AI in SAS II

Emre H Brookes

EMBO Practical :
Small Angle Neutron and X-ray Scattering
from biomacromolecules in solution

19 September 2024

UNIVERSITY OF MONTANA

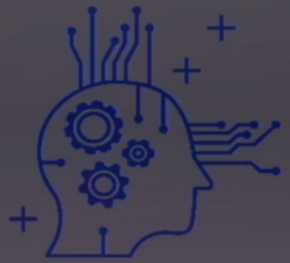
SCHOOL OF MEDICINE
UT HEALTH SCIENCE CENTER®
DEPARTMENT OF BIOCHEMISTRY

University of Lethbridge
FIAT LUX

AI

Artificial Intelligence

Any technique that enables computers to mimic human behavior

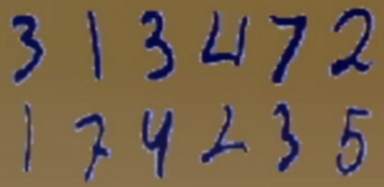


Machine Learning

Ability to learn without explicitly being programmed

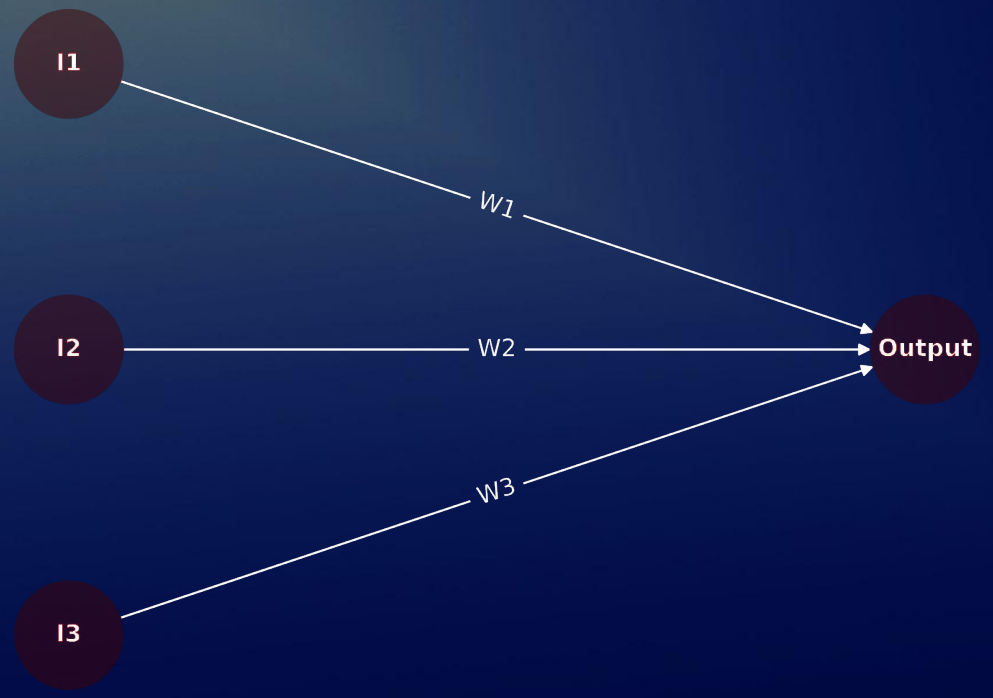
Deep Learning

Extract patterns from data using neural networks



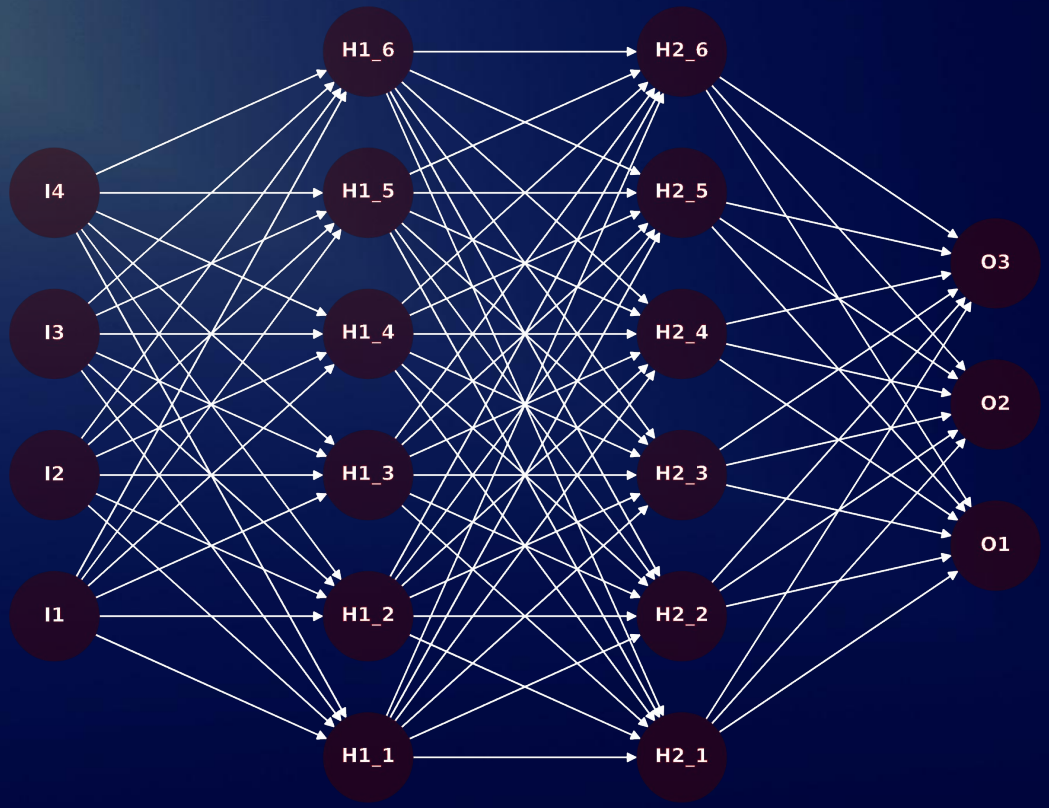
Credit: MIT 6.S191 introtodeeplearning.com

Perceptron

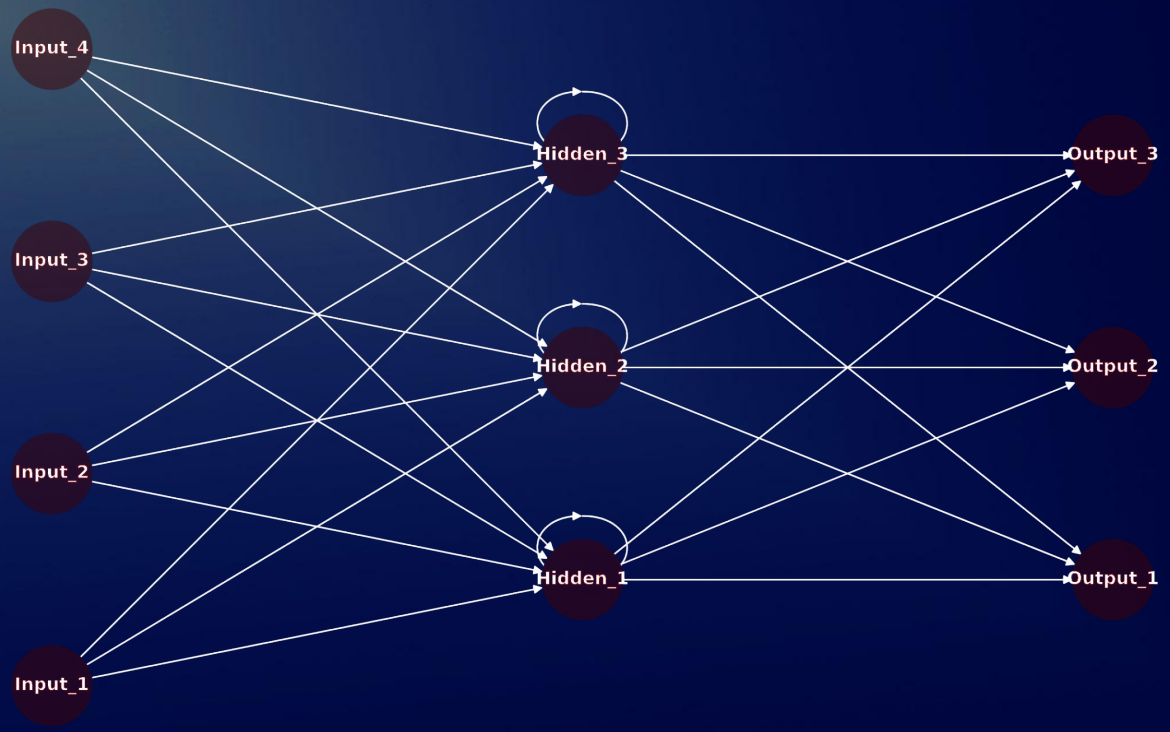


$$O = g\left(\sum_i w_i I_i\right)$$

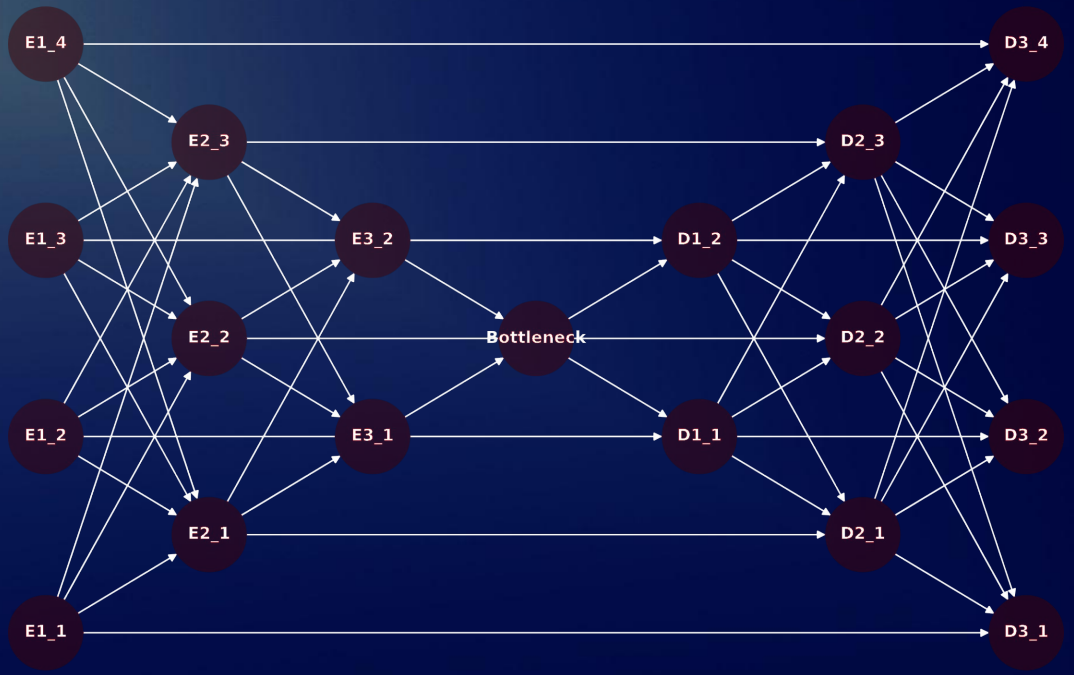
Neural Networks - Dense 4-6-6-3



Neural Networks - 4-3-3 Recurrent NN (RNN)

