrates our understanding of molecular driving forces into either thermodynamic or kinetic simulation of protein physics[16] or statistical approximations thereof[17]. Although theoretically very appealing, this approach has proved highly challenging for even moderate-sized proteins due to the computational intractability of molecular simulation, the context dependence of protein stability and the difficulty of producing sufficiently accurate models of protein physics. The evolutionary programme has provided an alternative in recent years, in which the constraints on protein structure are derived from bioinformatics analysis of the evolutionary history of proteins, homology to solved structures[18,19] and pairwise evolutionary correlations[20–24]. This bioinformatics approach has benefited greatly from the steady growth of experimental protein structures deposited in the Protein Data Bank (PDB)[5], the explosion of genomic sequencing and the rapid development of deep learning techniques to interpret these correlations. Despite these advances, contemporary physical and evolutionary-history-based approaches produce predictions that are far short of experimental accuracy in the majority of cases in which a close homologue has not been solved experimentally and this has limited their utility for many biological applications.

In this study, we develop the first, to our knowledge, computational approach capable of predicting protein structures to near experimental accuracy in a majority of cases. The neural network AlphaFold that we developed was entered into the CASP14 assessment (May–July 2020; entered under the team name 'AlphaFold2' and a completely different model from our CASP13 AlphaFold system[10]). The CASP assessment is carried out biennially using recently solved structures that have not been deposited in the PDB or publicly disclosed so that it is a blind test

[1]DeepMind, London, UK. [2]School of Biological Sciences, Seoul National University, Seoul, South Korea. [3]Artificial Intelligence Institute, Seoul National University, Seoul, South Korea. [4]These authors contributed equally: John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Demis Hassabis. ✉e-mail: jumper@deepmind.com; dhcontact@deepmind.com

Nature | Vol 596 | 26 August 2021 | 583



Supplementary Information for: Highly accurate protein structure prediction with AlphaFold

John Jumper[1*+], Richard Evans[1*], Alexander Pritzel[1*], Tim Green[1*], Michael Figurnov[1*], Olaf Ronneberger[1*], Kathryn Tunyasuvunakool[1*], Russ Bates[1*], Augustin Žídek[1*], Anna Potapenko[1*], Alex Bridgland[1*], Clemens Meyer[1*], Simon A A Kohl[1*], Andrew J Ballard[1*], Andrew Cowie[1*], Bernardino Romera-Paredes[1*], Stanislav Nikolov[1*], Rishub Jain[1*], Jonas Adler[1], Trevor Back[1], Stig Petersen[1], David Reiman[1], Ellen Clancy[1], Michal Zielinski[1], Martin Steinegger[2], Michalina Pacholska[1], Tamas Berghammer[1], Sebastian Bodenstein[1], David Silver[1], Oriol Vinyals[1], Andrew W Senior[1], Koray Kavukcuoglu[1], Pushmeet Kohli[1], and Demis Hassabis[1*+]

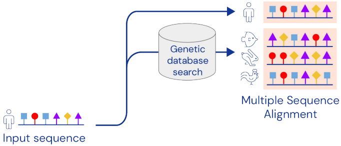
[1]DeepMind, London, UK
[2]School of Biological Sciences and Artificial Intelligence Institute, Seoul National University, Seoul, South Korea
[*]These authors contributed equally
[+]Corresponding authors: John Jumper, Demis Hassabis

Contents

1

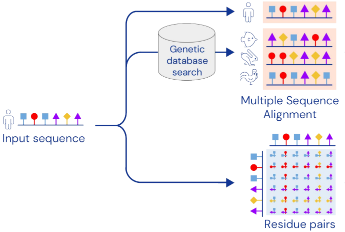
See Jumper et al. 2021 (especially the SI) for details

Inputs



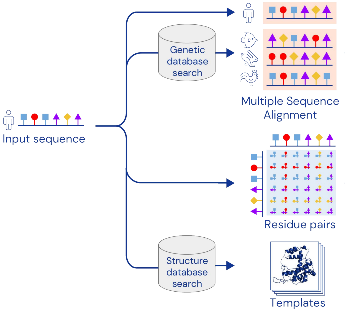
A key AlphaFold input is the MSA, containing sequences evolutionarily related to the target. Related sequences are found using standard tools and public databases.

Inputs



The input sequence is used to create an array of numbers representing all residue pairs.

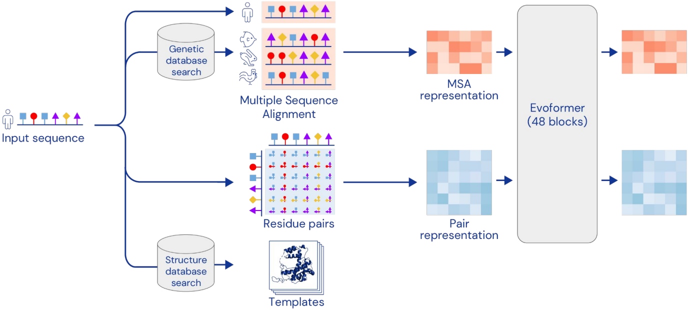
Inputs



AlphaFold can also use template structures from the PDB, found using standard tools.
However, it often produces accurate predictions without a template.

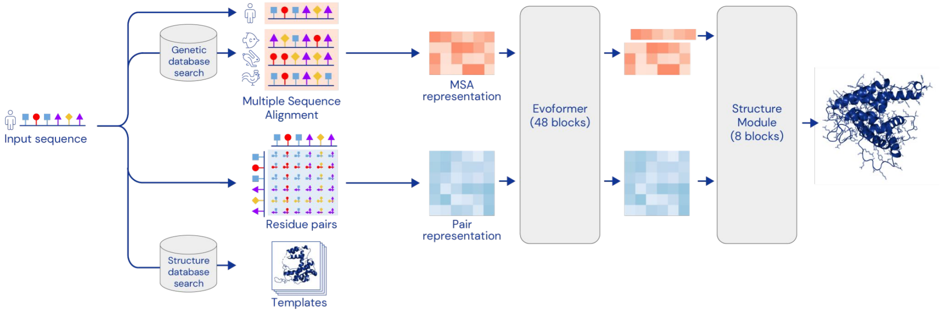
Network

Private & Confidential



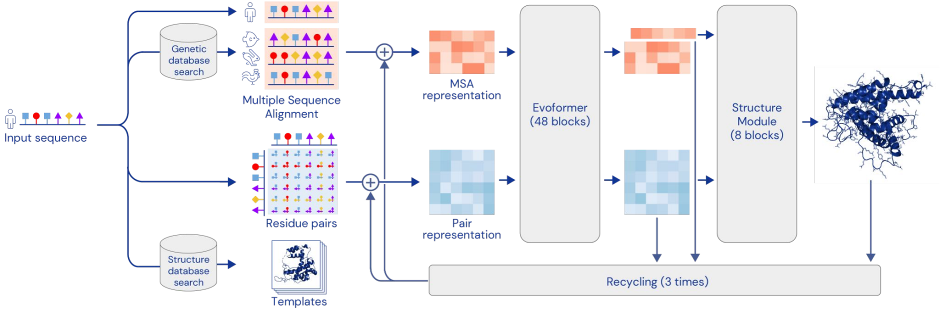
The Evoformer blocks extract information about the relationship between residues.
The MSA representation can update the pair representation and vice versa.

