IBPM Annual Meeting 2022/02/1-2

FUTURE@IBPM 2022

Discussing projects

Sharing skills

Generating knowledge

Protein structure prediction:
Is AlphaFold the solution to the folding problem?

Bioinformatics@IBPM

Teresa Colombo, Veronica Morea, Allegra Via



Ph.D. students: **Chiara Pacelli, Gianmarco Pascarella**


Consiglio Nazionale delle Ricerche

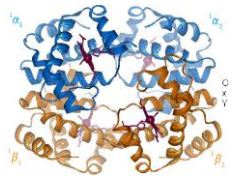
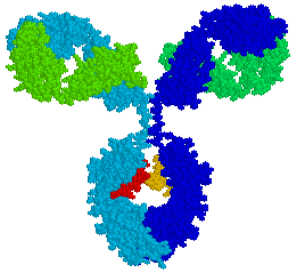
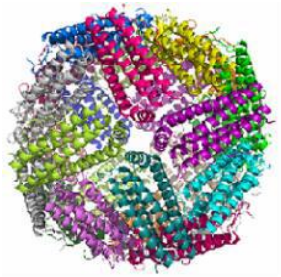
INFRASTRUTTURA ITALIANA DI BIOINFORMATICA

SAPIENZA UNIVERSITÀ DI ROMA

Department of Biochemical Sciences «A. Rossi Fanelli»

Proteins

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

BIO-MACRO-MOLECULES	Hemoglobin	Antibody	Ferritin
amino acid sequences	>2DN3_1|Chain A|Hemoglobin alpha subunit|Homo sapiens (9606) VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTK TYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMPNAL SALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPA VHASLDKFLASVSTVLTSKYR >2DN3_2|Chain B|Hemoglobin beta subunit|Homo sapiens (9606) VHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQR FFESFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDN LKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGK EFTPPVQAAYQKVVAGVANALAHKYH	>1HZH:M|PDBID|CHAIN|SEQUENCE EIVLTQSPGTLSLSPGERATFSCRSSHSIRSRRVAWYQHK PGQAPRLVIHGVSNRASGISDRFSGSGSGTDFTLTITRVE PEDFALYYCQVYGASSYTFGQGTKLERKRTVAAPSVFIFP PSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNS QESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQ GLRSPVTKSFNRGEC >1HZH:K|PDBID|CHAIN|SEQUENCE QVQLVQSGAEVKKPGASVKVSCQASGYRFSNFVIHWVRQA PGQRFEWMGWINPYNGNKEFSAKFQDRVTFTADTSANTAY MELRSLRSADTAVYYCARVGPYSWDDSPQDNYYMDVWGKG TTVIVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYF PEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPS SSLGTQTYICNVNHKPSNTKVDKKAEPKSCDKTHTCPPCP APELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHED PEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLH QDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYT LPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENN YKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHE ALHNHYTQKSLSLSPGK	(24 x) >4OYN_1|Chain A|Ferritin heavy chain|Homo sapiens (9606) MTTASTSQVRQNYHQDSEAAINRQINLELYASYVYLSMSY YFDRDDVALKNFAKYFLHQSHEEREHAEKLMKLQNQRGGR IFLQDIKKPDCDDWESGLNAMECALHLEKNVNQSLLELHK LATDKNDPHLCDFIETHYLNEQVKAIKELGDHVTNLRKMG APESGLAEYLFDKHTLGDSDNES
↓			
3D structures			
↓			
biological functions	Oxygen transport	Defense	Iron storage

[(https://pdb101.rcsb.org/learn/videos/what-is-a-protein-video)](https://pdb101.rcsb.org/learn/videos/what-is-a-protein-video)

Proteins

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

Sequence and structure databases

amino acid sequences

↓

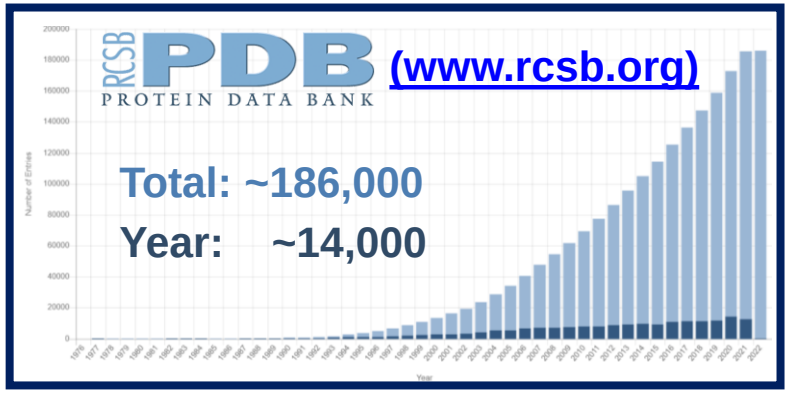
3D structures

NCBI (www.ncbi.nlm.nih.gov)

> 35,000,000

0.5% ↙

99.5% ↘



Folding problem

(https://pdb101.rcsb.org/learn/videos/what-is-a-protein-video)

The «folding problem»

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

«Given a protein sequence, what 3D structure will it assume?»

↓

Protein structure prediction methods => 3D models

↓

How accurate are protein models?

↓

Protein 3D model (M) with 3D structure of target (T) comparison

Protein model accuracy assessment

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

3D model (M) with 3D structure of target (T) comparison

1° scenario:
Target 3D structure available from PDB

↓

Information can affect prediction

2° scenario:
Target 3D structure *not* available from PDB

↓

Model accuracy cannot be evaluated

Until 1994

3° scenario:
Target 3D structure *not* available from PDB *but* available for model evaluation

↓

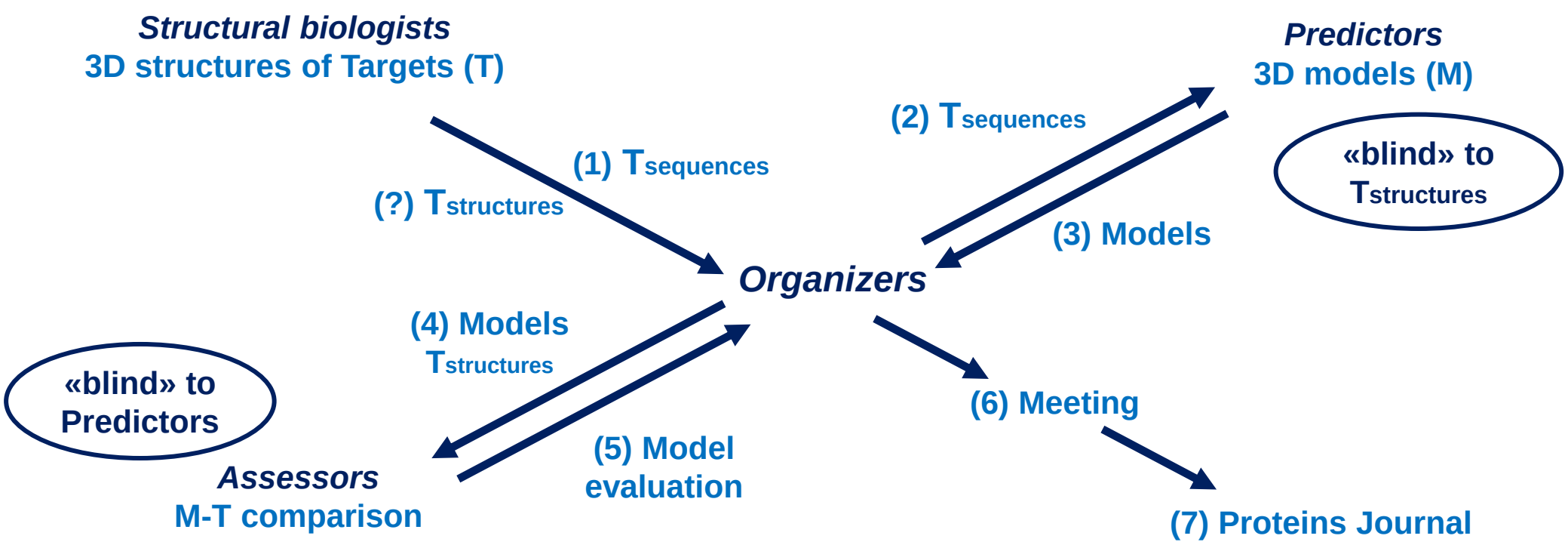
Since 1994: CASP

Protein model accuracy assessment

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

3D model (M) with 3D structure of target (T) comparison

CASP: «Critical Assessment of Structure Prediction»



(https://predictioncenter.org/index.cgi?page=proceedings)

Protein model accuracy assessment

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

3D model (M) with 3D structure of target (T) comparison

CASP: «Critical Assessment of Structure Prediction»

- ✓ **Double-blind:**
  - **Predictors: «blind» to Target proteins 3D structures**
  - **Assessors: «blind» to Predictors identity**
- ✓ **World-wide:**
  - **Predictors: majority of groups active in the field**
- ✓ **Long-standing:**
  - **Every two years since 1994**

«Gold-standard» Protein Structure Prediction Assessment

↙ **Peers** ↘ **Funding Agencies**

Issue: Statistical Significance?