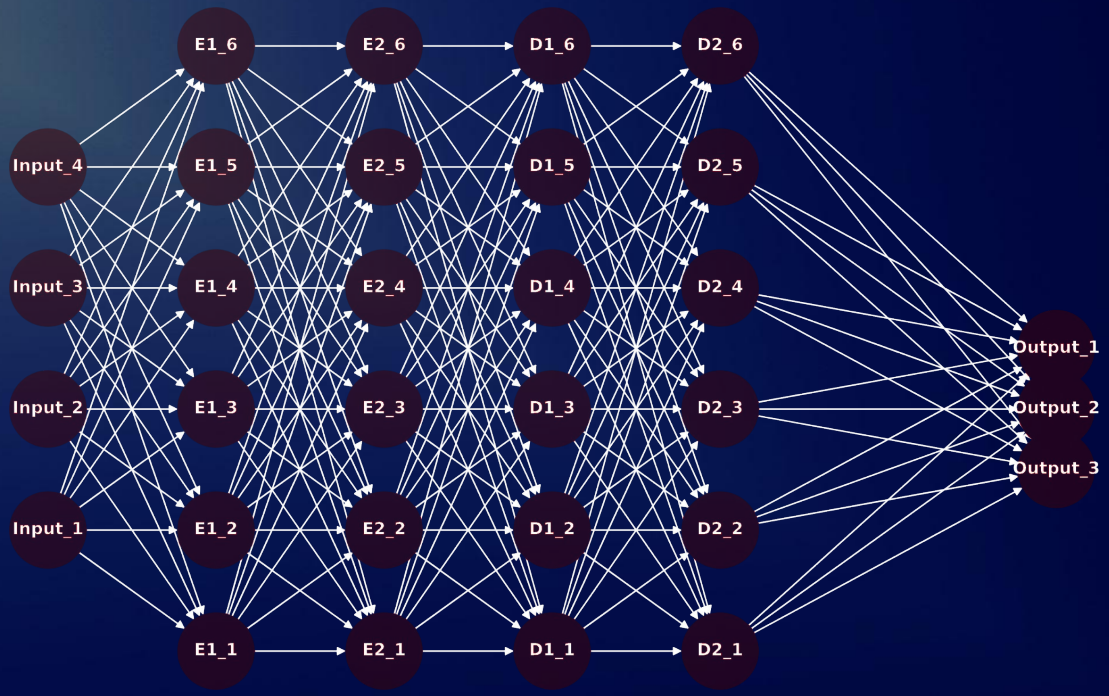


Neural Networks - U-Net



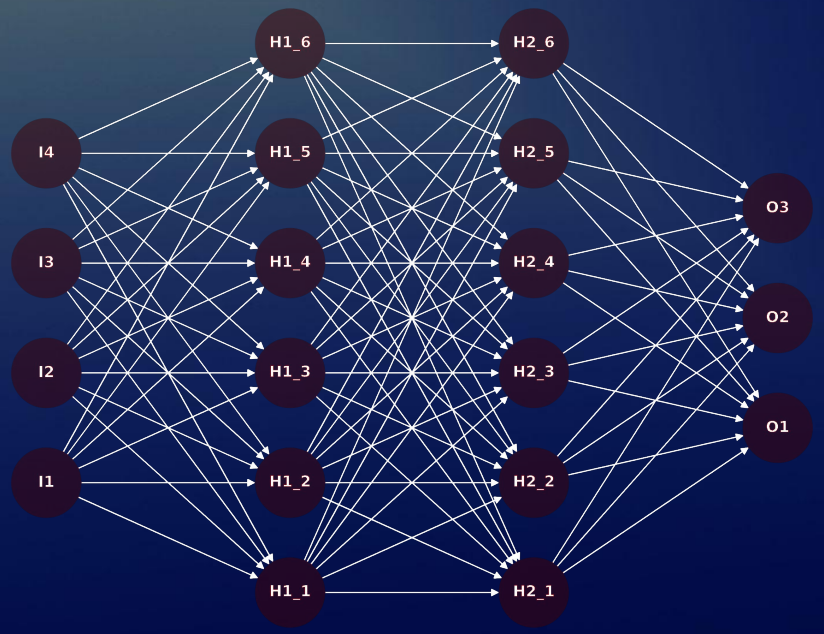
Ronneberger et al., 2015. U-net., MICCAI 2015 proceedings, part III 18 (pp. 234-241). Springer

Transformer Network



Vaswani, A., 2017. Attention is all you need. Advances in Neural Information Processing Systems.

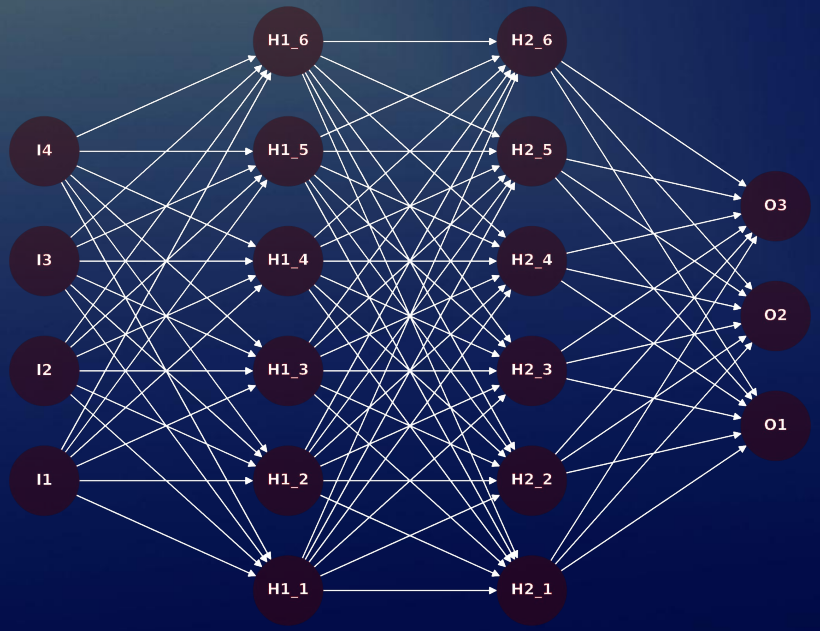
Training

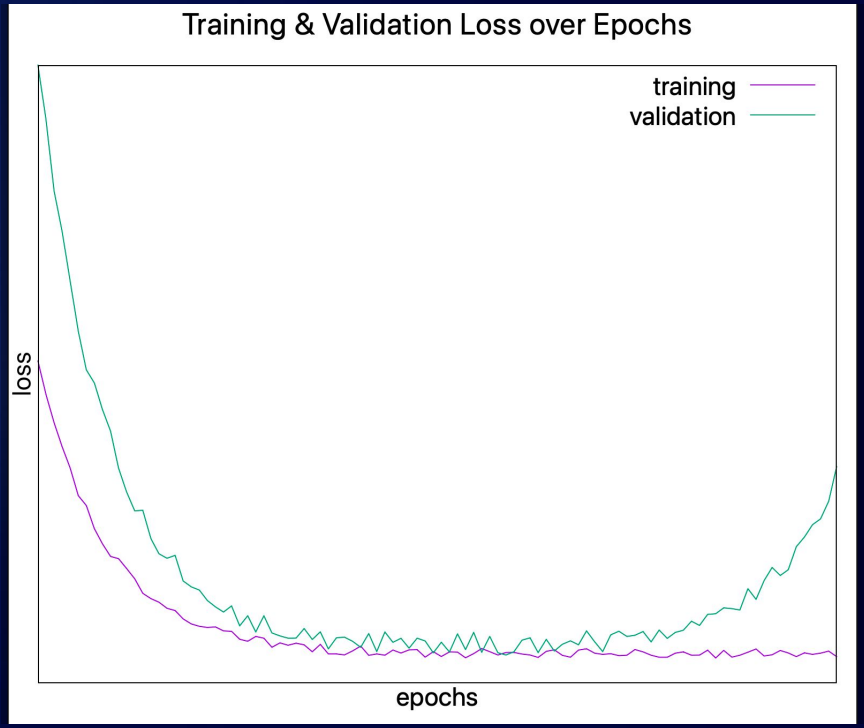


- Given:
  - training data
  - test data

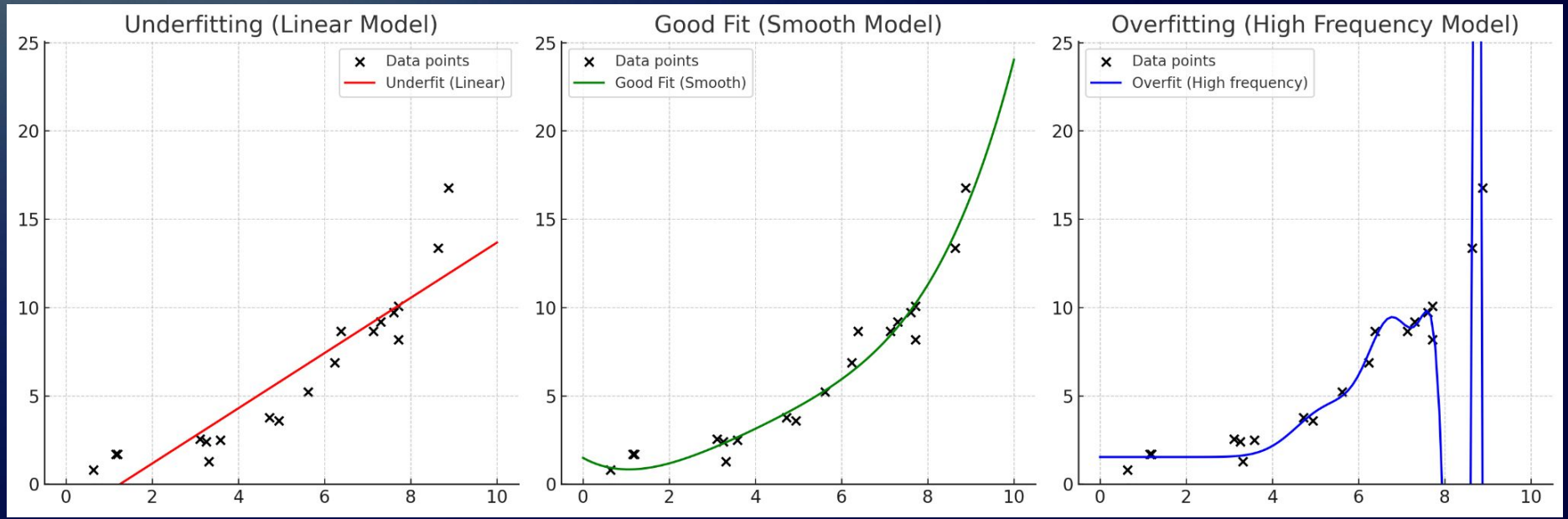
- initialize random weights
- optimize the weights:
  - repeat until satisfied:
    - compute loss
    - backpropagate to update weights

Training


I4
I3
I2
I1
H1_6
H1_5
H1_4
H1_3
H1_2
H1_1
H2_6
H2_5
H2_4
H2_3
H2_2
H2_1
O3
O2
O1


Training & Validation Loss over Epochs
training
validation
loss
epochs

Overfitting



ChatGPT (Generalized Pretrained Transformer)

Text → Response

- Multiple transformer blocks
- ChatGPT 3
  - *Brown, T.B., 2020. Language models are few-shot learners. arXiv preprint arXiv:2005.14165.*
  - 175 billion weights
  - 96 layers
  - The raw training set was 45TB of compressed plaintext. After filtering, it was ~570GB.
- ChatGPT 4
  - estimated 1 - 2 trillion weights

AI/ML in Facilities

- Synchrotron optimization
  - Leemann, S.C., Liu, S., Hexemer, A., Marcus, M.A., Melton, C.N., Nishimura, H. and Sun, C., 2019. Demonstration of machine learning-based model-independent stabilization of source properties in synchrotron light sources. Physical review letters, 123(19), p.194801.
- Beamline optimization
  - Morris, T.W., Rakitin, M., Giles, A., Lynch, J., Walter, A.L., Nash, B., Abell, D., Moeller, P., Pogorelov, I. and Goldring, N., 2022, October. On-the-fly optimization of synchrotron beamlines using machine learning. In Optical System Alignment, Tolerancing, and Verification XIV (Vol. 12222, pp. 171-175). SPIE.

NIST Center for Neutron Research

- Neutron Instrument Control Environment (NICE)
- 15 NICE instruments
- 585 total control parameters
- 818 experiment types
- Trajectory files need to be created to run the instruments
  - instrument-specific details
  - time consuming for users to master
- Given - Large amounts of documentation and example Trajectory files:

NIST Center for Neutron Research

User text
e.g.
“Create magik trajectory angleChecks that starts with a time of 10 and loops through sampleAngle from 2 to 10 in steps of 0.25”

Properly formatted trajectory file to control the instrument
e.g.

```
{"filePrefix": "mb111",
"init": [
["counter.countAgainst", "'TIME'"]],
"loops": [{
"vary": [
["sampleAngle", {"range": {"start": 2,
"step": 0.25, "stop": 10}}],
]}]
```

NIST Center for Neutron Research

User text
e.g.
"Create magik trajectory angleChecks that starts with a time of 10 and loops through sampleAngle from 2 to 10 in steps of 0.25"