Network



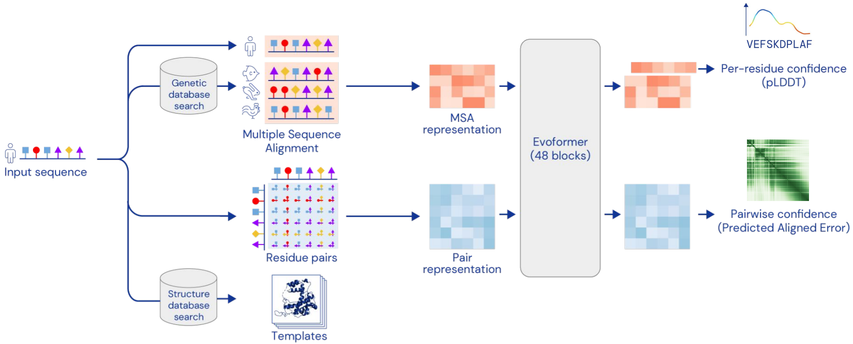
The Structure Module predicts a rotation + translation to place each residue.
A small network predicts side chain chi angles. The final structure is run through a relaxation process.

Network



Feeding certain outputs back through the network again improves performance

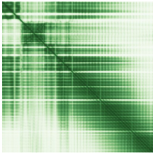
Other outputs



As well as a predicted structure, AlphaFold produces two confidence estimates

Private & Confidential

Interpreting predictions



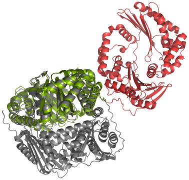
The short version: use **both** confidence metrics

Predicted LDDT: definition

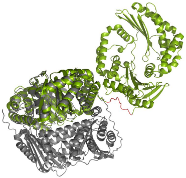
AlphaFold’s per-residue prediction of its lDDT-Cα score*

Roughly, lDDT measures the percentage of correctly predicted interatomic distances, not how well the predicted and true structures can be superimposed.

It rewards **locally correct** structures, and **getting individual domains right**.
pLDDT behaves similarly, as a measure of **local confidence**



Alignment-based metric



lDDT

*Mariani et al. 2013

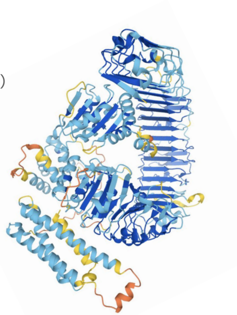
Predicted LDDT: format

pLDDT ranges from 0 to 100 (100 is most confident)

We use a consistent “confidence bands” color scheme when displaying predictions

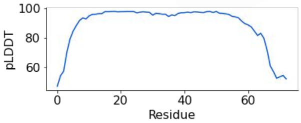
A pLDDT plot is also displayed by some of our tools

- ⬛ Very high (>90)
- 🟦 Confident (70-90)
- 🟨 Low (50-70)
- 🟧 Very low (<50)



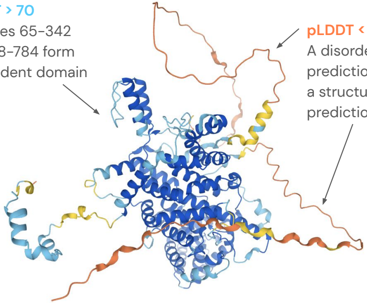
Prediction files always contain pLDDT in the B-factors
Therefore a **higher** B-factor is better!

```
MODEL        0
ATOM      1  N   MET A   1      -9.212  -5.798  33.490  1.00 63.75           N
ATOM      2  CA  MET A   1     -10.075  -6.130  32.322  1.00 63.75           C
ATOM      3  C   MET A   1     -11.469  -6.615  32.714  1.00 63.75           C
ATOM      4  CB  MET A   1      -9.419  -7.112  31.341  1.00 63.75           C
ATOM      5  O   MET A   1     -12.429  -6.075  32.184  1.00 63.75           O
ATOM      6  CG  MET A   1      -8.311  -6.411  30.547  1.00 63.75           C
ATOM      7  SD  MET A   1      -7.766  -7.280  29.061  1.00 63.75           S
ATOM      8  CE  MET A   1      -7.045  -8.751  29.798  1.00 63.75           C
ATOM      9  N   ALA A   2     -11.624  -7.579  33.634  1.00 66.38           N
ATOM     10  CA  ALA A   2     -12.951  -8.096  34.007  1.00 66.38           C
```



Predicted LDDT: usage

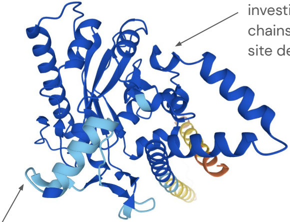
Identifying domains & possible disordered regions



pLDDT > 70
Residues 65-342 and 418-784 form a confident domain

pLDDT < 50
A disorder prediction not a structure prediction

Assessing confidence within a domain

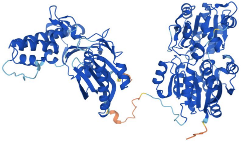


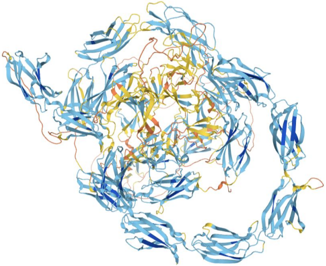
pLDDT > 90
Reasonable to investigate side chains / active site details

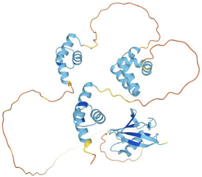
pLDDT > 70
Lower confidence on these specific parts

Predicted LDDT: pitfalls

High pLDDT on all domains does **not** imply AlphaFold is confident of their relative positions







Predicted Aligned Error: definition

AlphaFold's prediction of its position error at residue *x*,
if the predicted and the true structures were aligned on residue *y*

PAE aims to measure confidence in the **relative positions** of **pairs of residues**

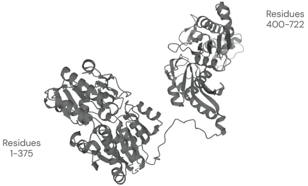
Mainly used to assess relative domain positions, but applicable whenever pairwise confidence is relevant

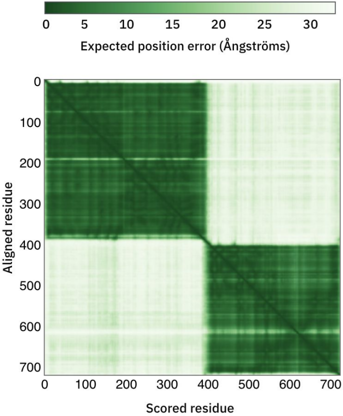
Predicted Aligned Error: format

PAE is displayed as a 2D plot.

Suppose residue *y* were aligned to the true structure
and we measured the position error at residue *x*
The color at (x, y) is AlphaFold's prediction of that error

In this case the squares correspond to two domains



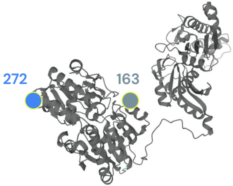


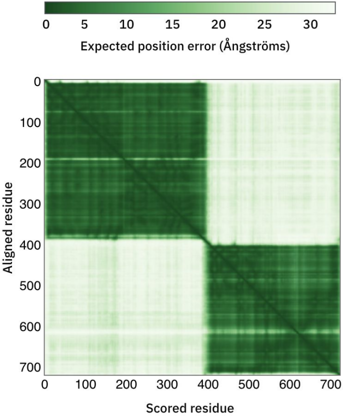
Predicted Aligned Error: format

PAE is displayed as a 2D plot.

Suppose residue *y* were aligned to the true structure
And we measured the position error at residue *x*
The color at (*x*, *y*) is AlphaFold's prediction of that error

AlphaFold is confident in relative positions within each domain...