(https://predictioncenter.org/index.cgi?page=proceedings)

Protein model accuracy assessment

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

CASP: «Gold Standard» Assessment

CASP Year–Round	Proteins J. Year–Vol.	Targets (T) Seq	Targets (T) 3D	Groups All	Groups 3D	Models All	Models 3D
1994–01	1995–23	35	30	40	30	200	200
1996–02	1997–29	40	25	70	60	2,000	900
1998–03	1999–37	45	40	100	60	4,000	2,000
2000–04	2001–45	45	40	160	110	10,000	5,000
2002–05	2003–53	70	60	220	180	30,000	25,000
2004–06	2005–61	90	70	210	170	40,000	25,000
2006–07	2007–69	100	95	250	180	65,000	50,000
2008–08	2009–77	130	120	240	160	80,000	60,000
2010–09	2011–79	130	120	250	170	90,000	60,000
2012–10	2014–82	160	100	210	150	70,000	50,000
2014–11	2016–84	210	90	210	140	60,000	40,000
2016–12	2018–86	150	70	190	130	55,000	40,000
2018–13	2019–87	110	75	190	110	60,000	35,000
2020–14	2021–89	90	70	220	150	70,000	45,000
2022–15							

Issue: Statistical Significance? RCSB PDB PROTEIN DATA BANK Total: ~186,000 Year: ~14,000

(https://predictioncenter.org/index.cgi?page=proceedings)

Protein structure prediction methods

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

CASP: many different protein structure prediction methods

Template-Based Modelling (TBM)

Need

Template Free Methods (TFM)

Do not need

Template(s): Experimentally determined 3D structure(s) of proteins homologous to the Target

(https://predictioncenter.org/index.cgi?page=proceedings)

Protein structure prediction methods

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

Template-Based Modelling (TBM)

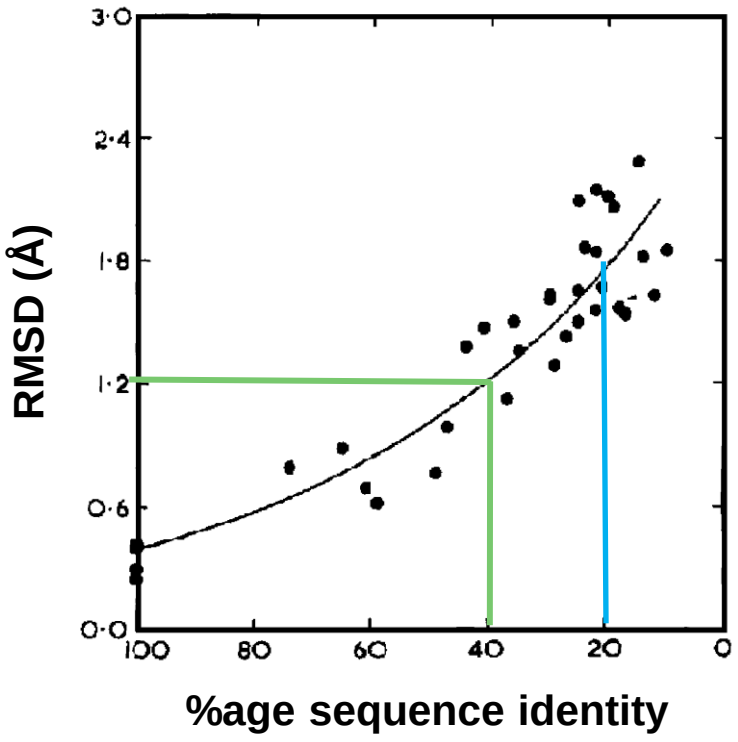
Based on Evolutionary Information (Homology or Comparative modelling)

Similar amino acid sequences

↓

Similar 3D structures

The **lower** the **RMSD** value, the **higher** the **structure similarity**



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

δ_i : **distance** between atom *i* of structure A and atom *i* of structure B
Usually C_α or main-chain atoms (N, C_α, C, O)

Drawback: high sensitivity to poorly aligned regions

1 Angstrom (Å) = 10^{-10} m = 0.0000000001 m

Protein structure prediction methods

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

Template-Based Modelling (TBM)

Based on Evolutionary Information (Homology or Comparative modelling)

Similar amino acid sequences

↓

Similar 3D structures

If **Sequence of Target is ≈ Sequence of Template(s)**

And **3D Structure of Template(s) is known**

↓

Build 3D Model of Target:

1) Align sequence of Target and sequence(s) of Template(s)
2) Copy co-ordinates of conserved regions from 3D structure of Template(s) to Model of Target
3) Use other methods to model non conserved regions

↓ Only protein regions similar to already known ones

↑ Known structures cluster into ~800 folds (~1,000 predicted)

↑ Known Model Accuracy

Protein structure prediction methods

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledge ments

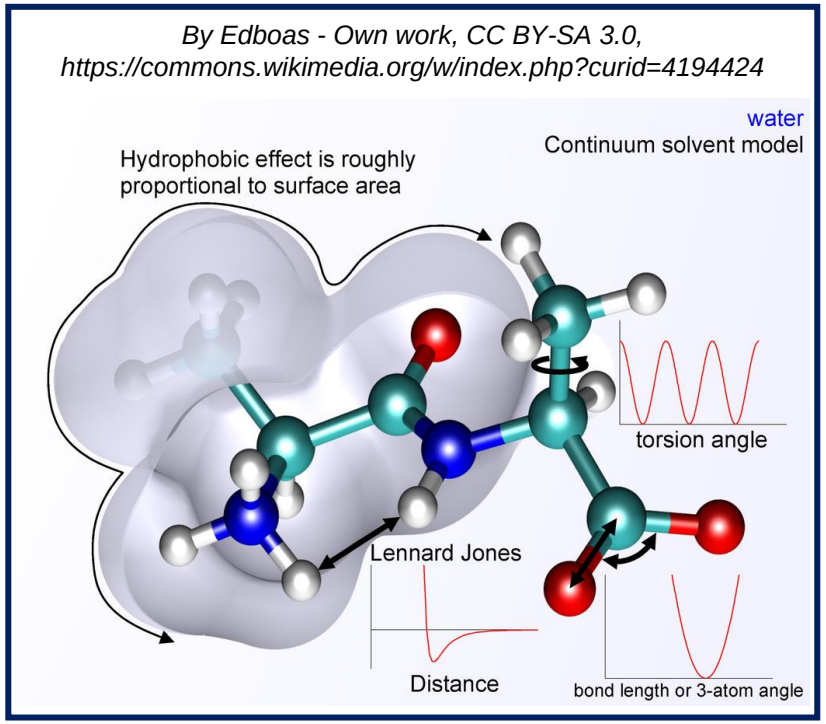
Template Free (TF) 1: *Ab initio* methods

Based on

Chemical-physical principles

Use

1) **Molecular Dynamics (MD) / stochastic methods => generate different conformations**
2) **Energy minimization (EM) => identify local lowest energy conformation (native structure)**
3) **Empirical energy functions (force fields)**



$$E = E_{\text{covalent}} + E_{\text{noncovalent}} \quad \begin{cases} E_{\text{covalent}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} \\ E_{\text{noncovalent}} = E_{\text{electrostatic}} + E_{\text{van der Waals}} \end{cases}$$

Protein structure prediction methods

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

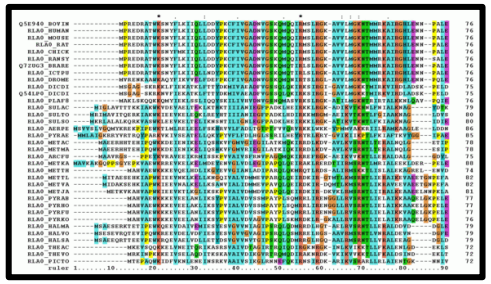
Template Free (TF) 2: New Fold (NF) prediction methods