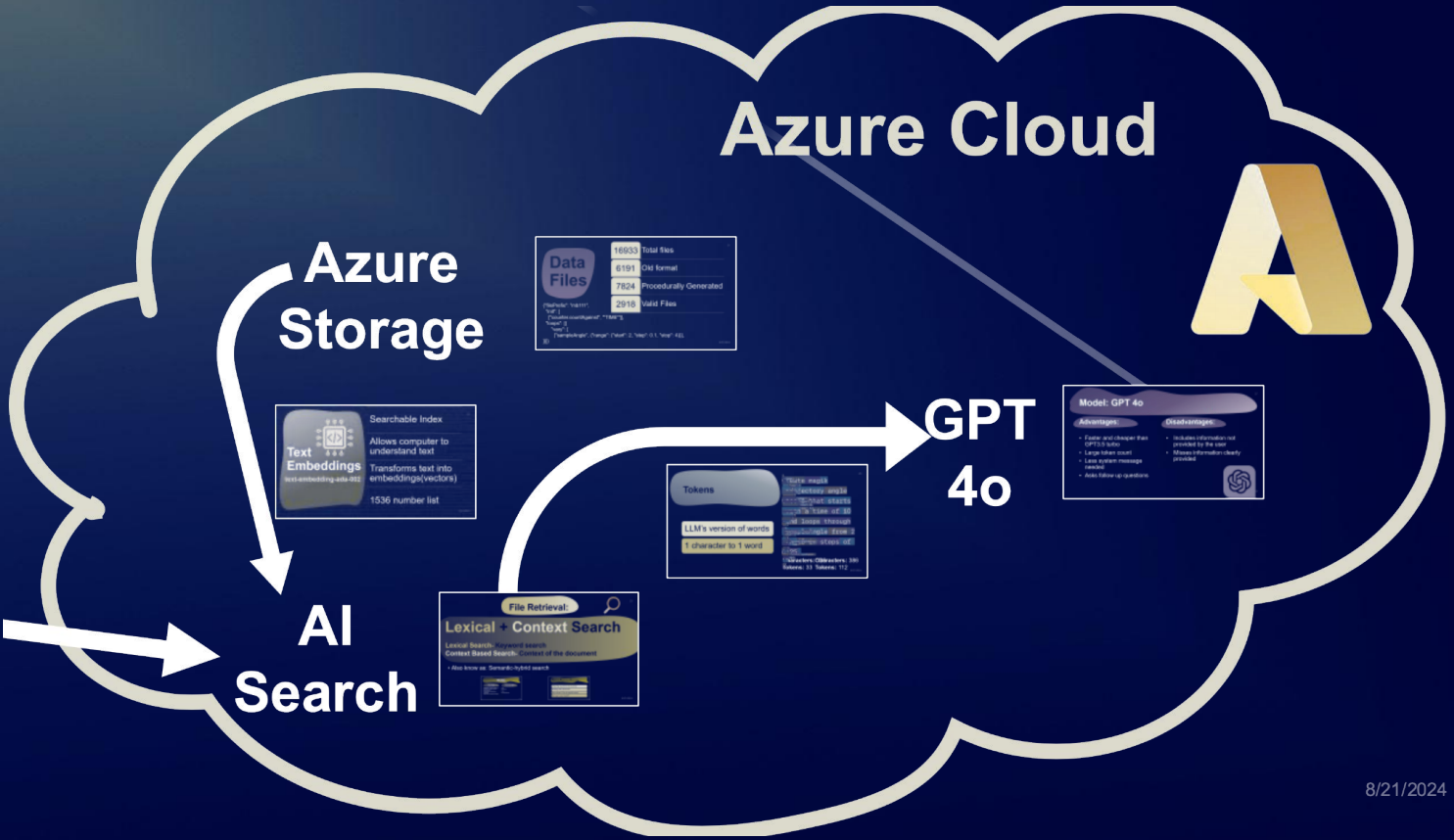
→

Properly formatted trajectory file to control the instrument
e.g.

```
{"filePrefix": "mb111",
"init": [
["counter.countAgainst", "'TIME'"]],
"loops": [{
"vary": [
["sampleAngle", {"range": {"start": 2,
"step": 0.25, "stop": 10}}],
]}]
```

Designing and training an LLM for this would be expensive!

Retrieval-Augmented Generation



8/21/2024

Credit: Jack Campbell, NIST, jack.campbell@nist.gov

Retrieval-Augmented Generation

- ~500 lines of code
- Data curation

Costs

$6.20 To create searchable database

Per Message $0.02

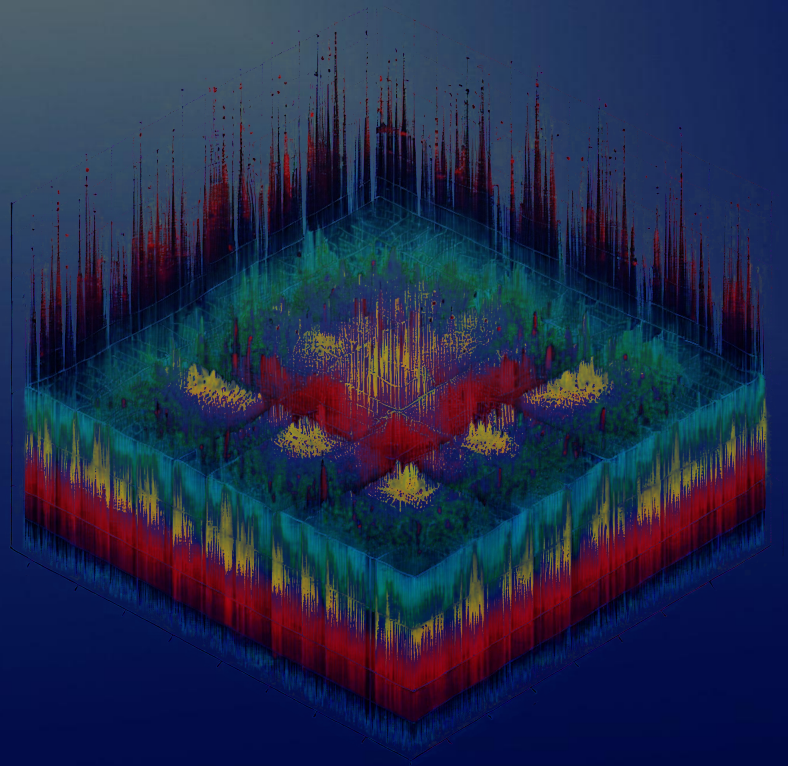
Per Month $96.09

$0.14 Per Conversation

$1,153 Per Year

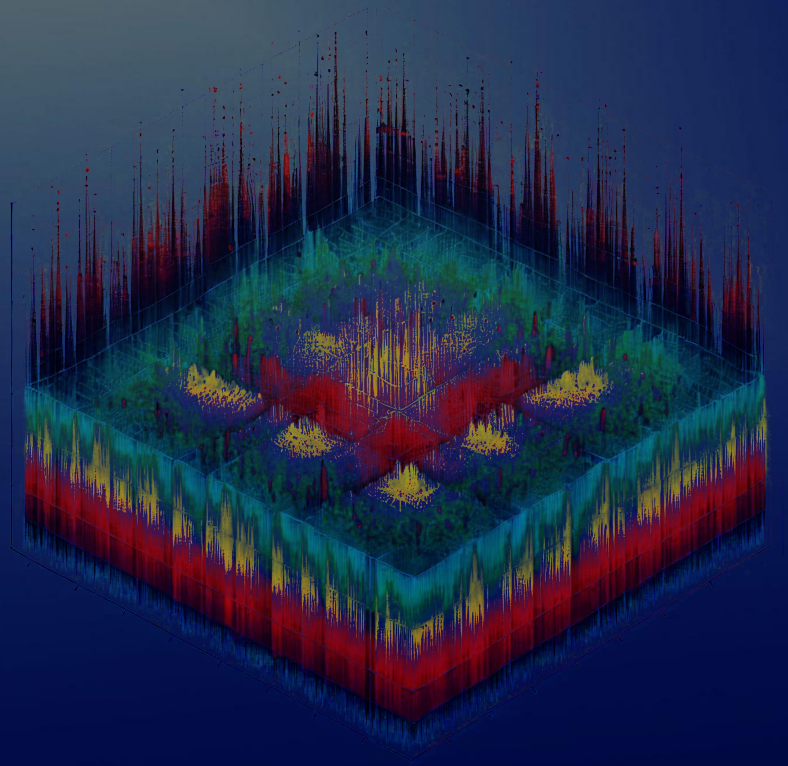
Credit: Jack Campbell, NIST, jack.campbell@nist.gov

A Common Science Problem



→ Relevant description of the data

A Common Science Problem



Relevant description of the data
e.g.

- parameters
- models
- categorization

AlphaFold / Google DeepMind

Sequence → Model

- v2 - Evoformer
- v3 - Pairformer

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

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]

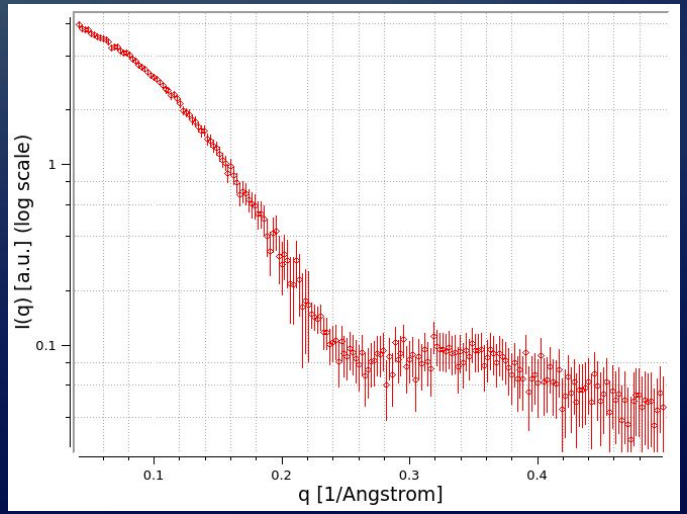
Article

Highly accurate protein structure prediction for the human proteome

https://doi.org/10.1038/s41586-021-03828-1
Received: 11 May 2021
Accepted: 16 July 2021
Published online: 22 July 2021
Open access

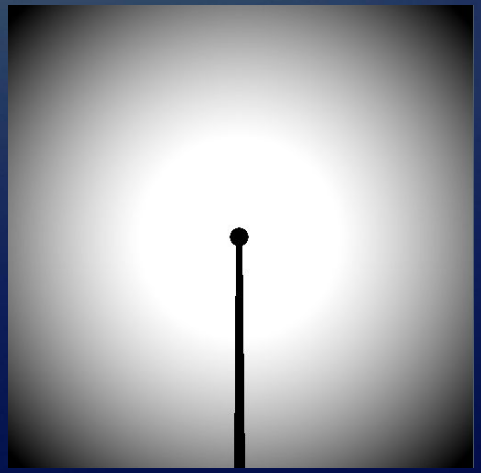
Kathryn Tunyasuvunakool[1], Jonas Adler[1], Zachary Wu[1], Tim Green[1], Michal Zielinski[1], Augustin Žídek[1], Alex Bridgland[1], Andrew Cowie[1], Clemens Meyer[1], Agata Laydon[1], Sameer Velankar[2], Gerard J. Kleywegt[2], Alex Bateman[2], Richard Evans[1], Alexander Pritzel[1], Michael Figurnov[1], Olaf Ronneberger[1], Russ Bates[1], Simon A. A. Kohl[1], Anna Potapenko[1], Andrew J. Ballard[1], Bernardino Romera-Paredes[1], Stanislav Nikolov[1], Rishub Jain[1], Ellen Clancy[1], David Reiman[1], Stig Petersen[1], Andrew W. Senior[1], Koray Kavukcuoglu[1], Ewan Birney[2], Pushmeet Kohli[1], John Jumper[1,3] & Demis Hassabis[1,3]

Biological SAS I(q)



- Quality assessment
- Rg, I(o)
- Dmax, P(r)
- Shape classification
- Ab initio model
- Informed model

Biological SAS I(q)



- Quality assessment
- Rg, I(o)
- Dmax, P(r)
- Shape classification
- Ab initio model
- Informed model

Machine Learning (ML) vs “Classical”

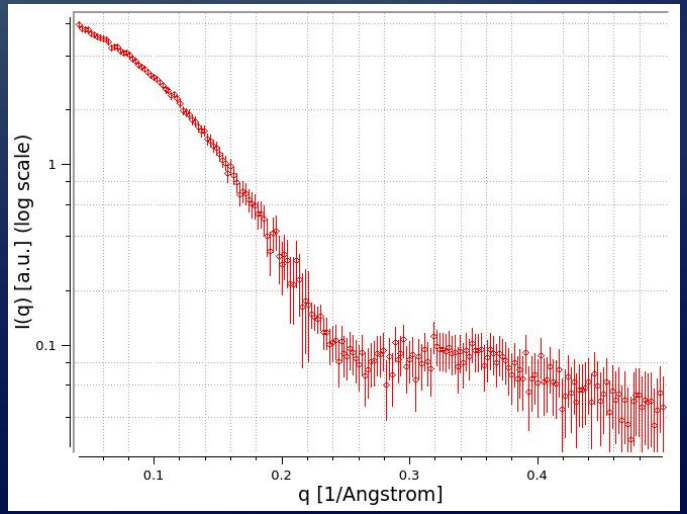
- ML typically requires good training & test data
  - and potentially large amounts
- ML can be fast
- If the data is outside of the range of the training & test data, results will likely be wrong
- Confidence of ML results is apparently still an open question
  - Papadopoulos, G., Edwards, P.J. and Murray, A.F., 2001. Confidence estimation methods for neural networks: A practical comparison. IEEE transactions on neural networks, 12(6), pp.1278-1287.
- ML can be a “black-box”