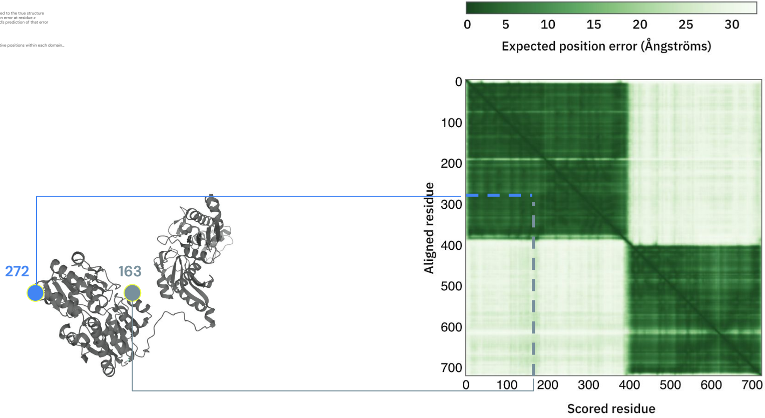


Predicted Aligned Error: format

PAE is displayed as a 2D plot.

Suppose residue *y* were aligned to the true structure
And we measured the position error at residue *x*
The color at (*x*, *y*) is AlphaFold's prediction of that error

AlphaFold is confident in relative positions within each domain...

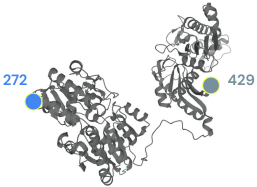


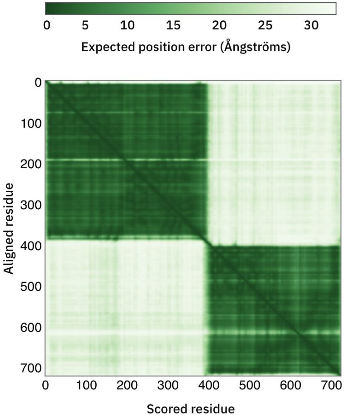
Predicted Aligned Error: format

PAE is displayed as a 2D plot.

Suppose residue *y* were aligned to the true structure
And we measured the position error at residue *x*
The color at (*x*, *y*) is AlphaFold's prediction of that error

...but not between domains.



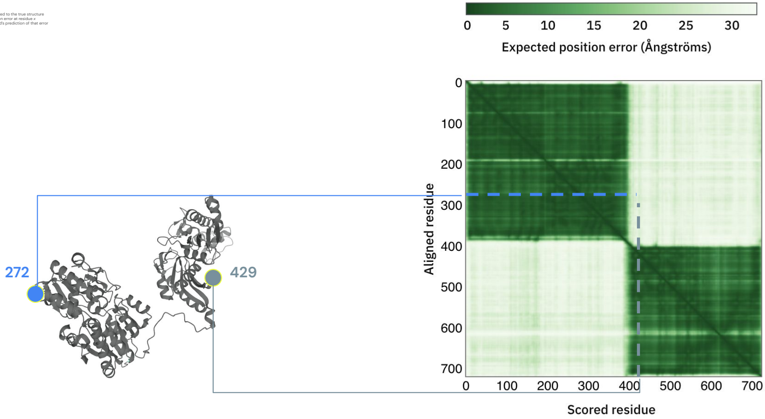


Predicted Aligned Error: format

PAE is displayed as a 2D plot.

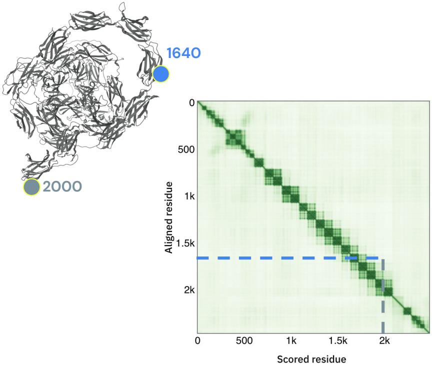
Suppose residue *y* were aligned to the true structure
And we measured the position error at residue *x*
The color at (*x*, *y*) is AlphaFold's prediction of that error

...but not between domains.

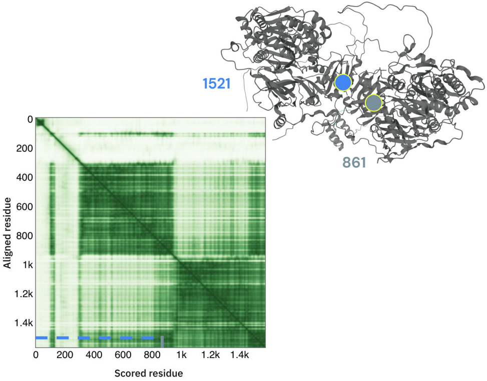


Predicted Aligned Error: usage

No confidence in relative domain positions



Predicted domain packing



Things to be aware of

Uncertain domain placement

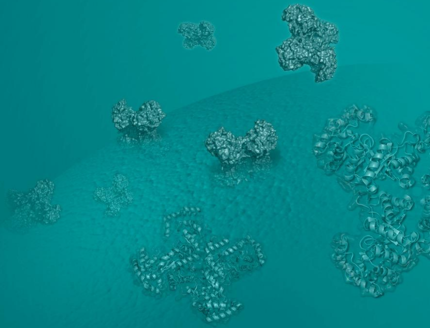
- If AlphaFold is uncertain, it won’t necessarily place domains sensibly relative to each other
  - Membrane proteins won’t leave space for the cell membrane
  - Clashes can occur

Complexes

- For proteins that exist in complex, AlphaFold is missing context about their binding partners
  - Heteromers more problematic than homomers
  - Worst case: the protein is flexible in isolation

- Some have had success predicting complexes by joining 2 sequences with a linker
  - We think it is possible to extend the ideas in AlphaFold to complexes
  - However, this linker setup remains to be benchmarked

AlphaFold database



Sameer Velankar

EMBL-EBI

DeepMind

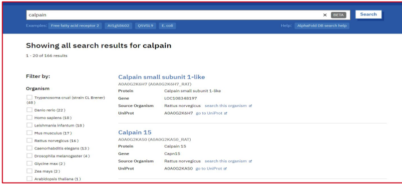
EMBL-EBI

AlphaFold Database

- ~365K predicted structures for proteins from 21 model organisms
- For the organisms currently covered, predicted structures available for sequences in the UniProt reference proteome that are between 16 and 2700 amino acids long and contain only standard amino acids
- Only one prediction (out of 5 independent predictions) is made available in the release
- Accession – AF-P12345-F[1-N]
- Files
  - AF-P12345-F[1-N]-model_V[1-N].[pdb, mmcif]
  - AF-P12345-F[1-N]-predicted-aligned-error_V[1-N].json

AlphaFold web pages

- Basic search system
- Allows search using UniProt accession, UniProt id, protein name, gene name and organism
- Clear indication that the structure shown is a prediction
- Allow easy download of structure data
- Basic information about protein
- Clearly indicates if there are experimental structures available
- Display residue-quality information in 3D viewer (pLDDT – predicted Local Distance Difference Test)
- Predicted Aligned Error (PAE viewer)



Showing all search results for calpain

1 - 20 of 166 results

Filter by:

Organism

- □ Trypanosoma cruzi (strain CL Brener) (48)
- □ Danio rerio (22)
- □ Homo sapiens (18)
- □ Leishmania infantum (18)
- □ Mus musculus (17)
- □ Rattus norvegicus (16)
- □ Caenorhabditis elegans (13)
- □ Drosophila melanogaster (4)
- □ Glycine max (2)
- □ Zea mays (2)
- □ Arabidopsis thaliana (1)

Calpain small subunit 1-like

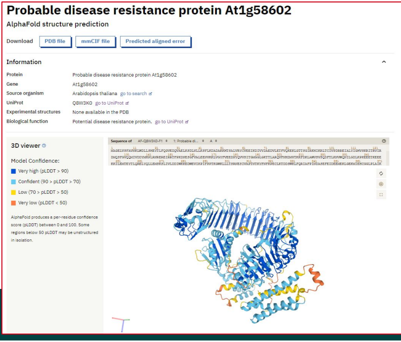
A0A0G2K6H7 (A0A0G2K6H7_RAT)