Based on

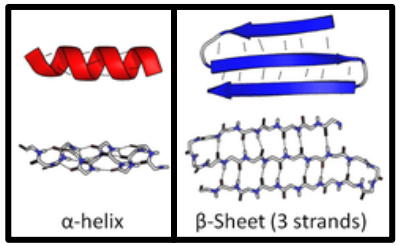
1) Evolutionary information (not whole templates)
2) Chemical-physical principles

Use

Artificial Intelligence (AI)
Neural Networks (NN), Deep Learning (DL), …

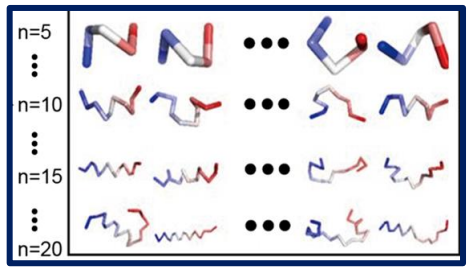


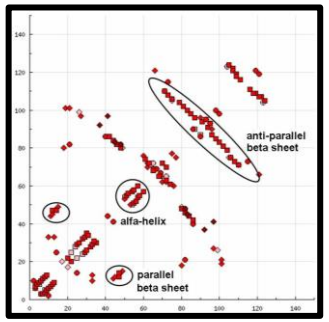
Multiple sequence alignments (MSA)



Secondary Structure Elements (SSE)

Structural Fragments





Inter-residue contacts (RR)

How accurate are protein models?

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

CASP1 (1994): The First Results

1) Template free methods (TFM): no useful 3D models
2) Template-based methods (TBM): high quality models
   - Template identification
   - Correct Target-Template sequence alignment production
   - Target-Template structurally conserved regions only (copying from Nature)

Disappointing: Protein Folding Problem not solved (despite claims)

Invaluable:

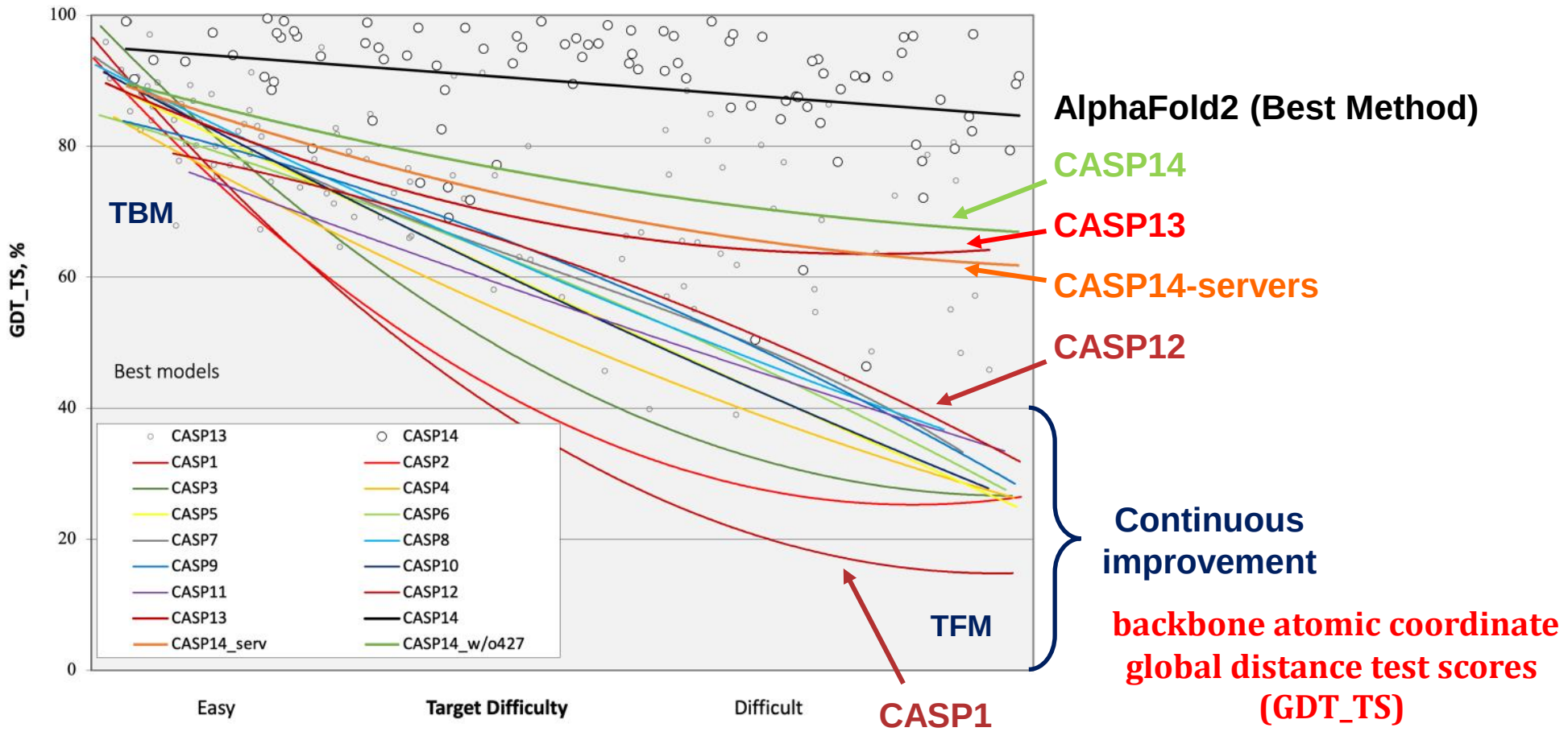
1) First objective assessment of methods accuracy
2) Promoted cross-fertilization of ideas
3) Stimulated research on method development/improvement

(https://predictioncenter.org/decoysets2019/description.cgi?casp=CASP1)

How accurate are protein models?

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

CASP13 & CASP14: Breakthrough!!!



By John Moult - Original publication: CASP 14 introductory presentation, slide 19
Immediate source: https://predictioncenter.org/casp14/doc/presentations/2020_11_30_CASP14_Introduction_Moult.pdf,
Fair use, https://en.wikipedia.org/w/index.php?curid=65998083

How accurate are protein models?

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

Improvements until CASP12 ➡➡➡ CASP13 & CASP14: Breakthrough!!!

CASP Year–Round	Proteins Year–Vol.
1994–01	1995–23
1996–02	1997–29
1998–03	1999–37
2000–04	2001–45
2002–05	2003–53
2004–06	2005–61
2006–07	2007–69
2008–08	2009–77
2010–09	2011–79
2012–10	2014–82
2014–11	2016–84
2016–12	2018–86
2018–13	**2019–87**
2020–14	**2021–89**

CASP13 (2018)

Science
The Guardian
The New York Times
DeepMind blog
Forbes
M. Alquraishi blog

CASP14 (2020)

Science
The Guardian
The New York Times
DeepMind blog
Fortune
Nature
CNBC news
Bloomberg
Financial Post
MIT Technology Review
CASP Press Release
The Telegraph
Daily Mail
Tech Crunch
Venture Beat
New Scientist
SciTech Daily
Eureka Alert
News Medical
MedCity News
BBC news
The Verge

(https://predictioncenter.org/casp14/results.cgi)

How accurate are protein models?