Machine Learning vs “Classical” - ChatGPT 4o

Tradeoffs for using machine learning vs classical approaches for modeling

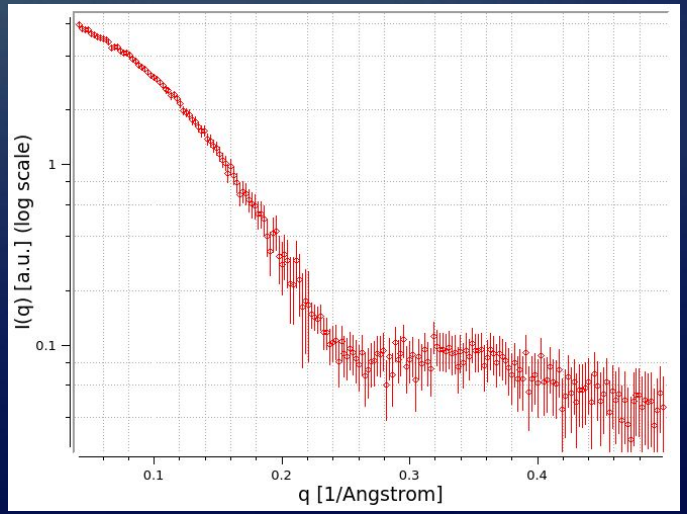
Category	Machine Learning	Classical Approach
Data Requirements	Requires large datasets	Works with limited data
Interpretability	Less interpretable (black-box)	Highly interpretable
Generalization	Good at interpolation	Good for extrapolation
Computational Complexity	Computationally intensive	Less computationally demanding
Flexibility	Highly flexible	Less flexible
Model Robustness	Can be less robust	More robust
Development Time	Longer development time	Faster development time
Scalability	Highly scalable	Less scalable across domains
Handling Nonlinearity	Great for handling nonlinearity	Limited handling of nonlinearity

SAS I(q)



- **Quality assessment**
- **Rg, I(o)**
- Dmax, P(r)
- Shape classification
- Ab initio model
- Informed model

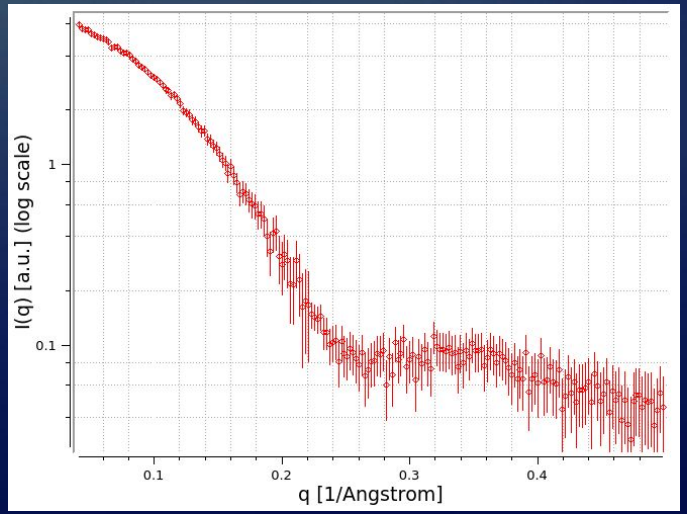
SAS I(q)



- **Quality assessment**
- **Rg, I(o)**
- Dmax, P(r)
- Shape classification
- Ab initio model
- Informed model

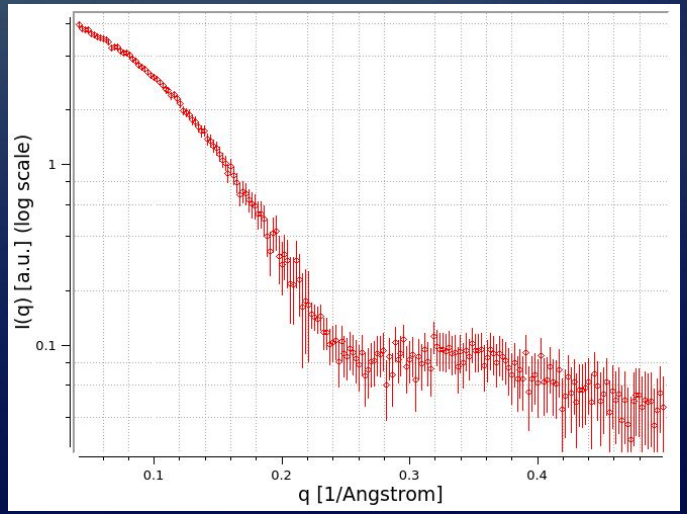
Prof. André Guinier
1911-2000
Orsay, France

SAS I(q)



- Quality assessment
- Rg, I(0)
- **Dmax, P(r)**
- Shape classification
- Ab initio model
- Informed model

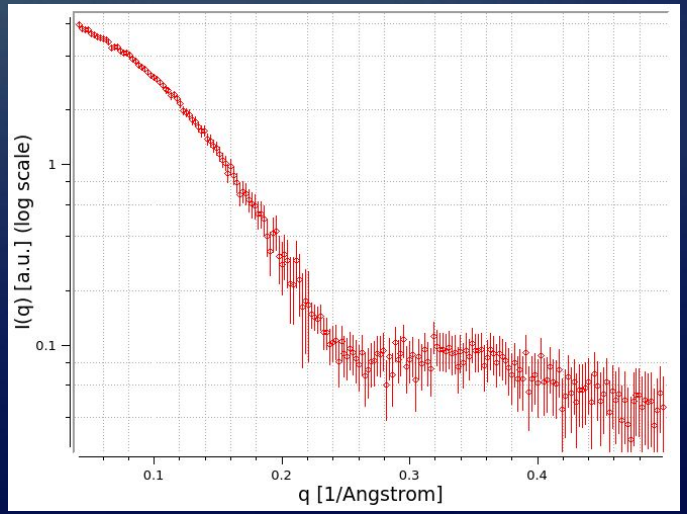
SAS I(q)



- Quality assessment
- Rg, I(0)
- **Dmax, P(r)**
- Shape classification
- Ab initio model
- Informed model

- Glatter, O. (1977) J. Appl. Cryst. 10, 415-421.
- GNOM - Svergun D.I. (1992) J. Appl. Cryst. 25, 495-503.
- Bayesian Fitting - Hansen, S. (2000) J. Appl. Cryst. 33, 1415-1421

SAS I(q)

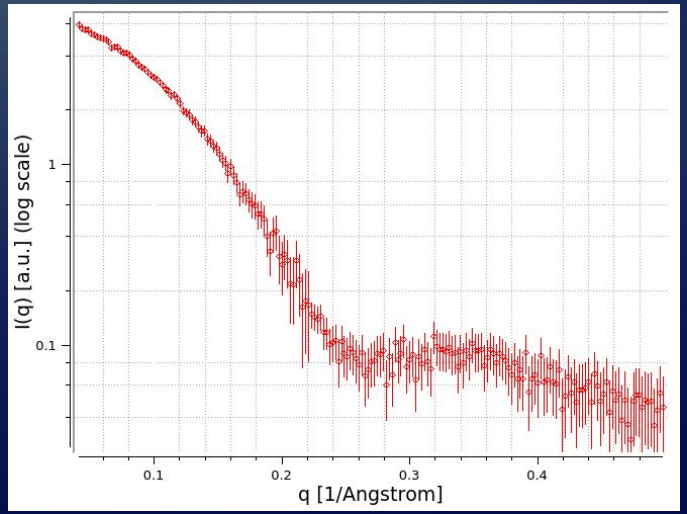


- Quality assessment
- Rg, I(o)
- Dmax, P(r)
- **Shape classification**
- Ab initio model
- Informed model

SAS I(q) - Shape Classification

- Classical:
  - P(r)
  - Kratky
  - Normalized Kratky
    - *Pérez, J., Vachette, P., Russo, D., Desmadril, M. and Durand, D., 2001. J. Mol. Bio., 308(4), pp.721-743.*
- ML:
  - *Franke, D., Jeffries, C.M. and Svergun, D.I., 2018. Machine learning methods for X-ray scattering data analysis from biomacromolecular solutions. Biophys. J. 114(11), pp.2485-2492.*
    - 3D feature vector space used for classification

SAS I(q)

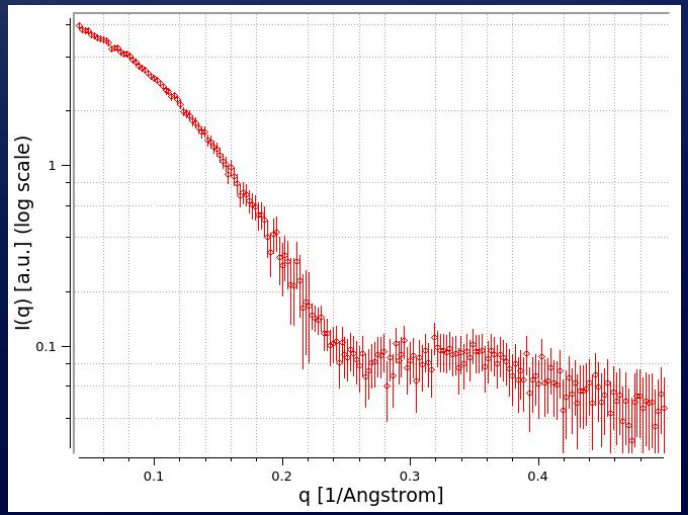


- Quality assessment
- Rg, I(o)
- Dmax, P(r)
- Shape classification
- **Ab initio model**
- Informed model

Information content in SAS curves

Svergun, D.I. & Koch, M.H.J. (2003) Small-angle scattering studies of biological macromolecules in solution. Rep. Prog. Phys. 66 1735-82

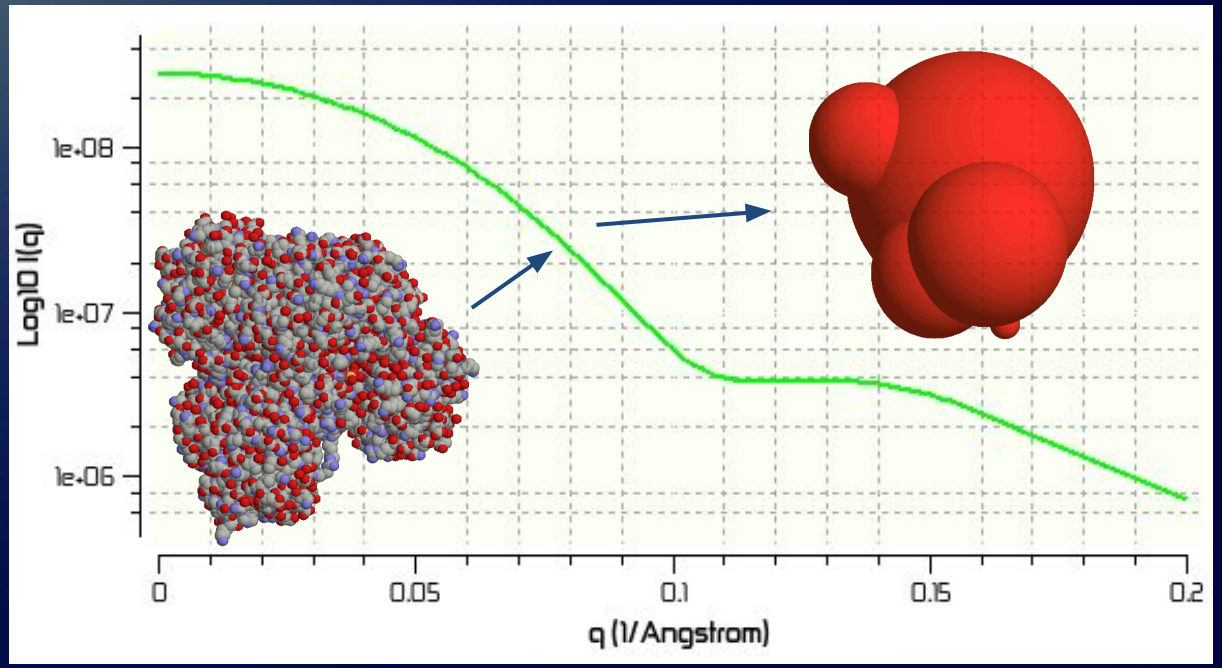
- Shannon channels = $D_{max} \cdot q\text{-}range / \pi$
- "the number of [obtainable parameters] typically does not exceed **10–15**"



Lyzosyme $D_{max} \sim 48$ Angstroms

Shannon channels = $48 * 0.5 / \pi \sim 8$

Parsimonious Models



Brookes, E., Parsimonious Spatial Models from Small Angle Scattering of Biological Macromolecules, SAS 2012, Sydney