Protein	Calpain small subunit 1-like
Gene	LOC108348197
Source Organism	Rattus norvegicus search this organism
UniProt	A0A0G2K6H7 go to UniProt

Calpain 15

A0A0G2KAS0 (A0A0G2KAS0_RAT)

Protein	Calpain 15
Gene	Capn15
Source Organism	Rattus norvegicus search this organism
UniProt	A0A0G2KAS0 go to UniProt



Probable disease resistance protein At1g58602

AlphaFold structure prediction

Information

Protein	Probable disease resistance protein At1g58602
Gene	At1g58602
Source organism	Arabidopsis thaliana go to search
UniProt	Q8W3K0 go to UniProt
Experimental structures	None available in the PDB
Biological function	Potential disease resistance protein. go to UniProt

3D viewer

Model Confidence:

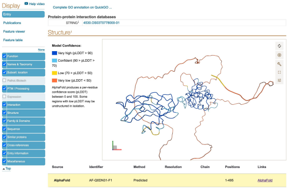
- 🟦 Very high (pLDDT > 90)
- 🟦 Confident (90 > pLDDT > 70)
- 🟨 Low (70 > pLDDT > 50)
- 🟧 Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

DeepMind

Models available across EBI resources

UniProt



Display Help video

Entry
Publications
Feature viewer
Feature table

None

☑ Function
☑ Names & Taxonomy
☑ Subcell. location
□ Pathol./Biotech
☑ PTM / Processing
□ Expression
☑ Interaction
☑ Structure
☑ Family & Domains
☑ Sequence
☑ Similar proteins
☑ Cross-references
☑ Entry information
☑ Miscellaneous

▲ Top

Complete GO annotation on QuickGO ...

Protein-protein interaction databases

STRING[i]	4530.OS03T0778000-01

Structure[i]

Model Confidence:

🟦 Very high (pLDDT > 90)

🟦 Confident (90 > pLDDT > 70)

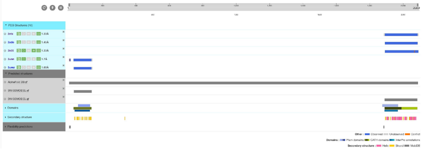
🟨 Low (70 > pLDDT > 50)

🟧 Very low (pLDDT < 50)

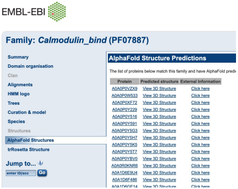
AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions with low pLDDT may be unstructured in isolation.

Source	Identifier	Method	Resolution	Chain	Positions	Links
AlphaFold	AF-Q0DN31-F1	Predicted			1-495	AlphaFold

PDBe-KB



Pfam



EMBL-EBI

Family: *Calmodulin_bind* (PF07887)

Summary
Domain organisation
Clan
Alignments
HMM logo
Trees
Curation & model
Species
Structures
AlphaFold Structures
trRosetta Structure

Jump to...

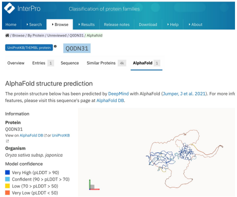
enter ID/acc Go

AlphaFold Structure Predictions

The list of proteins below match this family and have AlphaFold predi

Protein	Predicted structure	External Information
A0A0P0VZX9	View 3D Structure	Click here
A0A0P0W533	View 3D Structure	Click here
A0A0P0XF72	View 3D Structure	Click here
A0A0P0Y229	View 3D Structure	Click here
A0A0P0Y516	View 3D Structure	Click here
A0A0P0Y591	View 3D Structure	Click here
A0A0P0Y5G3	View 3D Structure	Click here
A0A0P0Y5H7	View 3D Structure	Click here
A0A0P0Y5K5	View 3D Structure	Click here
A0A0P0Y5T7	View 3D Structure	Click here
A0A0P0YBV0	View 3D Structure	Click here
A0A0R0KNR6	View 3D Structure	Click here
A0A1D6E9U4	View 3D Structure	Click here
A0A1D6F486	View 3D Structure	Click here
A0A1D6GE14	View 3D Structure	Click here

InterPro



InterPro Classification of protein families

Home ‣ Search ‣ Browse ‣ Results Release notes Download ‣ Help ‣ About

/ Browse / By Protein / Unreviewed / Q0DN31 / Alphafold

UniProtKB/TrEMBL protein Q0DN31

Overview Entries 1 Sequence Similar Proteins 4k **AlphaFold** 1

AlphaFold structure prediction

The protein structure below has been predicted by DeepMind with AlphaFold (Jumper, J et al. 2021). For more info features, please visit this sequence's page at AlphaFold DB.

Information

Protein
Q0DN31
View on AlphaFold DB or UniProtKB

Organism
Oryza sativa subsp. japonica

Model confidence

🟦 Very High (pLDDT > 90)
🟦 Confident (90 > pLDDT > 70)
🟨 Low (70 > pLDDT > 50)
🟧 Very Low (pLDDT < 50)

AlphaFold Database – limitations

- Information on complexes with other proteins, nucleic acids (DNA or RNA) or ligands. In some cases, the single-chain prediction may correspond to the structure adopted in a complex. The missing context from surrounding molecules may lead to an uninformative prediction
  - AlphaFold does not make any predictions about any of the non-protein components such as cofactors, metals, ligands including drug-like molecules, ions, carbohydrates and other post-translational modifications
- Protein dynamics - AlphaFold will usually only produce one of multiple conformations
- AlphaFold has not been trained or validated for predicting the effect of mutations
- May (or may not) lead to hypotheses about protein function – any hypotheses have to be tested by further experimentation

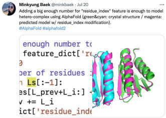
What’s next – under discussion

- Remove signal peptides from predictions
- Making 5 independent predictions available for each protein
- Additional metadata
  - MSA – need to consider data size
  - information on templates
  - quality criteria e.g. predicted TM score
- Updating database to UniRef90 dataset (~130 million structures)

Impact of AlphaFold database on life science research

Structural bioinformatics – (structure/function)

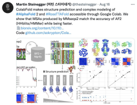
- Predicting complexes between macromolecules
  - Homo- and Hetero- Protein-protein; Protein-nucleic acid complexes
  - Intrinsically disordered proteins
- Provide information on protein dynamics
  - Relevant conformational states
- Functionally important residues
  - Impact of mutation; Binding sites; Conformationally important residues
  - Interfaces
- Ligand prediction – What binds?
  - What might bind in a pocket



Minkyung Baek @minkbaek · Jul 20

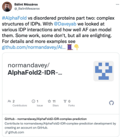
Adding a big enough number for "residue_index" feature is enough to model hetero-complex using AlphaFold (green&cyan: crystal structure / magenta: predicted model w/ residue_index modification).
#AlphaFold #alphafold2

```
 enough number to
 feature_dict['re
0
ber of residues i
n Ls[:-1]:
es[L_prev+L_i:] +
v += L_i
ict['residue_inde
```



Martin Steinegger (마틴 스타이네거) @thesteinegger · Aug 16

ColabFold makes structure prediction and complex modeling of #AlphaFold 2 and #RoseTTAFold accessible through Google Colab. We show that MSAs produced by MMseqs2 match the accuracy of AF2 (HHblits/HMMer) while being faster.
biorxiv.org/content/10.110...
Code github.com/sokrypton/Cola...



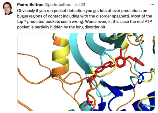
Bálint Mészáros
@_BalintMeszaros

#AlphaFold vs disordered proteins part two: complex structures of IDPs. With @Daveyab we looked at various IDP interactions and how well AF can model them. Some work, some don't, but all are enlighting. For details and more examples see github.com/normandavey/Al... 🧵👇

normandavey/
AlphaFold2-IDR-...