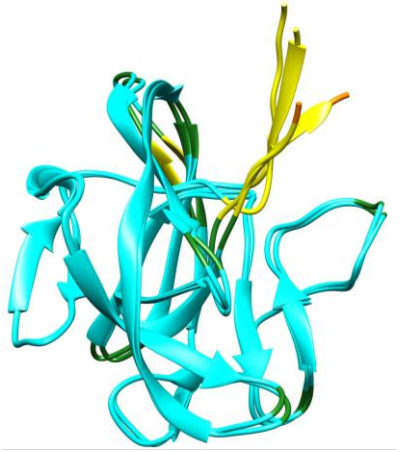
Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

TBM: Best Models Much better than Best template & ≈ Experimental Structure

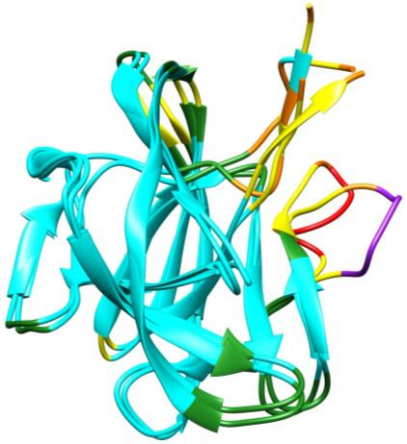
Cα-Cα distance from Target

- < 1 Å
- < 2 Å
- < 4 Å
- < 8 Å
- > 8 Å (Model)
- > 8 Å (Target)

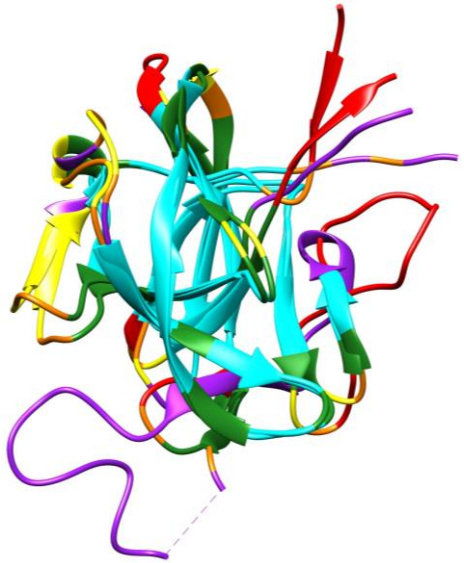
Easy Target: T1034-D1 (156 a.a.)



Best model (AlphaFold2)
GDT-TS = 94
LGA = 96, RMSD = 0.87



2° Best model (MULTICOM)
GDT-TS = 87
LGA = 90, RMSD = 1.15



Best-template: 6FRH_A
LGA = 82, RMSD = 1.6

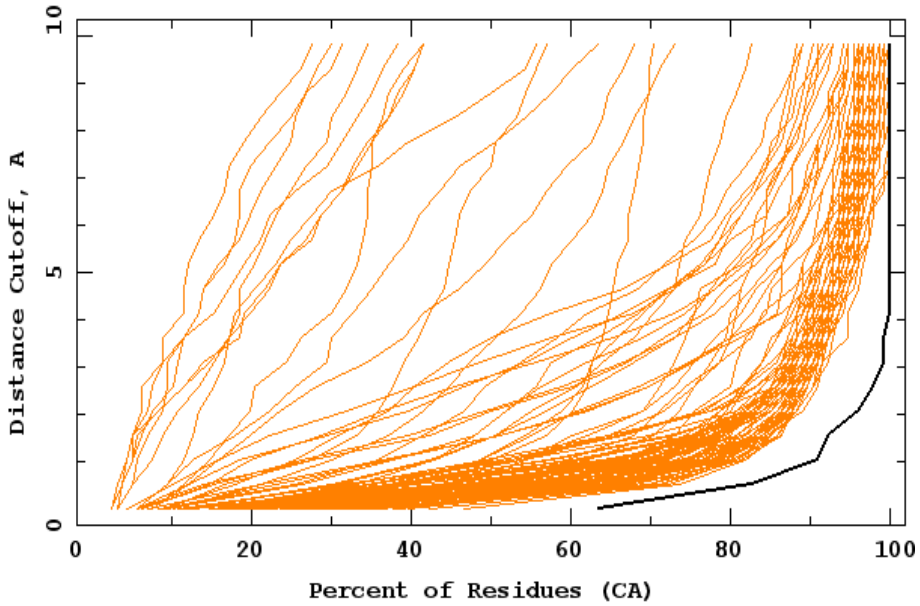
(https://predictioncenter.org/casp14/results.cgi)

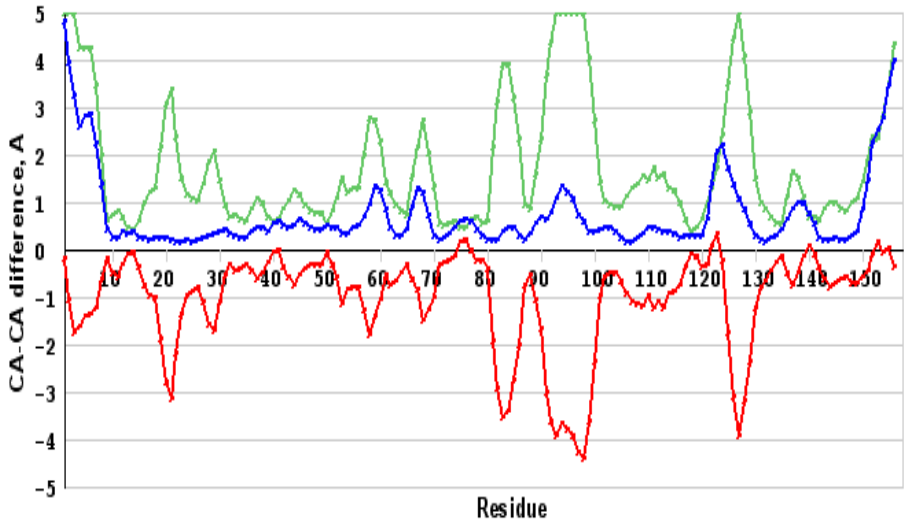
How accurate are protein models?

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

TBM: Best Models Much better than Best template & ≈ Experimental Structure

Easy Target: T1034-D1 (156 a.a.)





Other Models GDT-TS = 87-9

Best-template: 6FRH_A

Best model (AlphaFold2) GDT-TS = 94

(https://predictioncenter.org/casp14/results.cgi)

How accurate are protein models?

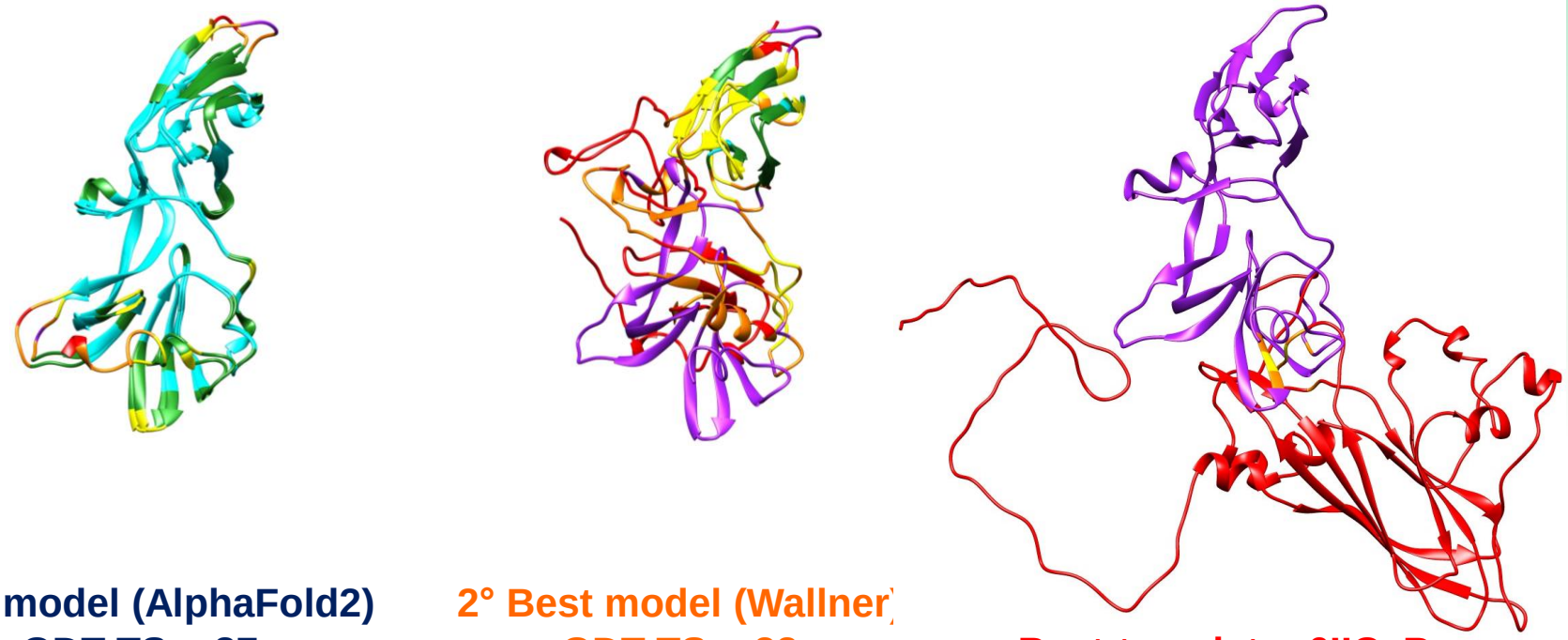
Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

TFM: Best Models with no template ≈ Experimental Structure

Difficult Target: T1038-D1 (189 a.a.)