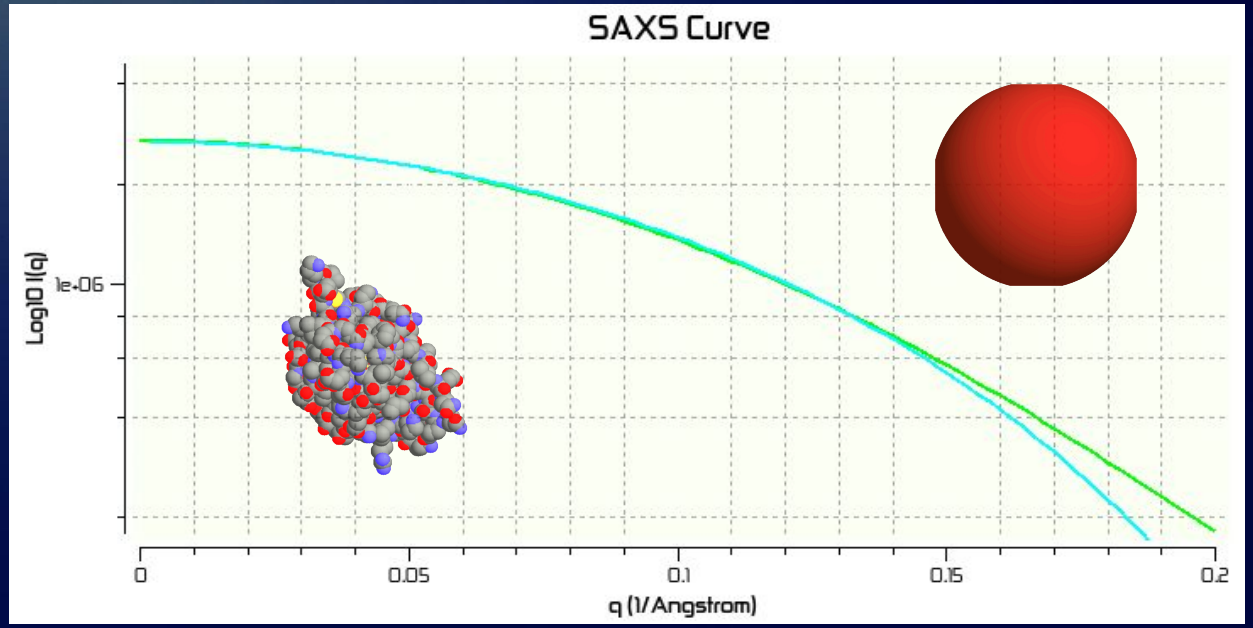
Parsimonious Models

PDB	MW in Daltons	Description
8RAT.PDB	13,683.87	CRYSTALLOGRAPHIC STUDIES OF THE PROTEIN RIBONUCLEASE-A
1A4V.PDB	14,152.00	ALPHA-LACTALBUMIN
1DWR.PDB	17,682.20	MYOGLOBIN (HORSE HEART) WILD-TYPE COMPLEXED WITH CO
1HCO.PDB	32,279.78	HUMAN CARBONMONOXY HAEMOGLOBIN
1BEB.PDB	35,305.26	BOVINE BETA-LACTOGLOBULIN
1CTS.PDB	49,129.58	CITRATE SYNTHASE
2CGA.PDB	51,318.72	BOVINE CHYMOTRYPSINOGEN
1GZX.PDB	64,575.52	OXY T STATE HAEMOGLOBIN: OXYGEN BOUND AT ALL FOUR HAEMS
5LDH.PDB	74,917.32	ACTIVE TERNARY COMPLEX OF PIG HEART LACTATE DEHYDROGENASE WITH S-LAC-NAD
2GD1.PDB	144,427.77	OXIDOREDUCTASE(ALDEHYDE(D)-NAD(A))
1GD1.PDB	147,077.69	HOLO-GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE FROM BACILLUS STEAROTHERMOPHILUS
1ADO.PDB	157,287.20	FRUCTOSE 1,6-BISPHOSPHATE ALDOLASE FROM RABBIT MUSCLE
1OVA.PDB	169,965.56	UNCLEAVED OVALBUMIN

Brookes, E., Parsimonious Spatial Models from Small Angle Scattering of Biological Macromolecules, SAS 2012, Sydney

Parsimonious Models

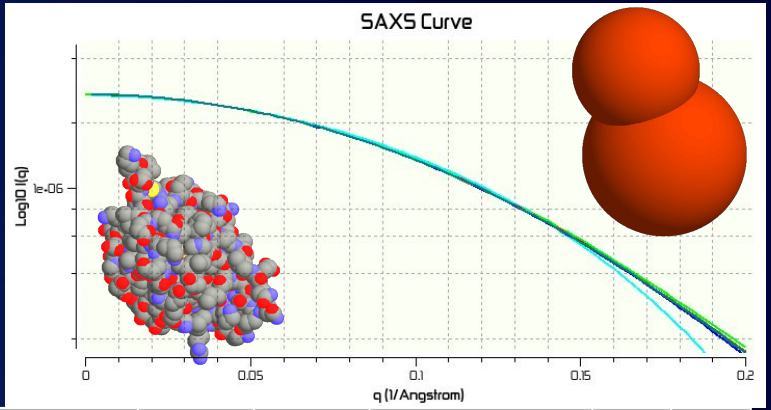
1A4V - 1 sphere



Brookes, E., Parsimonious Spatial Models from Small Angle Scattering of Biological Macromolecules, SAS 2012, Sydney

Parsimonious Models

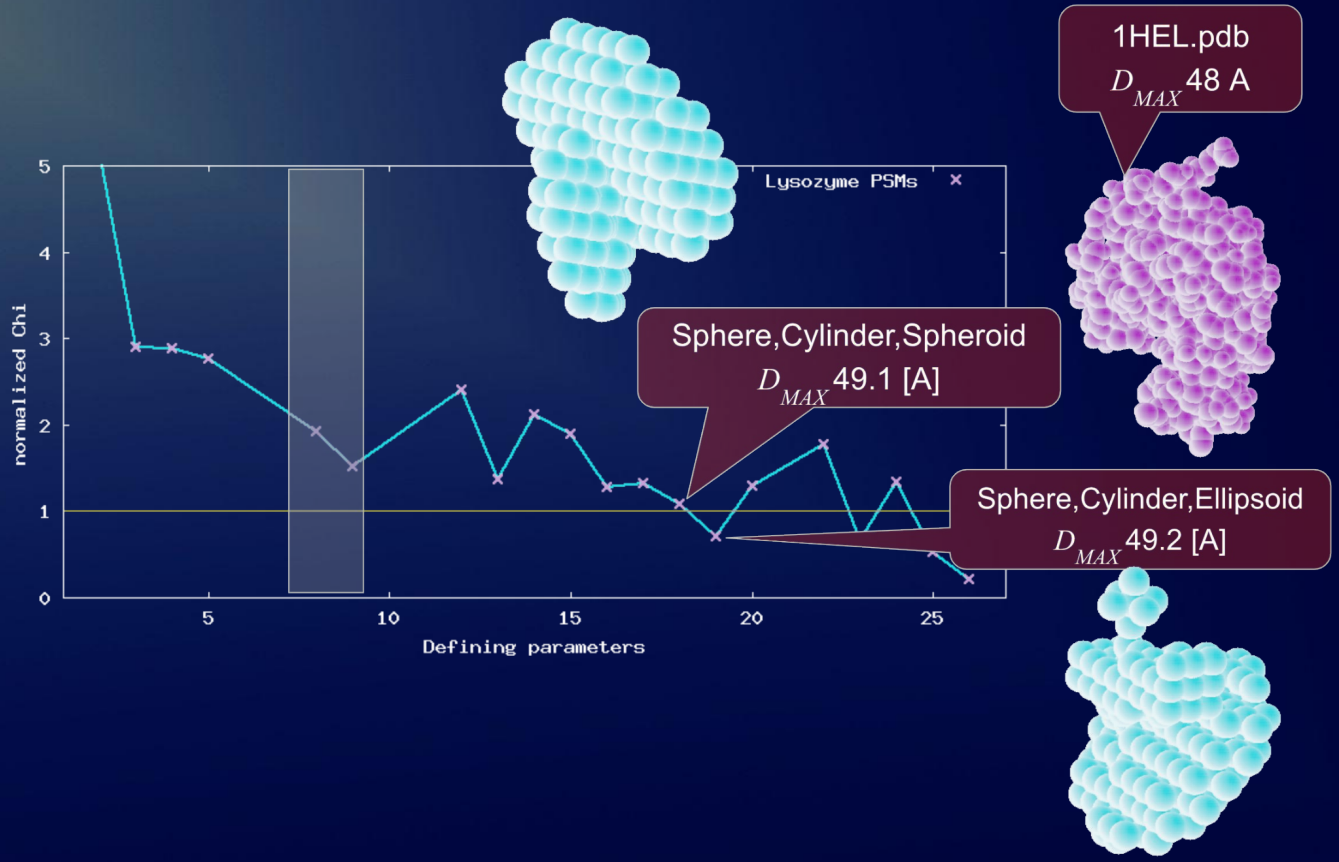
1A4V - 7 spheres



Model name	D(tr) [cm/sec^2]	Rg [nm]	Max extensions X [nm]	Y [nm]	Z [nm]	Axial ratios X:Z	X:Y	Y:Z
1A4V_1-db_1sa-10_1	1.210e-06	1.37	3.54	3.54	3.54	1.00	1.00	1.00
1A4V_1-db_2sa-10_1	1.200e-06	1.45	4.74	3.15	3.15	1.50	1.50	1.00
1A4V_1-db_3sa-10_1	1.140e-06	1.52	4.69	3.50	2.94	1.60	1.34	1.19
1A4V_1-db_4sa-10_1	1.170e-06	1.48	4.67	3.15	3.08	1.51	1.48	1.02
1A4V_1-db_5sa-10_1	1.140e-06	1.51	4.70	4.01	3.23	1.46	1.17	1.24
1A4V_1-db_6sa-10_1	1.140e-06	1.50	4.87	3.69	3.28	1.49	1.32	1.12
1A4V_1-db_7sa-10_1	1.160e-06	1.49	4.74	3.21	3.21	1.48	1.48	1.00
1A4V_1-so	1.137e-06	1.48	5.67	3.54	3.36	1.69	1.60	1.05

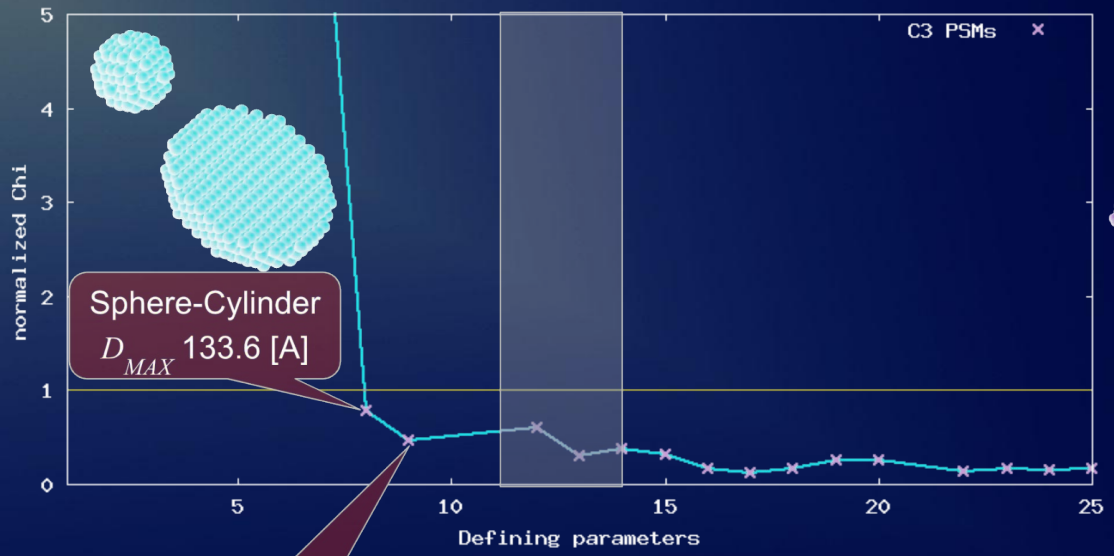
Brookes, E., Parsimonious Spatial Models from Small Angle Scattering of Biological Macromolecules, SAS 2012, Sydney

Parsimonious Modeling



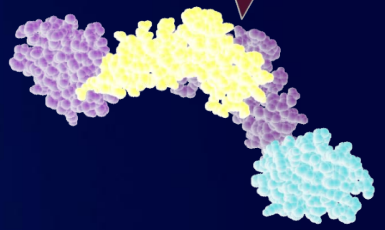
Brookes, E., Progress in Parsimonious Spatial Modeling of Biological SAS Experimental Data, ACA 2014, Albuquerque

Parsimonious Modeling

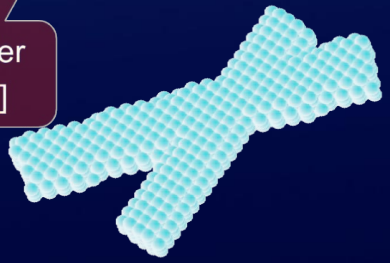


Sphere-Cylinder
D_{MAX} 133.6 [A]

Homology Model
D_{MAX} 142 [A]