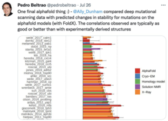
2 Contributors · 1 Issue · 11 Stars · 0 Forks

GitHub - normandavey/AlphaFold2-IDR-complex-prediction
Contribute to normandavey/AlphaFold2-IDR-complex-prediction development by creating an account on GitHub.
github.com



Pedro Beltrao @pedrobeltrao · Jul 23

Obviously if you run pocket detection you get lots of over predictions on bugus regions of contact including with the disorder spaghetti. Most of the top 7 predicted pockets seem wrong. Worse even, in this case the real ATP pocket is partially hidden by the long disorder bit.

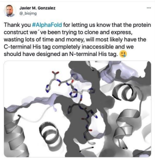


Pedro Beltrao @pedrobeltrao · Jul 26

One final alphafold thing :) - @Ally_Dunham compared deep mutational scanning data with predicted changes in stability for mutations on the alphafold models (with FoldX). The correlations observed are typically as good or better than with experimentally derived structures

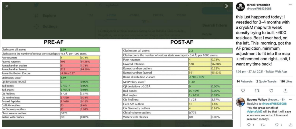
Structural biology

- Accelerating structure studies
  - Improved construct design
  - Starting model for structure determination
  - Fitting models in low resolution EM maps
  - Time resolved studies to understand mechanism



Javier M. Gonzalez
@_biojmg

Thank you #AlphaFold for letting us know that the protein construct we´ve been trying to clone and express, wasting lots of time and money, will most likely have the C-terminal His tag completely inaccessible and we should have designed an N-terminal His tag. 🙂



PRE-AF

Clashscore, all atoms:	2.39	
Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Poor rotamers	4	0.74%
Favored rotamers	496	91.18%
Ramachandran outliers	11	1.78%
Ramachandran favored	522	84.60%
Rama distribution Z-score	-3.98 ± 0.27	
MolProbity score^	1.69	
Cβ deviations >0.25Å	0	0.00%
Bad bonds:	0 / 5017	0.00%
Bad angles:	2 / 6772	0.03%
Cis Prolines:	1 / 28	3.57%
Cis nonProlines:	2 / 590	0.34%
Twisted Peptides:	1 / 618	0.16%
CaBLAM outliers	53	8.6%
CA Geometry outliers	12	1.95%
Chiral volume outliers	0/778	
Waters with clashes	0/0	0.00%

POST-AF

Clashscore, all atoms:	2.2	
Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Poor rotamers	4	0.73%
Favored rotamers	528	96.88%
Ramachandran outliers	2	0.32%
Ramachandran favored	591	95.63%
Rama distribution Z-score	-1.90 ± 0.27	
MolProbity score^	1.30	
Cβ deviations >0.25Å	0	0.00%
Bad bonds:	0 / 4993	0.00%
Bad angles:	1 / 6732	0.01%
Cis Prolines:	1 / 28	3.57%
CaBLAM outliers	10	1.6%
CA Geometry outliers	4	0.65%
Chiral volume outliers	0/776	
Waters with clashes	0/0	0.00%

Israel Fernandez
@IsraelF96135088

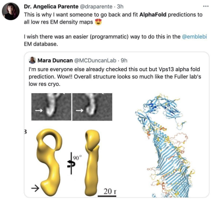
this just happened today: I wrestled for 3-4 months with a cryoEM map with weak density trying to built ~600 residues. Best I ever had, on the left. This morning, got the AF prediction, minimum adjustment to fit into the map + refinement and right...shit, I want my time back!

1:06 pm · 27 Jul 2021 · Twitter Web App

14 Retweets 2 Quote Tweets 110 Likes

Eugene Valkov @euge... · 8h
Replying to @IsraelF96135088
Yes, the great benefit of #alphafold2 will be that it will save enormous amounts of time (and research money).

1 6



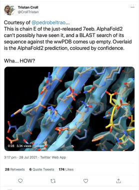
Dr. Angelica Parente @draparente · 3h

This is why I want someone to go back and fit **AlphaFold** predictions to all low res EM density maps 😍

I wish there was an easier (programmatic) way to do this in the @emblebi EM database.

Mara Duncan @MCDuncanLab · 9h

I'm sure everyone else already checked this out but Vps13 alpha fold prediction. Wow!! Overall structure looks so much like the Fuller lab's low res cryo.



Tristan Croll
@CrollTristan

Courtesy of @pedrobeltrao...
This is chain E of the just-released 7eeb. AlphaFold2 can't possibly have seen it, and a BLAST search of its sequence against the wwPDB comes up empty. Overlaid is the AlphaFold2 prediction, coloured by confidence.

Wha... HOW?

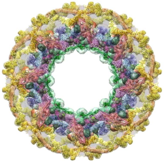
3:17 pm · 28 Jul 2021 · Twitter Web App

28 Retweets 6 Quote Tweets 174 Likes

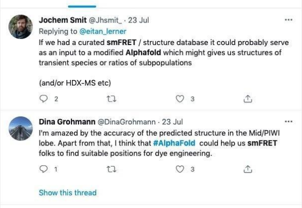
Structural biology

- Integrative/hybrid methods
  - Models for individual components

- Combination of sparse experimental data and predicted model may lead to actionable data to test hypothesis
  - Chemical foot printing
  - Hydrogen-Deuterium exchange
  - smFRET - Single molecule fluorescence resonance energy transfer



I/H Methods Structures
552-protein yeast Nuclear Pore Complex
Kim et al. (2018) *Nature 555*, 475-82
PDBDEV_00000010; PDBDEV_00000011; PDBDEV_00000012



Jochem Smit @Jhsmit_ · 23 Jul
Replying to @eitan_lerner
If we had a curated **smFRET** / structure database it could probably serve as an input to a modified **Alphafold** which might gives us structures of transient species or ratios of subpopulations

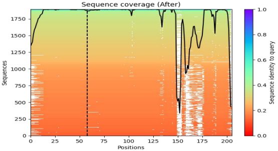
(and/or HDX-MS etc)

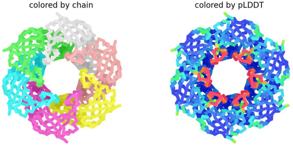
Dina Grohmann @DinaGrohmann · 23 Jul
I'm amazed by the accuracy of the predicted structure in the Mid/PIWI lobe. Apart from that, I think that #AlphaFold could help us **smFRET** folks to find suitable positions for dye engineering.

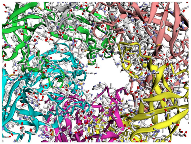
ColabFold

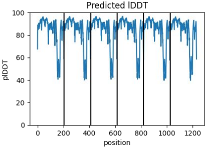
Making Protein folding accessible to all via Google Colab

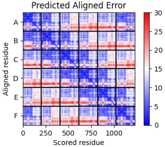
(and the unintended uses of AlphaFold)







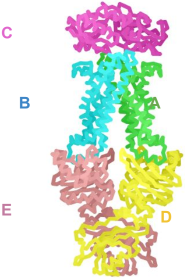


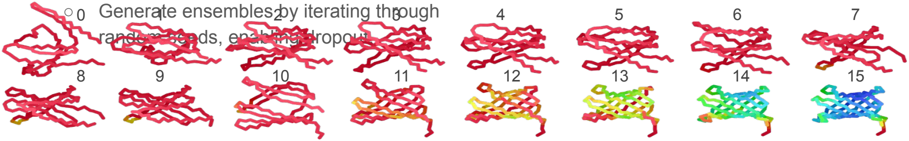


github.com/sokrypton/ColabFold

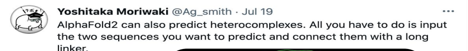
ColabFold - Advanced options

- Modify MSA input
  - Custom or MMseqs2 (much faster)
  - Trim
- **Complexes**
  - **Homo-oligomers**
  - **Hetero-oligomers**
- Fine control
  - Number of recycles
- Sample (Output more than 5 models)
  - Generate ensembles by iterating through random seeds, enabling dropout



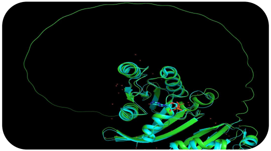


Can predict protein-protein/peptide interactions

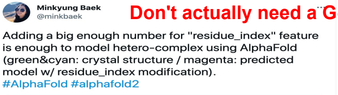


Yoshitaka Moriwaki @Ag_smith · Jul 19

AlphaFold2 can also predict heterocomplexes. All you have to do is input the two sequences you want to predict and connect them with a long linker.