Cα-Cα distance from Target

- < 1 Å
- < 2 Å
- < 4 Å
- < 8 Å
- > 8 Å (Model)
- > 8 Å (Target)



Best model (AlphaFold2)
GDT-TS = 87
LGA = 91, RMSD = 1.23

2° Best model (Wallner)
GDT-TS = 32
LGA = 32, RMSD = 2.51

Best-template: 6IIC_B
LGA = 28, RMSD = 2.3

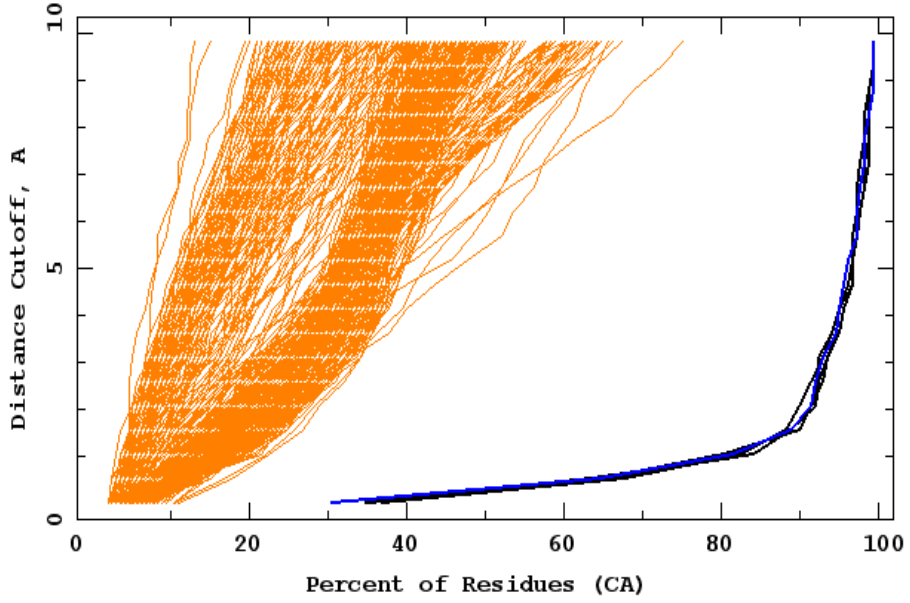
(https://predictioncenter.org/casp14/results.cgi)

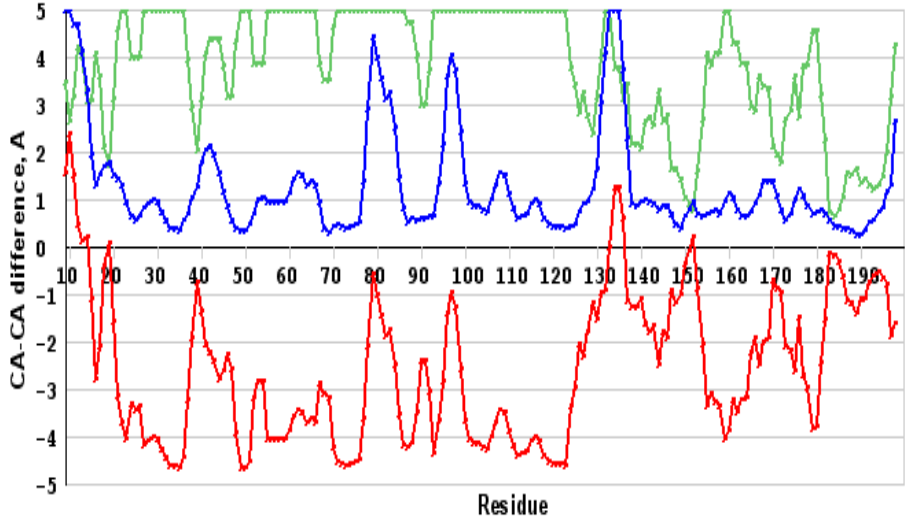
How accurate are protein models?

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

TFM: Best Models with no template ≈ Experimental Structure

Difficult Target: T1038-D1 (189 a.a.)





Other Models GDT-TS = 32-7

Best-template: 6IIC_B

Best model (AlphaFold2) GDT-TS = 87

(https://predictioncenter.org/casp14/results.cgi)

How accurate are protein models?

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

Prediction Methods	CASP1 (1994)	CASP2-12 (1996-2016)	CASP13-14 (2018-2020)
Template-based (TBM)	Best models only as good as the best template structure	Some of best models include loops	Consistent models quality ≈ experimental structures
Evolutionary (TFM)	-	Sporadically good results	Consistent models quality ≈ experimental structures
Ab initio (TFM)	no useful results	Sporadically loops & refinement	Sporadically loops & refinement
Protein Folding Problem	far from being solved	far from being solved	«giant leap forward» ***BUT:*** Unpredicted regions? (~1/3) Different proteins? Human intervention? Theoretical understanding?

(https://predictioncenter.org/casp14/results.cgi)

How accurate is «my» model?

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

Variables

1) Program
2) Human intervention
3) Predicted accuracy
4) Homologous structure
5) Specific protein
6) Other features

How accurate is «my» model?

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

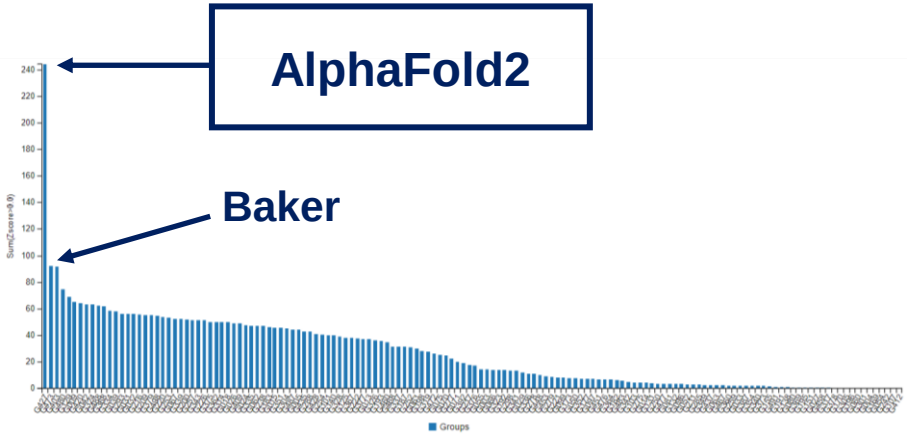
Variables

1) **Program**
2) Human intervention
3) Predicted accuracy
4) Homologous structure
5) Specific protein
6) Other features

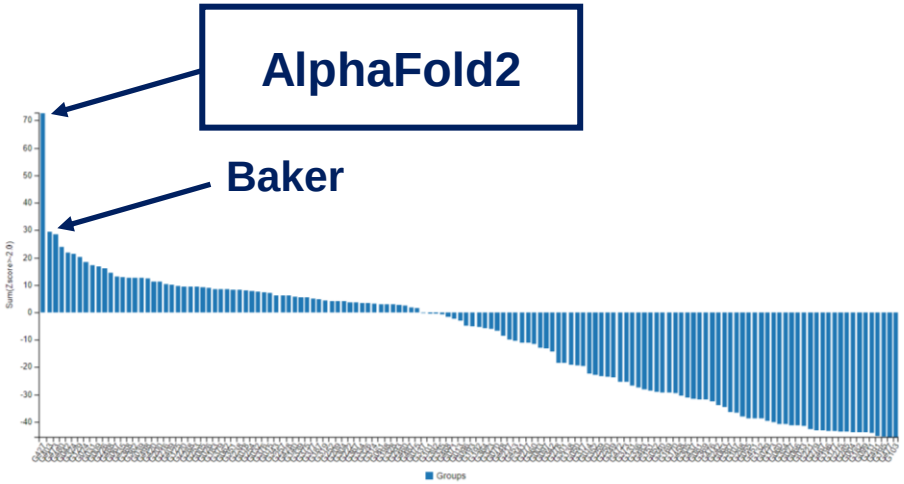
Yaxis: Sum of Zscores for GDT_TS values on ALL targets (i.e., TBM easy and hard, TBM/TFM, TFM)

Xaxis: Group number

Zscore $z = (x - \mu)/\sigma$ number of standard deviations (σ) from average (μ)



Zscore Automatic evalutation: GDT-TS Cα



Zscore Assessors: additional parameters (*e.g.*, side-chains, steric clashes, quality prediction)

(https://predictioncenter.org/casp14/results.cgi)

How accurate is «my» model?