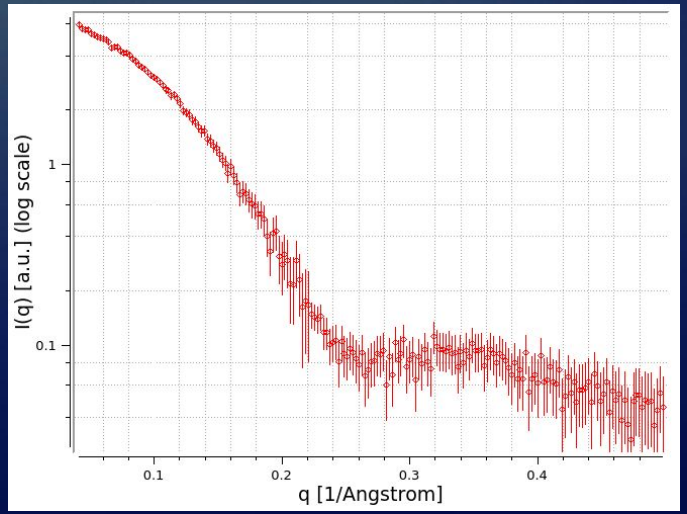


Cylinder-Cylinder
D_{MAX} 138.6 [A]



Brookes, E., Progress in Parsimonious Spatial Modeling of Biological SAS Experimental Data, ACA 2014, Albuquerque

SAS I(q)

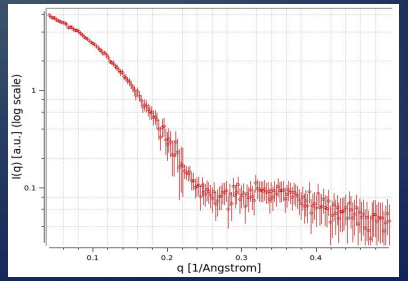


- Quality assessment
- Rg, I(o)
- Dmax, P(r)
- Shape classification
- **Ab initio model**
- Informed model

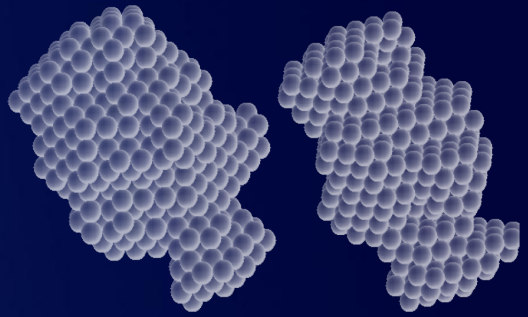
Ab Initio Models

DAMMIN

D. I. Svergun (1999) Biophys J. 2879-2886.

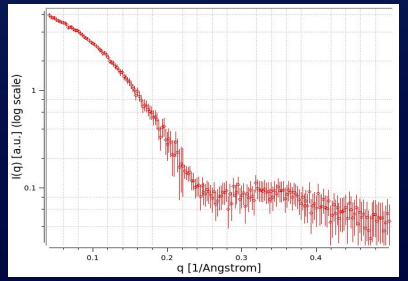


→

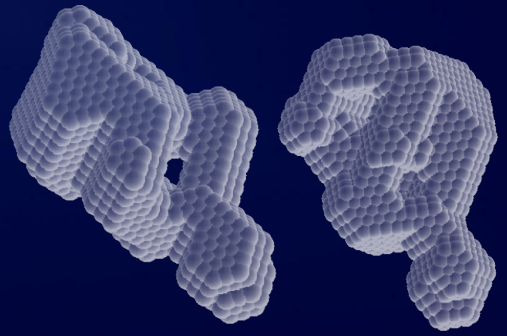


DAMMIF

Franke, D. and Svergun, D.I. (2009) J. Appl. Cryst., 42, 342-346.



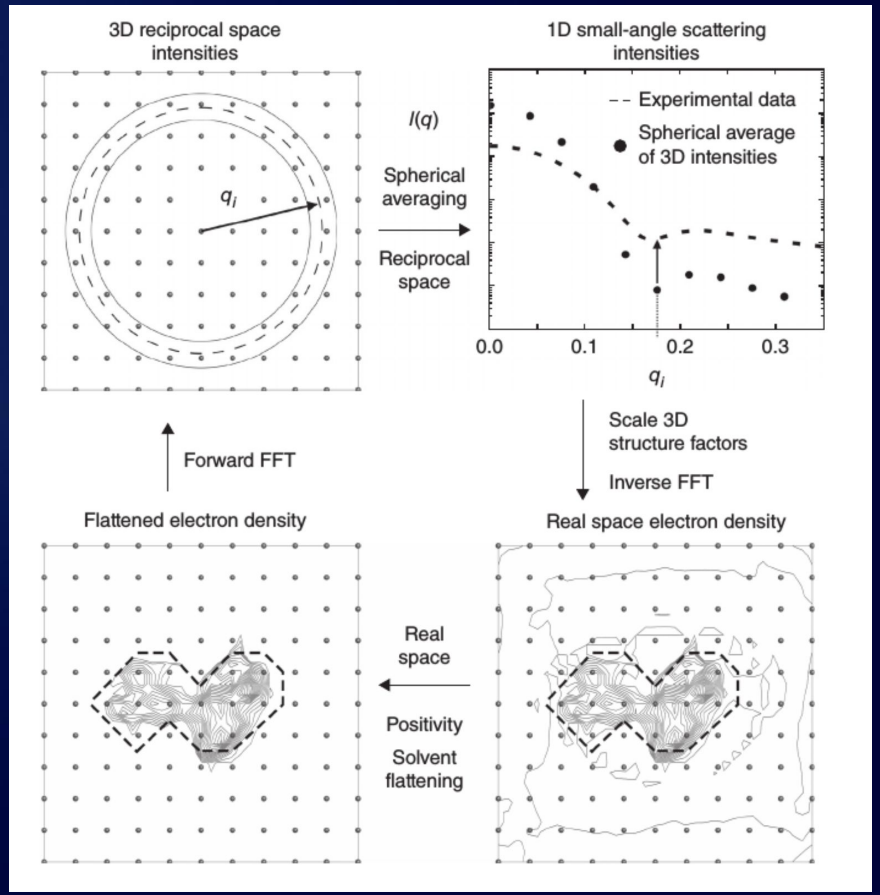
→



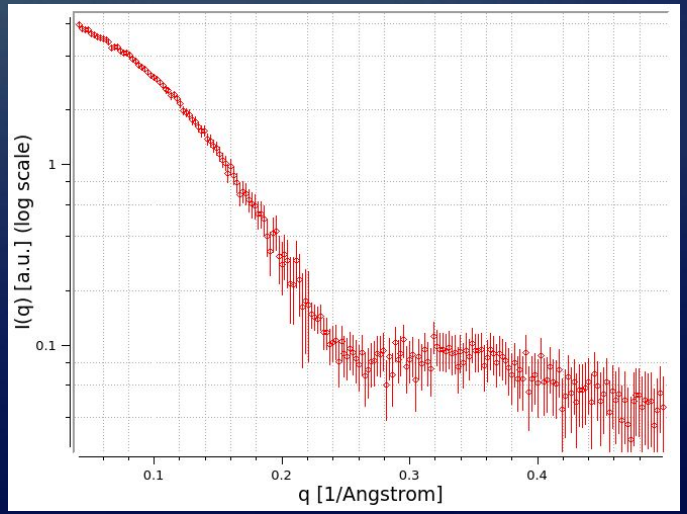
Ab Initio Models

DENSS

T. D. Grant (2018) Nat. Meth. 15:3 191-193



SAS I(q)

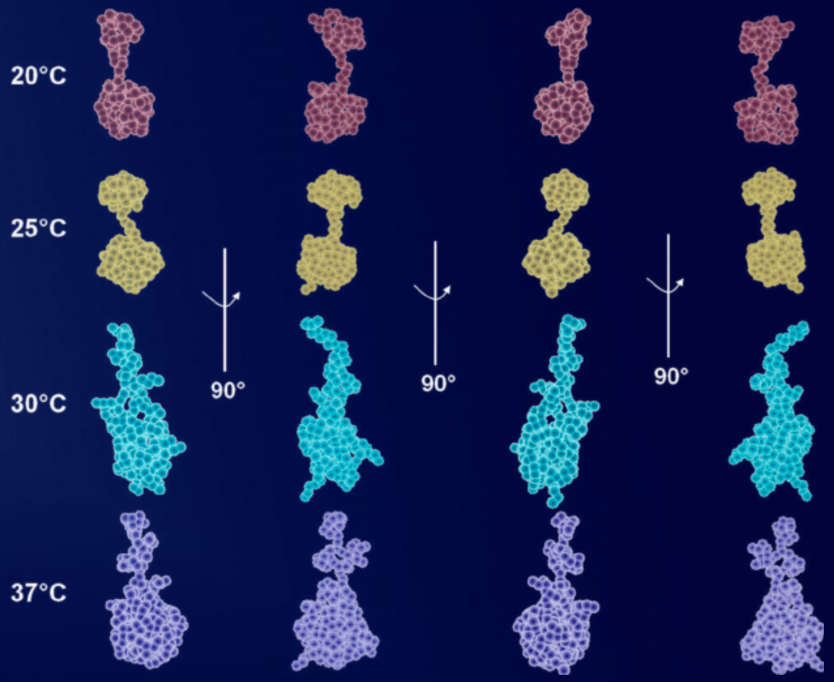


- Quality assessment
- Rg, I(o)
- Dmax, P(r)
- Shape classification
- Ab initio model
- **Informed model**

GASBOR

Svergun, D.I. et al. (2001) Biophys. J., 80, 2946-2953.

ab initio reconstruction of protein structure
by a chain-like ensemble of dummy residues



Various views of the ab initio 3D models obtained using GASBOR and by averaging ten single models for each sample by using DAMAVER for rSdrFB1-4 at different temperatures.

Dipoto, Antonella et al., (2015). Appl. Microbiology and Biotech.

Information Content Revisited

Jochen S Hub. Curr. Op. in Struct. Bio. 2018, 49:18-26

"the interpretation of solution scattering data by computational methods is complicated by ***the low information content of the data***, by scattering contributions from the hydration layer, and by unknown systematic errors."

"***The physical information in atomistic force fields complements the low-information SWAXS data***; explicit-solvent MD may be used to predict solvent scattering, and the MD-related sampling methods may guide the structure refinement against SWAXS data."

"Because SWAXS curves are smooth and one-dimensional (1D), they contain quite a limited amount of information. How the information is distributed over the q-range is a matter of ongoing research, but ***it is generally accepted that experimental SWAXS curves do not contain more than 10–30 independent data points***. Hence, ***the number of backbone angles of biomolecules exceeds the number of independent data points of SWAXS curves by roughly two orders of magnitude***. This precludes any straightforward fitting of protein structures against SWAXS data, but instead it leads to a ***high risk of overfitting***."

Starting Structure + Experimental Data → Representative structure(s) that are Consistent with the Experimental Data

- Create a pool of structures from the starting structure
  - Molecular Dynamics
  - Monte Carlo
  - etc.
- Compute simulated data on each structure
- Compare the simulated data with the experimental data to choose representative structures
  - Best Fit
  - Least Squares
  - etc.

Starting Structure + Experimental Data → Representative structure(s) that are Consistent with the Experimental Data

- Create a pool of structures from the starting structure
  - Molecular Dynamics
  - Monte Carlo
  - etc.
- Compute simulated data on each structure
- Compare the simulated data with the experimental data to choose representative structures
  - Best Fit
  - Least Squares
  - etc.

The resulting representative structures can only be said to be consistent with the data.