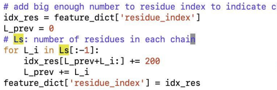
G-linker!



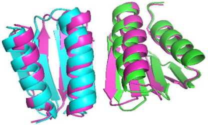
Minkyung Baek
@minkbaek

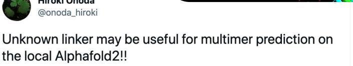
Don't actually need a G-linker!

Adding a big enough number for "residue_index" feature is enough to model hetero-complex using AlphaFold (green&cyan: crystal structure / magenta: predicted model w/ residue_index modification).
#AlphaFold #alphafold2



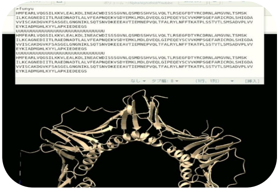
```
# add big enough number to residue index to indicate chain breaks
idx_res = feature_dict['residue_index']
L_prev = 0
# Ls: number of residues in each chain
for L_i in Ls[:-1]:
    idx_res[L_prev+L_i:] += 200
    L_prev += L_i
feature_dict['residue_index'] = idx_res
```





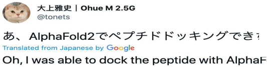
Hiroki Onoda
@onoda_hiroki

Unknown linker may be useful for multimer prediction on the local Alphafold2!!



```
>funyu
HMFEARLVQGSILKKVLEALKDLINEACWDISSSGVNLQSMDSSHVSLVQLTLRSEGFDTYRCDRNLAMGVNLTSMSK
ILKCAGNEDIITLRAEDNADTLALVFEAPNQEKVSDYEMKLMDLDVEQLGIPEQEYSCVVKMPSGEFARICRDLSHIGDA
VVISCAKDGVKFSASGELGNGNIKLSQTSNVDKEEEAVTIEMNEPVQLTFALRYLNFFTKATPLSSTVTLSMSADVPLVV
EYKIADMGHLKYYLAPKIEDEEGS
UUUUUUUUUUUUUUUUUUUUUUUUUUUUUU
HMFEARLVQGSILKKVLEALKDLINEACWDISSSGVNLQSMDSSHVSLVQLTLRSEGFDTYRCDRNLAMGVNLTSMSK
ILKCAGNEDIITLRAEDNADTLALVFEAPNQEKVSDYEMKLMDLDVEQLGIPEQEYSCVVKMPSGEFARICRDLSHIGDA
VVISCAKDGVKFSASGELGNGNIKLSQTSNVDKEEEAVTIEMNEPVQLTFALRYLNFFTKATPLSSTVTLSMSADVPLVV
EYKIADMGHLKYYLAPKIEDEEGS
UUUUUUUUUUUUUUUUUUUUUUUUUUUUUU
HMFEARLVQGSILKKVLEALKDLINEACWDISSSGVNLQSMDSSHVSLVQLTLRSEGFDTYRCDRNLAMGVNLTSMSK
ILKCAGNEDIITLRAEDNADTLALVFEAPNQEKVSDYEMKLMDLDVEQLGIPEQEYSCVVKMPSGEFARICRDLSHIGDA
VVISCAKDGVKFSASGELGNGNIKLSQTSNVDKEEEAVTIEMNEPVQLTFALRYLNFFTKATPLSSTVTLSMSADVPLVV
EYKIADMGHLKYYLAPKIEDEEGS
```

UNK-linker!

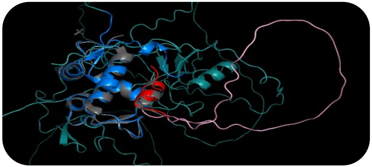


大上雅史 | Ohue M 2.5G
@tonets

あ、AlphaFold2でペプチドドッキングでき

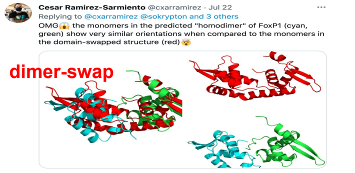
Translated from Japanese by Google

Oh, I was able to dock the peptide with AlphaFold2



Protein-peptide interaction

Can predict protein-protein/peptide interactions

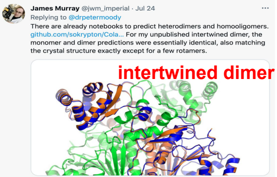


Cesar Ramirez-Sarmiento @cxarramirez · Jul 22

Replying to @cxarramirez @sokrypton and 3 others

OMG😱 the monomers in the predicted "homodimer" of FoxP1 (cyan, green) show very similar orientations when compared to the monomers in the domain-swapped structure (red) 🤯

dimer-swap

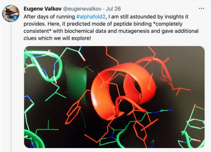


James Murray @jwm_imperial · Jul 24

Replying to @drpetermoody

There are already notebooks to predict heterodimers and homooligomers. github.com/sokrypton/Cola... For my unpublished intertwined dimer, the monomer and dimer predictions were essentially identical, also matching the crystal structure exactly except for a few rotamers.

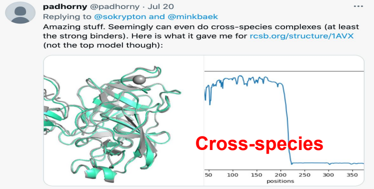
intertwined dimer



Eugene Valkov @eugenevalkov · Jul 26

After days of running #alphafold2, I am still astounded by insights it provides. Here, it predicted mode of peptide binding *completely consistent* with biochemical data and mutagenesis and gave additional clues which we will explore!

consistent w/ biochem data



padhorny @padhorny · Jul 20

Replying to @sokrypton and @minkbaek

Amazing stuff. Seemingly can even do cross-species complexes (at least the strong binders). Here is what it gave me for rcsb.org/structure/1AVX (not the top model though):

Cross-species

Preprints rolling in...

Can AlphaFold2 predict protein-peptide complex structures accurately?

Junsu Ko, Juyong Lee

bioRxiv 2021.07.27.453972; doi: https://doi.org/10.1101/2021.07.27.453972

Harnessing protein folding neural networks for peptide-protein docking

Tomer Tsaban, Julia Varga, Orly Avraham, Ziv Ben-Aharon, Alisa Khramushin, Ora Schueler-Furman

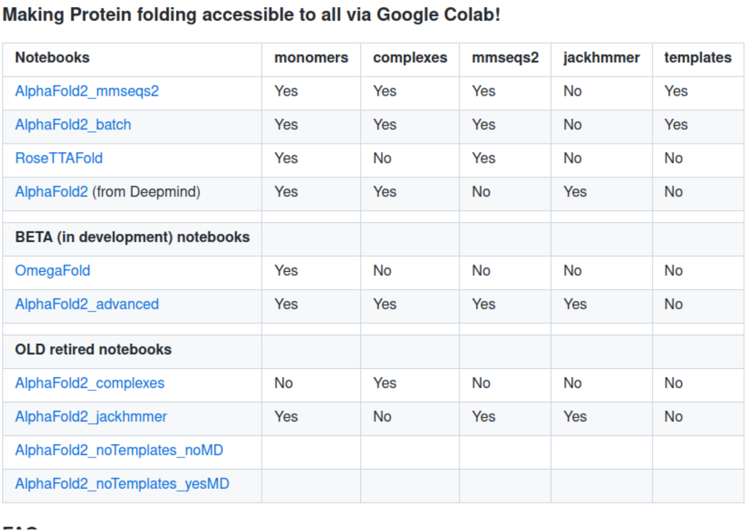
bioRxiv 2021.08.01.454656; doi: https://doi.org/10.1101/2021.08.01.454656