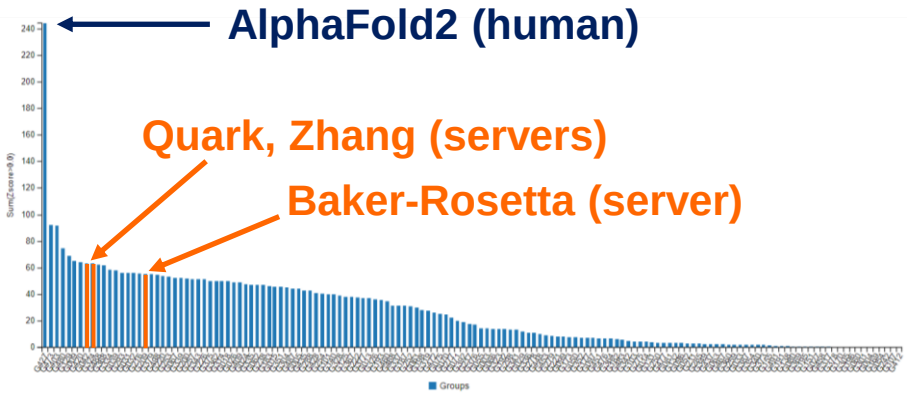
Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

Variables

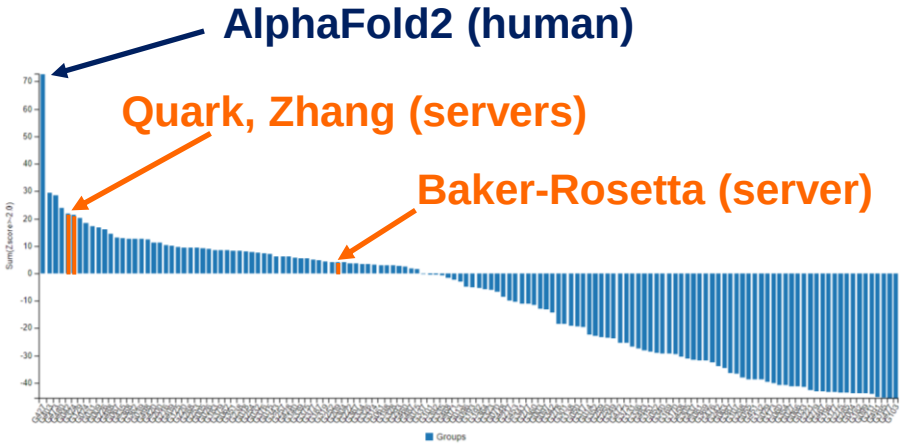
1) Program
2) **Human intervention**
3) Predicted accuracy
4) Homologous structure
5) Specific protein
6) Other features

Human Experts better than best programs

Best programs better than non expert users



Zscore: GDT-TS Cα



Zscore: additional parameters
(*e.g.*, side-chains, steric clashes, quality prediction)

(https://predictioncenter.org/casp14/results.cgi)

How accurate is «my» model?

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

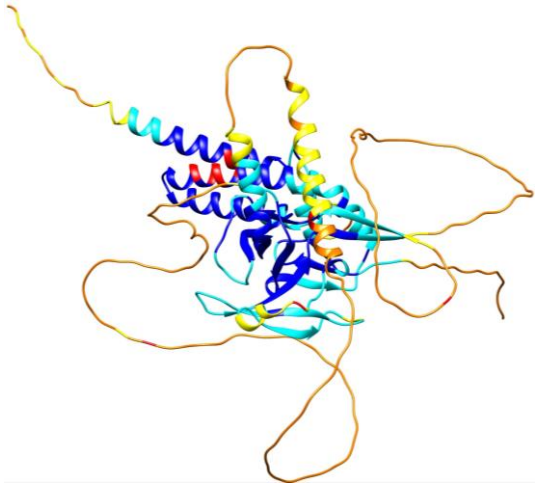
Variables

1) Program
2) Human intervention
3) **Predicted accuracy**
4) Homologous structure
5) Specific protein
6) Other features

CASP Assessment

Predicted Accuracy Values *(B-factor column)* compared with measured T-M distances

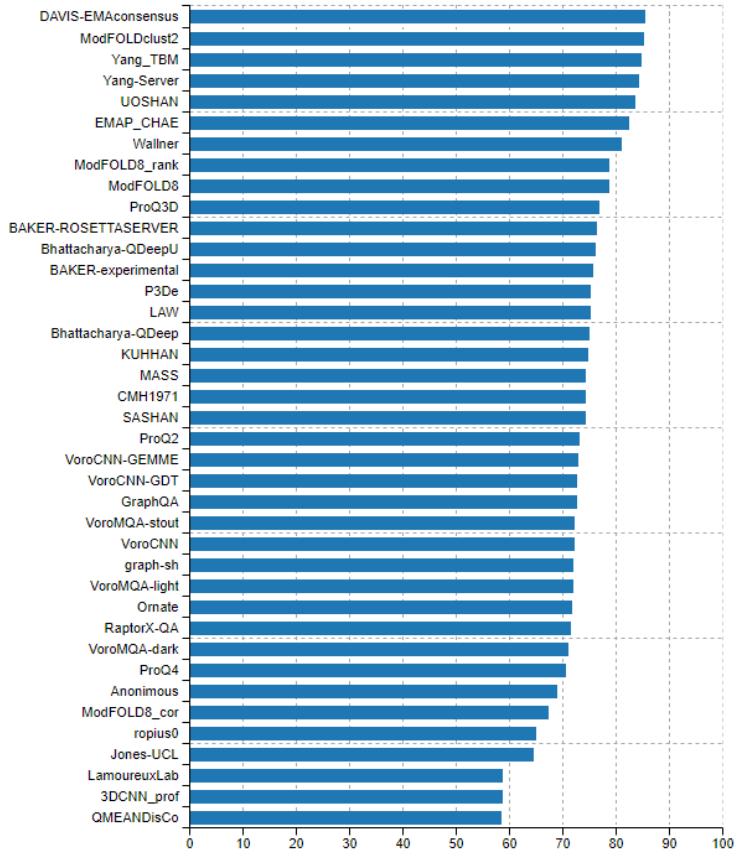
AlphaFold2 not in this category



AlphaFold2
Predicted Accuracy

- Very high
- High
- Low
- Very low (regions may be unstructured)

Group



Predicted accuracy

https://predictioncenter.org/casp14/qa_corr.cgi

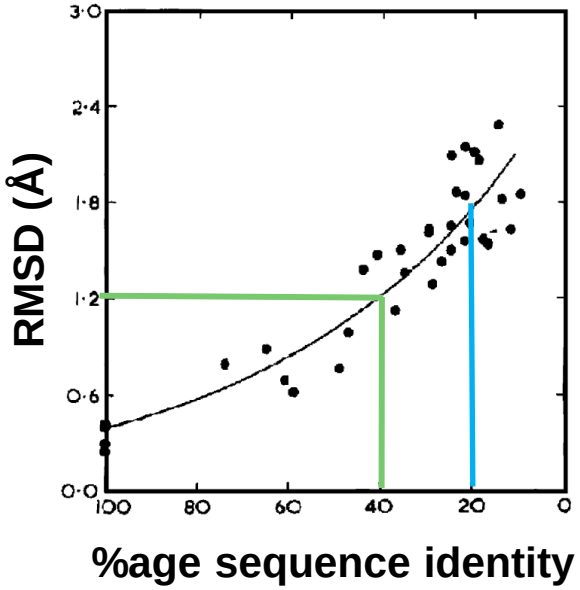
How accurate is «my» model?