AlphaFold / Google DeepMind

Sequence → Model

- v2 - Evoformer
- v3 - Pairformer

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

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]

Article

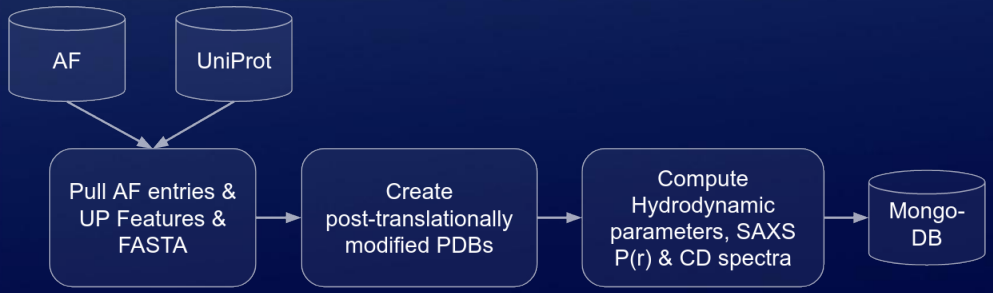
Highly accurate protein structure prediction for the human proteome

https://doi.org/10.1038/s41586-021-03828-1
Received: 11 May 2021
Accepted: 16 July 2021
Published online: 22 July 2021
Open access

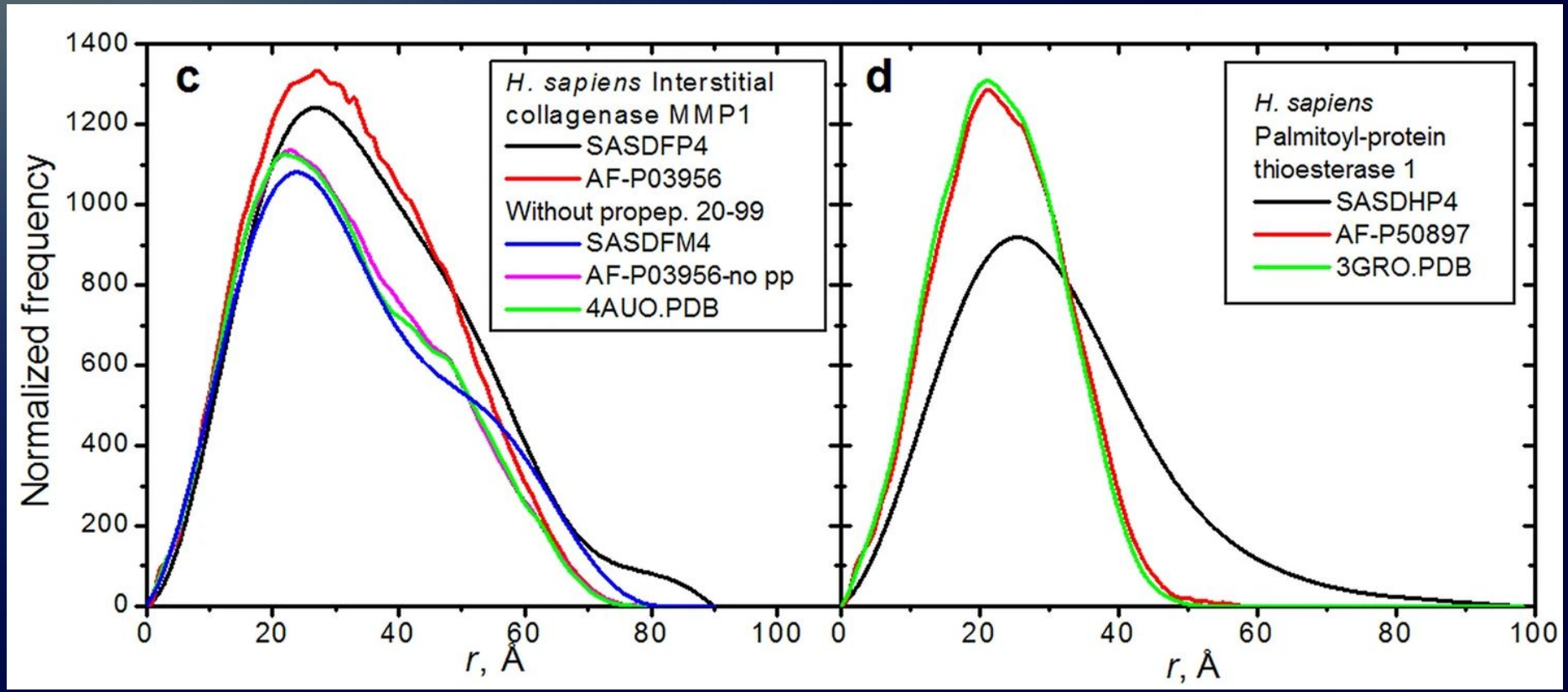
Kathryn Tunyasuvunakool[1], Jonas Adler[1], Zachary Wu[1], Tim Green[1], Michal Zielinski[1], Augustin Žídek[1], Alex Bridgland[1], Andrew Cowie[1], Clemens Meyer[1], Agata Laydon[1], Sameer Velankar[2], Gerard J. Kleywegt[2], Alex Bateman[2], Richard Evans[1], Alexander Pritzel[1], Michael Figurnov[1], Olaf Ronneberger[1], Russ Bates[1], Simon A. A. Kohl[1], Anna Potapenko[1], Andrew J. Ballard[1], Bernardino Romera-Paredes[1], Stanislav Nikolov[1], Rishub Jain[1], Ellen Clancy[1], David Reiman[1], Stig Petersen[1], Andrew W. Senior[1], Koray Kavukcuoglu[1], Ewan Birney[2], Pushmeet Kohli[1], John Jumper[1,3] & Demis Hassabis[1,3]

US-SOMO AlphaFold Database

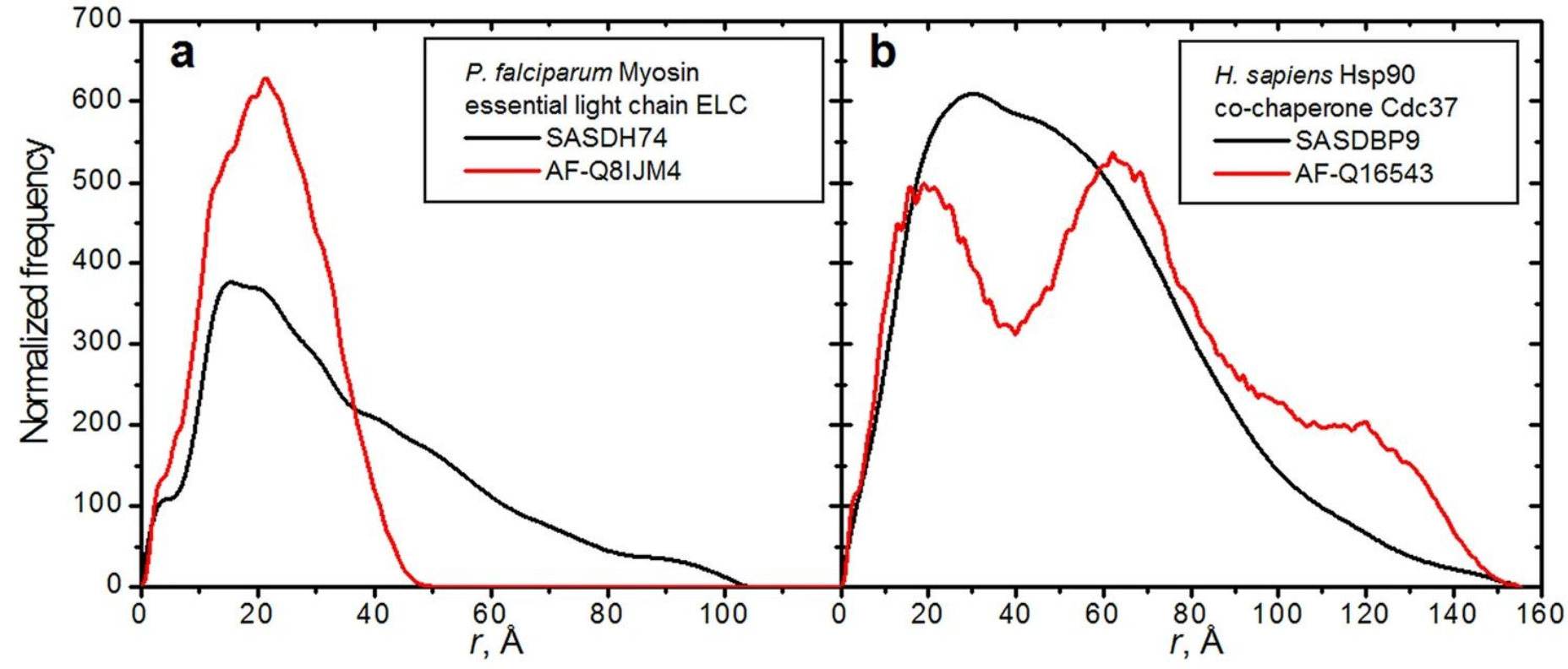
- The AlphaFold structures were predicted directly from the UniProt sequences, without any curing regarding post-translational modifications
  - Based on the UniProt annotations, removed the Initiator Methionine, Signal Peptide, and Transit Peptide(s) from the AlphaFold structures. Permuted structures with & without Propeptide(s) were also generated
- Post-translationally modified the entire AlphaFold v2 database & computed hydrodynamic, structural - incl. SAXS P(r), & circular dichroism calculations, (~1M structures)
- *https://somo.genapp.rocks*



US-SOMO AlphaFold Database



US-SOMO AlphaFold Database



JOURNAL OF APPLIED CRYSTALLOGRAPHY

Volume 56 | Part 4 | August 2023 | Pages 910-926
https://doi.org/10.1107/S1600576723005344
OPEN ACCESS
Cited by 8

ISSN: 1600-5767

AlphaFold-predicted protein structures and small-angle X-ray scattering: insights from an extended examination of selected data in the Small-Angle Scattering Biological Data Bank

Emre Brookes,[a*] Mattia Rocco,[b] Patrice Vachette[c] and Jill Trewhella[d*]

A Fast Ensemble Modeling Method Optimizing the Fit of Protein Structures with Flexible Regions to SAXS Data

- E Brookes, M Rocco, P Vachette, J Trewella
- Inputs - Experimental data & a predicted structure, e.g. AlphaFold etc.
- Outputs - A representative ensemble
- "FAST" - NNLS fits on P(r) from MC derived pool
  - J Curtis et al. - Monomer & Complex Monte Carlo
  - Currently identifying regions of flexibility by confidence levels or user supplied
- Eventually refined in I(q) space (e.g. WAXSiS)

JOURNAL OF APPLIED CRYSTALLOGRAPHY

Volume 56 | Part 4 | August 2023 | Pages 910-926
https://doi.org/10.1107/S1600576723005344
OPEN ACCESS
Cited by 8

ISSN: 1600-5767

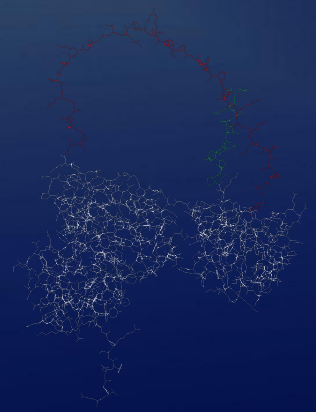
AlphaFold-predicted protein structures and small-angle X-ray scattering: insights from an extended examination of selected data in the Small-Angle Scattering Biological Data Bank

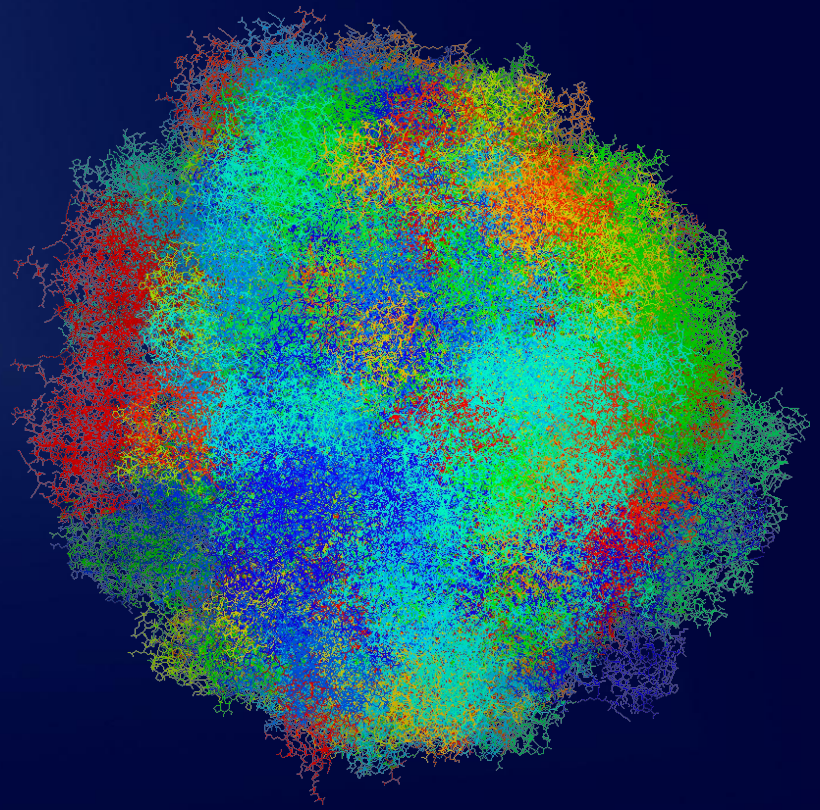
Emre Brookes,[a*] Mattia Rocco,[b] Patrice Vachette[c] and Jill Trewhella[d*]

A Fast Ensemble Modeling Method Optimizing the Fit of Protein Structures with Flexible Regions to SAXS Data