Improved Docking of Protein Models by a Combination of Alphafold2 and ClusPro

Usman Ghani, Israel Desta, Akhil Jindal, Omeir Khan, George Jones, Sergey Kotelnikov, Dzmitry Padhorny, Sandor Vajda, Dima Kozakov

bioRxiv 2021.09.07.459290; doi: https://doi.org/10.1101/2021.09.07.459290

ColabFold se diversifie



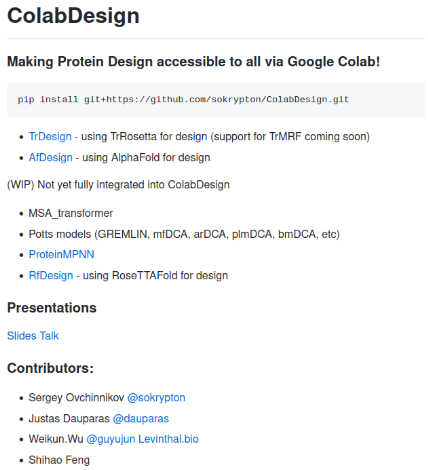
Making Protein folding accessible to all via Google Colab!

Notebooks	monomers	complexes	mmseqs2	jackhmmer	templates
AlphaFold2_mmseqs2	Yes	Yes	Yes	No	Yes
AlphaFold2_batch	Yes	Yes	Yes	No	Yes
RoseTTAFold	Yes	No	Yes	No	No
AlphaFold2 (from Deepmind)	Yes	Yes	No	Yes	No
BETA (in development) notebooks					
OmegaFold	Yes	No	No	No	No
AlphaFold2_advanced	Yes	Yes	Yes	Yes	No
OLD retired notebooks					
AlphaFold2_complexes	No	Yes	No	No	No
AlphaFold2_jackhmmer	Yes	No	Yes	Yes	No
AlphaFold2_noTemplates_noMD					
AlphaFold2_noTemplates_yesMD					

Pour aller plus loin: EMBL webinar

ColabDesign

```
https://github.com/sokrypton/ColabDesign
```



ColabDesign

Making Protein Design accessible to all via Google Colab!

```
pip install git+https://github.com/sokrypton/ColabDesign.git
```

- TrDesign - using TrRosetta for design (support for TrMRF coming soon)
- AfDesign - using AlphaFold for design

(WIP) Not yet fully integrated into ColabDesign

- MSA_transformer
- Potts models (GREMLIN, mfDCA, arDCA, plmDCA, bmDCA, etc)
- ProteinMPNN
- RfDesign - using RoseTTAFold for design

Presentations

Slides Talk

Contributors:

- Sergey Ovchinnikov @sokrypton
- Justas Dauparas @dauparas
- Weikun.Wu @guyujun Levinthal.bio
- Shihao Feng

AlphaFold-Multimer de DeepMind

Richard Evans et al. (Oct. 2021). en. preprint. Bioinformatics

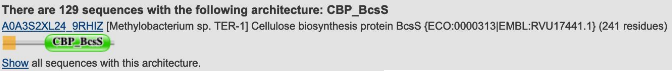
Some AlphaFold use cases

Alex Bateman

EMBL-EBI

Just because AlphaFold can fold it doesn’t mean nature can

- Pfam use case 2: CPB_BcsS family (PF17036)
- AlphaFold prediction of region matched by Pfam identifies incomplete domain
- Structurally similar to MBB clan

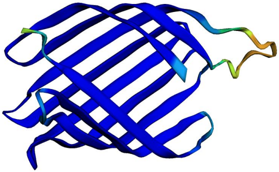


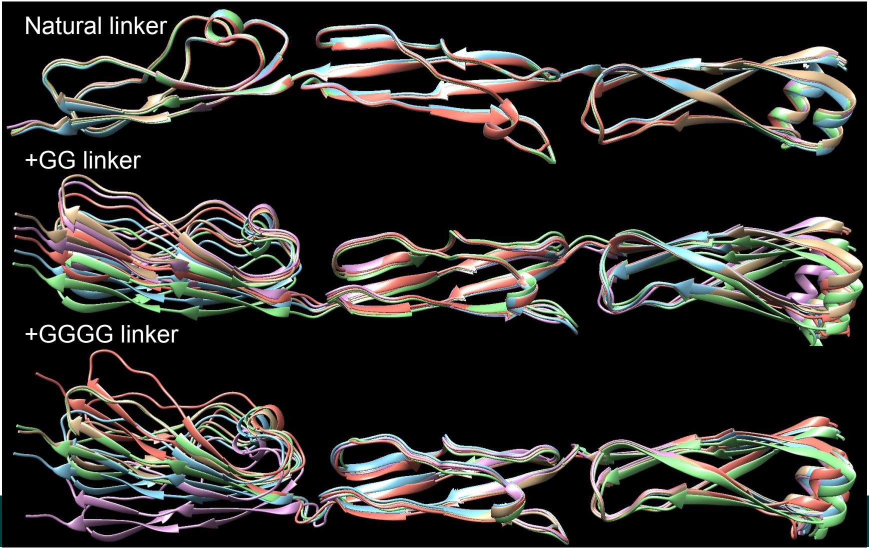
There are 129 sequences with the following architecture: CBP_BcsS

A0A3S2XL24_9RHIZ [Methylobacterium sp. TER-1] Cellulose biosynthesis protein BcsS {ECO:0000313|EMBL:RVU17441.1} (241 residues)

Show all sequences with this architecture.

- This will not be stable in vitro!





Natural linker
+GG linker
+GGGG linker

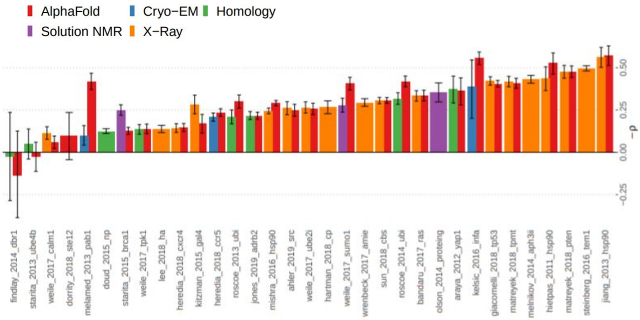
Application of AF2 structures for variant effect prediction and pocket detection

Pedro Beltrao, Group Leader

www.ebi.ac.uk/beltrao
@pedrobeltrao

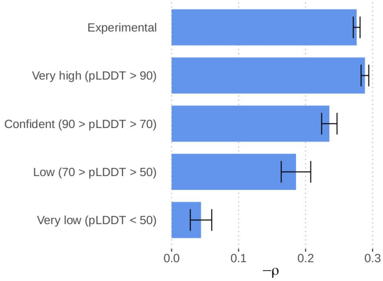
EMBL-EBI

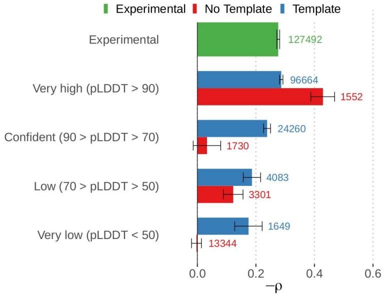
Comparing experimental vs structure based prediction of missense mutations