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

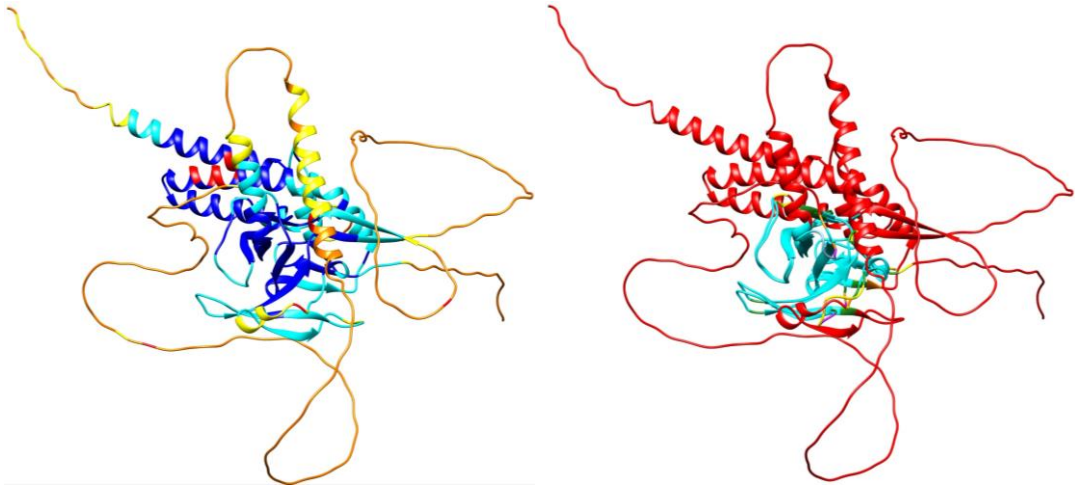
Variables

1) Program
2) Human intervention
3) Predicted accuracy
4) **Homologous structure**
5) Specific protein
6) Other features

Protein 3D structures are more conserved than amino acid sequences



Model-Homologue
Cα-Cα distance < 1.0 Å



Model-Homologue conserved regions (cyan):

RMSD with Target structure < value depending on %age sequence identity (1.0-2.0 Å)

How accurate is «my» model?

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

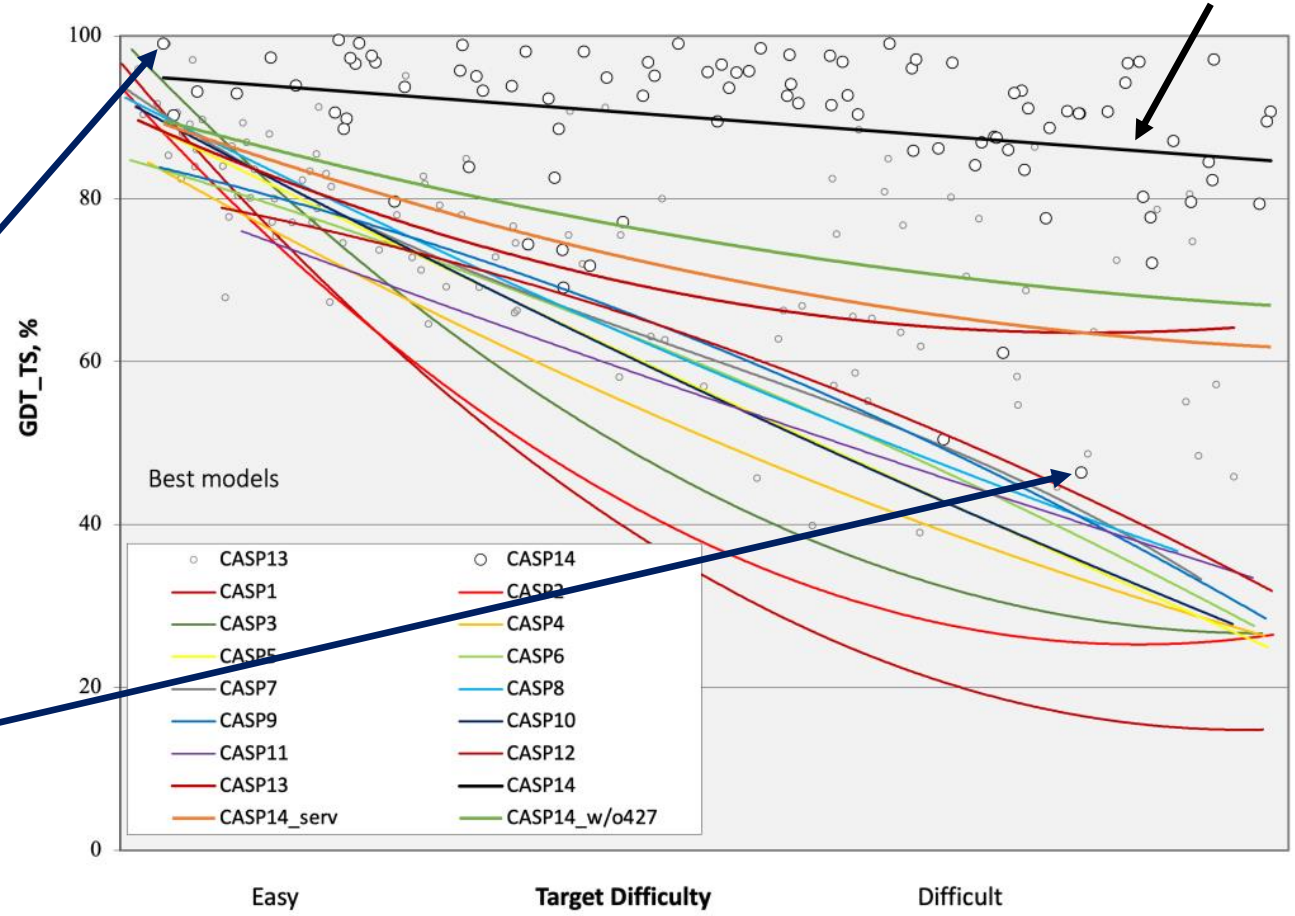
Variables

1) Program
2) Human intervention
3) Predicted accuracy
4) Homologous structure
5) **Specific protein**
6) Other features

CASP14 best case scenarios (GDT-TS ≈ 100)

CASP14 worst case scenario (GDT-TS ≈ 45)

AlphaFold2



(https://predictioncenter.org/casp14/results.cgi)

How accurate is «my» model?

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

Variables

1) Program
2) Human intervention
3) Predicted accuracy
4) Homologous structure
5) Specific protein
6) **Other features**

AlphaFold2 not these categories

Accuracy Estimation
- for overall models, domains and residues

Refinement
- Model improvement (Feig or Baker)

Assembly (with CAPRI)
- protein-protein, subunit-subunit, and domain-domain interactions

Data Assisted
- e.g., crosslinking data, SAXS, NMR

Biological Relevance
- answers to biological questions (e.g., experimental structure determination, function determinants)

(https://predictioncenter.org/casp14/results.cgi)

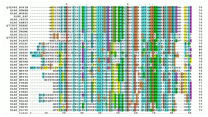
Protein structure prediction methods: AlphaFold2

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

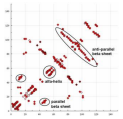
CASP14: best for 88/97 Targets & accuracy ≈ experimental structures

Previous Template Free

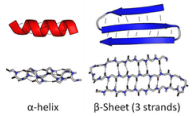
MSA



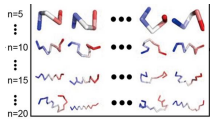
RR



SSE

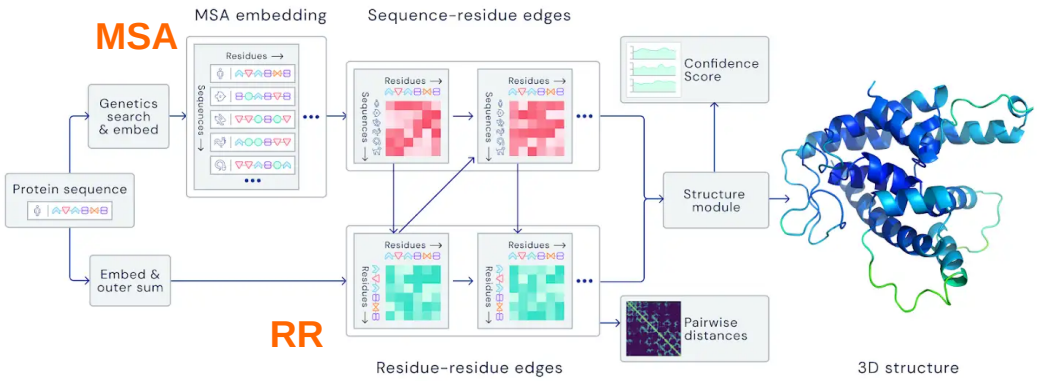


Fragments



Ab initio *Energy calculations*

AI system developed by DeepMind (https://deepmind.com/)