- Experimental impact of mutations from deep mutational scanning experiments (30 proteins)
- Predicted ddG of mutation using FoldX on alphafold structures or experimental structures
- Alphafold structures give equal or better predictions and it holds for regions with no templates

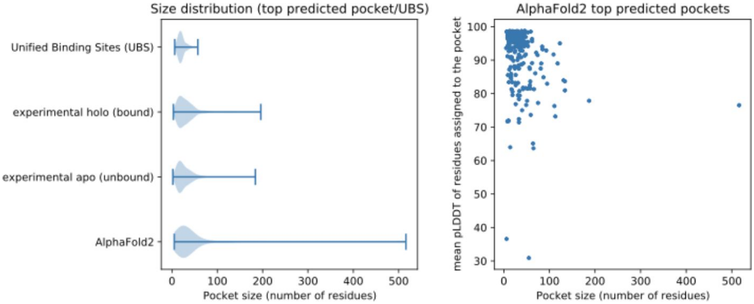
Comparing experimental vs structure based prediction of missense mutations





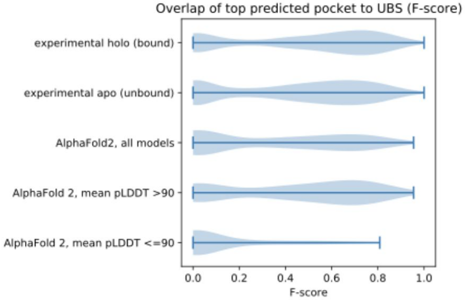
EMBL-EBI

Pocket detection and how to filter the models



We retained 230 of 304 proteins from a dataset by Clark et al., 2020. Pocket detection was performed using ghecom (Kawabata, 2010), as done previously in (Clark et al., 2020).

Pocket detection and how to filter the models



Filtering the pocket residues by confidence (likely also predicted aligned residue) improves pocket detection

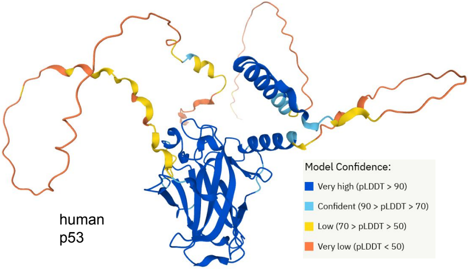
AlphaFold2

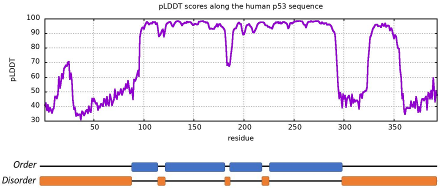
for detecting intrinsically disordered protein regions

Bálint Mészáros
EMBL Heidelberg
08/09/2021

AlphaFold2 indicates the presence of IDRs

AF2 generates coordinates for every residue, even ones that have no fixed structure





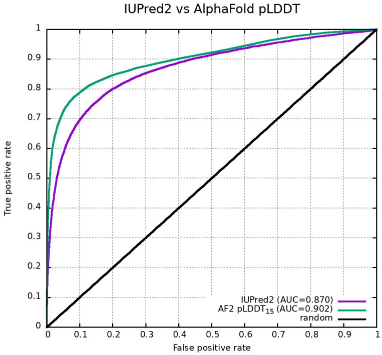
Two interpretations for low confidence:

- AF2 isn't good enough to predict the structure
- There is no structure to predict

pLDDT is a good indicator of disordered regions (in this case)

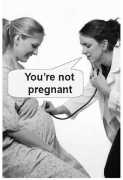
Let's test the generic case – binary disordered prediction

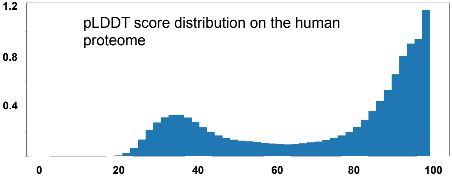
AlphaFold2 as a disorder prediction method


IUPred2 vs AlphaFold pLDDT

false positive

false negative




pLDDT score distribution on the human proteome

EMBL

L’avenir selon DeepMind

Une petite vidéo pour terminer...

https://www.youtube.com/watch?time_continue=8&v=KpedmJdrTpY&feature=emb_title
à regarder chez vous (regarder sur les écrans, vous reconnaissez le logiciel de visualisation que les chercheurs de DeepMind utilisent ?):
https://www.youtube.com/watch?v=gg7WjuFs8F4

The end

- MERCI pour votre attention!

7 Bibliographie

Bibliographie I

Akasaka, Kazuyuki, Ryo Kitahara, and Yuji O. Kamatari (Mar. 2013). “Exploring the folding energy landscape with pressure”. In: *Archives of Biochemistry and Biophysics*. Protein Folding and Stability 531.1–2, pp. 110–115.

Bowman, Gregory R, Vincent A Voelz, and Vijay S Pande (Feb. 2011). “Taming the complexity of protein folding”. en. In: *Current Opinion in Structural Biology* 21.1, pp. 4–11.

Chen, Yiwen et al. (Jan. 2008). “Protein folding: Then and now”. In: *Archives of Biochemistry and Biophysics*. Highlight Issue: Protein Folding 469.1, pp. 4–19.

Daggett, Valerie and Alan R. Fersht (2003). “Is there a unifying mechanism for protein folding?” In: *Trends in biochemical sciences* 28.1, pp. 18–25.

Dill, Ken A. et al. (June 2008). “The Protein Folding Problem”. en. In: *Annual Review of Biophysics* 37.1, pp. 289–316.

Evans, Richard et al. (Oct. 2021). *Protein complex prediction with AlphaFold-Multimer*. en. preprint. Bioinformatics.

Fersht, Alan R. and Valerie Daggett (2002). “Protein folding and unfolding at atomic resolution”. In: *Cell* 108.4, pp. 573–582.

Hosseinzadeh, Parisa et al. (Dec. 2017). “Comprehensive computational design of ordered peptide macrocycles”. en. In: *Science* 358.6369, pp. 1461–1466.

Bibliographie II

Huang, Po-Ssu, Scott E. Boyken, and David Baker (Sept. 2016). "The coming of age of *de novo* protein design". en. In: *Nature* 537.7620, pp. 320–327.

Kuhlman, Brian et al. (Nov. 2003). "Design of a Novel Globular Protein Fold with Atomic-Level Accuracy". en. In: *Science* 302.5649, pp. 1364–1368.

Lane, Thomas J et al. (Feb. 2013). "To milliseconds and beyond: challenges in the simulation of protein folding". en. In: *Current Opinion in Structural Biology* 23.1, pp. 58–65.

Muñoz, Victor and Michele Cerminara (Sept. 2016). "When fast is better: protein folding fundamentals and mechanisms from ultrafast approaches". In: *Biochem J* 473.17, pp. 2545–2559.

Piana, Stefano, John L Klepeis, and David E Shaw (Feb. 2014). "Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations". en. In: *Current Opinion in Structural Biology* 24, pp. 98–105.

Plaxco, Kevin W, Kim T Simons, and David Baker (Apr. 1998). "Contact order, transition state placement and the refolding rates of single domain proteins11Edited by P. E. Wright". In: *Journal of Molecular Biology* 277.4, pp. 985–994.

Thirumalai, D. et al. (2010). "Theoretical Perspectives on Protein Folding". In: *Annual Review of Biophysics* 39.1, pp. 159–183.

Bibliographie III

Voelz, Vincent A. et al. (Aug. 2012). “Slow Unfolded-State Structuring in Acyl-CoA Binding Protein Folding Revealed by Simulation and Experiment”. In: *J. Am. Chem. Soc.* 134.30, pp. 12565–12577.

Woodside, Michael T. and Steven M. Block (2014). “Reconstructing Folding Energy Landscapes by Single-Molecule Force Spectroscopy”. In: *Annual Review of Biophysics* 43.1, pp. 19–39.