*Deep learning method – trained on >170.000 3D structures (AlphaFold1: 29,000) – 2 main modules: **MSA** (red) and **RR** (green) – reciprocal feed (e.g., correlated mutations in large MSAs => contact maps => distance maps); relevant data are brought together and irrelevant data are filtered out ("attention algorithm") – first amino acid clusters, then clusters joining – **SSE** down-tuning (overfitting in AlphaFold1) – **no fragment** libraries – "**Ab initio**" energy refinement (AMBER): last step, only slight adjustments*

Protein Models
(https://alphafold.ebi.ac.uk/)

- ~50 proteomes (human, parasites and model organisms)
- SwissProt
- Aiming at UniRef90 (1,000,000 sequences)

Prediction Server
(https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb)

Source code
(https://github.com/deepmind/alphafold)

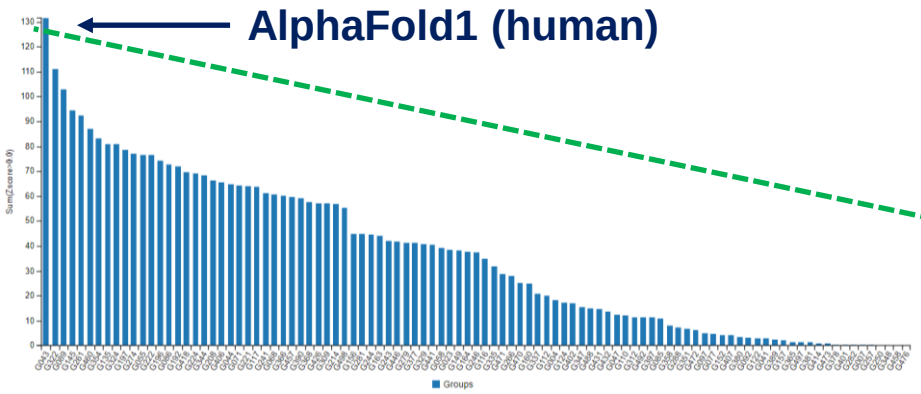
By DeepMind - Original publication: DeepMind blog: AlphaFold: a solution to a 50-year-old grand challenge in biology
Immediate source: https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology,
Fair use, https://en.wikipedia.org/w/index.php?curid=65993510

Future Perspectives

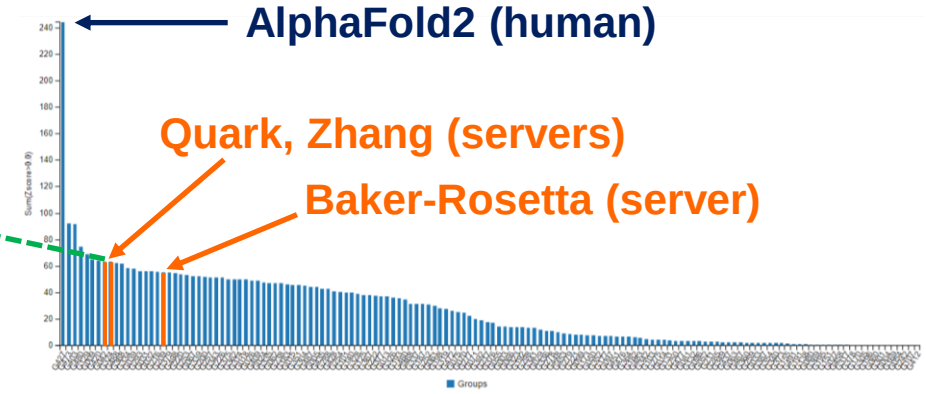
Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

CASP15 (2022): further improvements expected

CASP13 (2018)



CASP14 (2020)



Yaxis:
- Zscores for GDT_TS values on ALL targets (i.e., TBM easy and hard, TBM/TFM, TFM)

Xaxis:
- Group number

Best server in CASP14 performed as well as best humans in CASP13

Future Perspectives

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

Protein Structure Prediction

CASP15 (2022)

- Running (https://predictioncenter.org/news.cgi#409)
- Proposed changes (e.g., add RNA, dynamics)
- Experimentalists contribution essential for methods assessment and improvement

AlphaFold

- Expected to further improve: GDT-TS ≈ 100 on all Targets?

Folding Problem:

- A.I. provides accurate models for most proteins
- *Ab initio* principles not closer to being understood
- A.I. «black box» opening may help

Artificial intelligence (A.I.)

A.I.: “intelligence demonstrated by machines”

Intelligence: “ability to perceive the environment and take actions that maximize the chance of achieving goals”

A.I. ultimate goal: general (human-like) intelligence, i.e., ability to solve any problem

A.I. impact huge in specific fields since 2015

- biological sciences (AlphaFold2: Protein Structure Prediction)
- Deepmind: disease diagnosis, energy saving, web search (Google)
- Recommendation (Amazon, Netflix, YouTube)
- Human speech (Siri, Alexa)
- Self-driving cars (Tesla)

Acknowledgements

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

John Moult
University of Maryland
CASP founder and organizer

Anna Tramontano (1957-2017)
CASP organizer and assessor

CASP organizers, predictors, assessors and

3D structure providers:

please, provide 3D structures to next CASPs!!!

Cyrus Chothia (1942-2019)
Arthur M. Lesk (Pennsylvania State University)
Pioneering work on protein 3D structure analysis and prediction

Tavolo di lavoro del CNR su A.I.

Caudai C (ISTI), Galizia A (IMATI), Geraci F, ***Le Pera L*** *(IBIOM-IBPM ->ISS),* ***Morea V****, Salerno E (ISTI),* ***Via A, Colombo T.***

"A.I. applications in functional genomics"

Comput Struct Biotechnol J *(2021) 19: 5762-5790*

Gianmarco Pascarella
help with talk slides and AlphaFold2 use

How accurate are protein models?

Proteins
Folding Problem
Protein model accuracy assessment
Protein structure prediction methods
How accurate are protein models?
How accurate is my model?
Future perspectives
Acknowledgements

CASP: 3D model (M) with 3D structure (T: Target) comparison

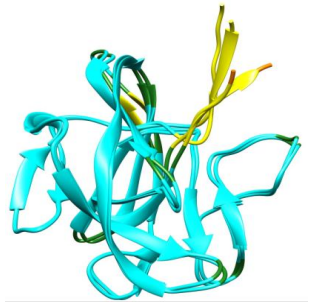
Structure Similarity Measure:

GDT-TS

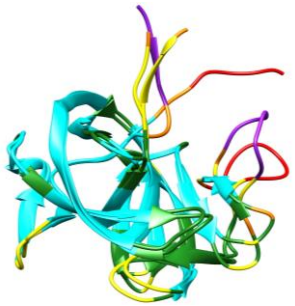
(Global Distance Test–Total Score)

GDT_TS = (GDT_P1 + GDT_P2 + GDT_P4 + GDT_P8) / 4

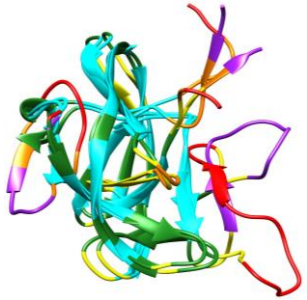
High GDT-TS values => High Model Quality



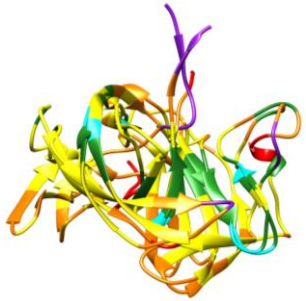
GDT-TS = 90



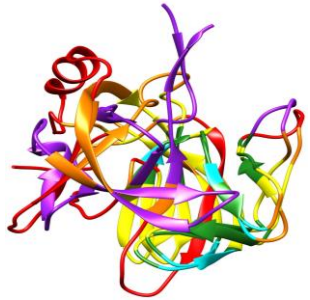
GDT-TS = 80



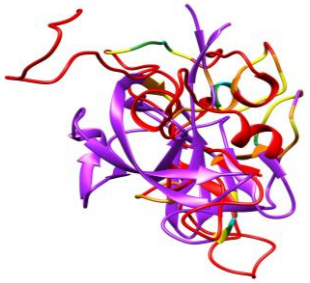
GDT-TS = 70



GDT-TS = 50



GDT-TS = 30



GDT-TS = 10

(https://predictioncenter.org/index.cgi?page=proceedings)

References