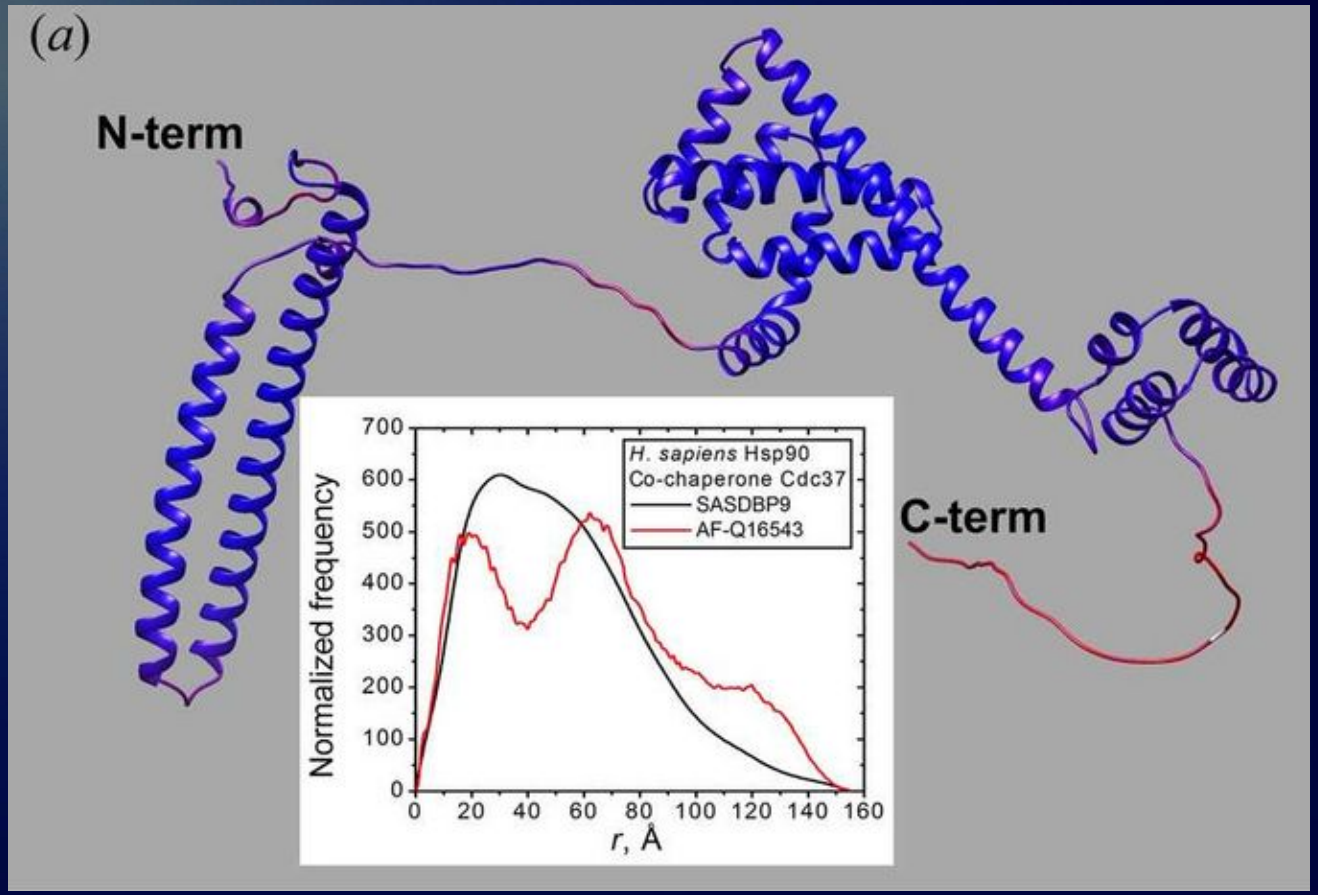
- Create a pool of structures from the starting AlphaFold (or other) structure
  - User selects flexible regions, possibly informed from AlphaFold confidence levels
  - Torsion Angle Monte Carlo
- Compute simulated data on pool of structures
  - CRYSOL
  - on selected structures - WAXSiS
    - Knight, C.J. and Hub, J.S., 2015. WAXSiS.,Nucleic acids research, 43(W1), pp.W225-W230.
- Compare the simulated data with the experimental data to choose representative structures
  - Non-negatively constrained Least Squares (NNLS)

Torsion Angle Monte Carlo on AF-Q15113

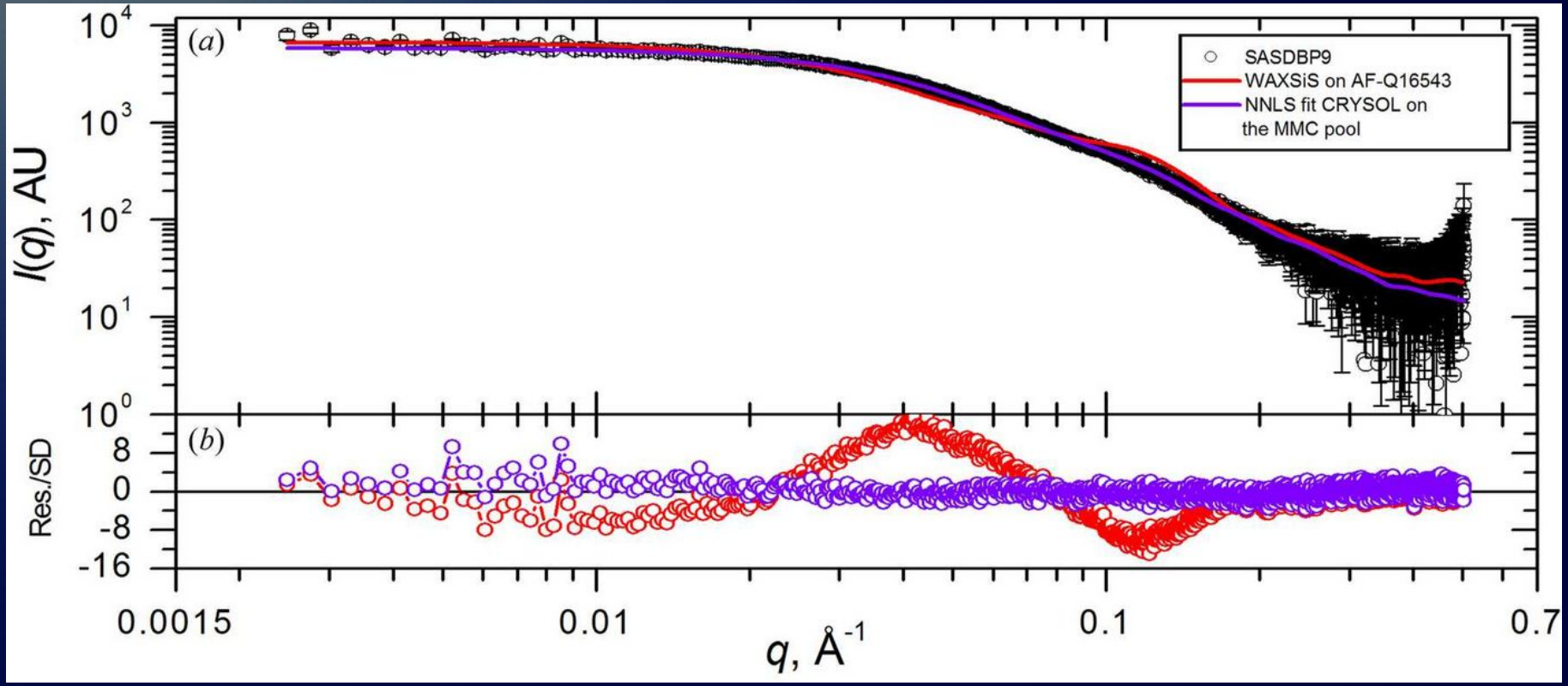




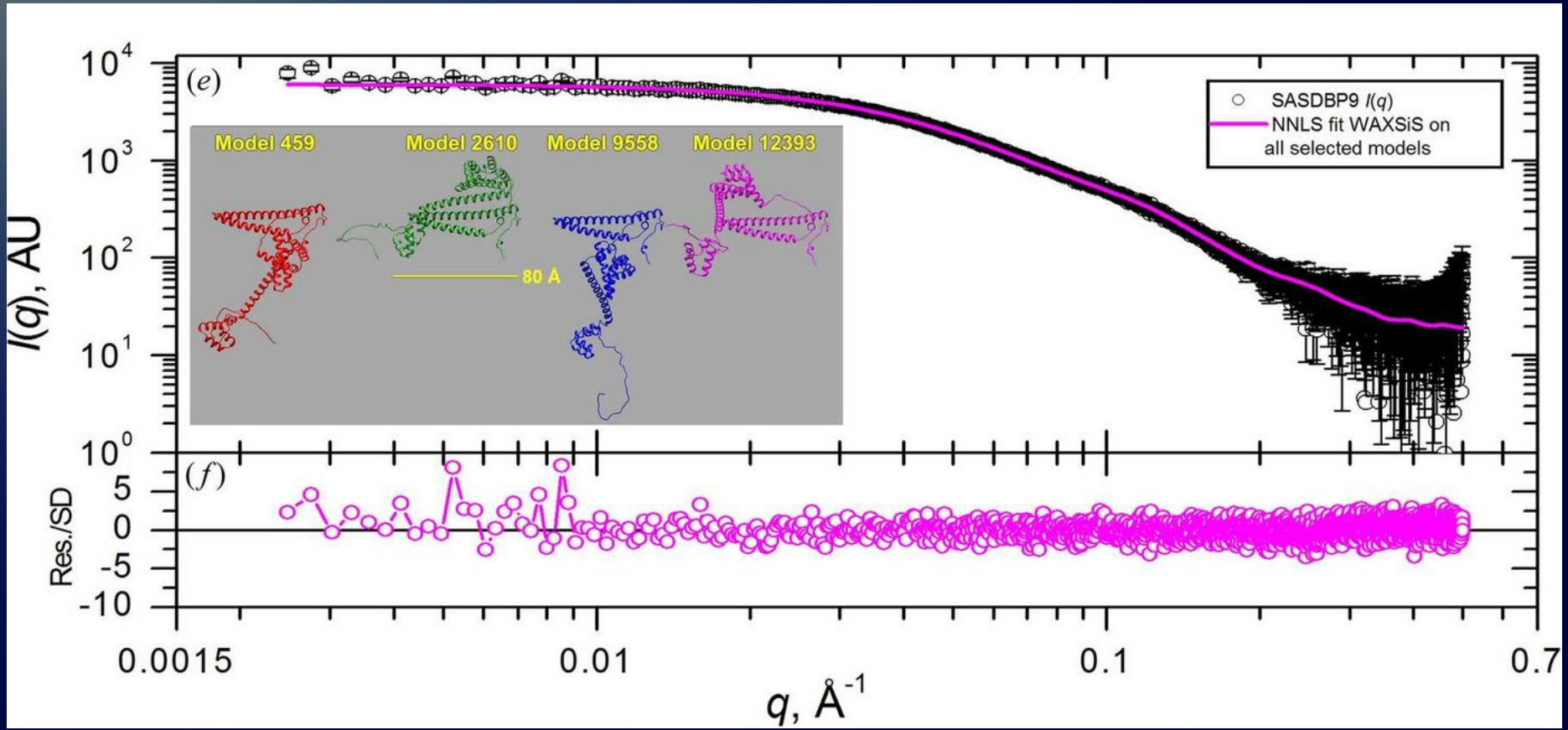
Curtis et al., 2012. SASSIE, Comp. Phy. Comm., 183(2), pp.382-389.
Perkins et al., 2016. J. App. Cryst., 49(6), pp.1861-1875.



Brookes et al., 2023. J Appl Cryst, 56(4), pp.910-926.



Brookes et al., 2023. J Appl Cryst, 56(4), pp.910-926.



Brookes et al., 2023. J Appl Cryst, 56(4), pp.910-926.

Conclusions

- AI is a powerful tool for modeling, but is not always the best choice
  - Maslow's law of the instrument:
    - *"If the only tool you have is a hammer, you tend to see every problem as a nail."*
- SAS I(q) data has limited information content, so, IMO, not likely a useful direct target for AI
- AI can be helpful for managing the beamline and providing user help
- AI derived structural models can provide starting structures for comparison with experimental data
- AlphaFold predicted structures do not always match solution SAS experimental data

Acknowledgments

Special Thanks!

- Frank Gabel

Discussions

- Jesse Hopkins
- Robert Rambo
- Tim Snow

Collaborators

- Joseph Curtis
- Susan Krueger
- Andreas Larsen
- Javier Pérez
- Mattia Rocco
- Jill Trewhella
- Patrice Vachette

Infrastructure Support

- Brian Beck
- Jeremy Fischer
- Zach Graber
- Danny Havert
- Mike Lowe
- Suresh Marru
- Mark Perri
- George Turner

Funding

- NIH K25GM090154, GM120600
- NSF CHE-1265817, OAC-1740097, OAC-1912444
- NSF XSEDE TG-MCB140255, TG-MCB170057
- NSF ACCESS MCB170057

Funded by a grant from the National Institute of General Medical Sciences of the National Institutes of Health

Thank you for listening!

Questions?

Publications

- *Wright DW, Nan R, Hui G, Curtis JE, Brookes EH, Perkins SJ. CCP-SAS - Novel Approaches for the Atomistic Modeling of Small Angle Scattering Data in Biology. Biophysical journal. 2015 January 27; 108(2):191a.*
- *Brookes E, Rocco M. A database of calculated solution parameters for the AlphaFold predicted protein structures. Scientific reports. 2022 May 5;12(1):7349.*
- *Brookes EH, Rocco M. Beyond the US-SOMO-AF database: a new website for hydrodynamic, structural, and circular dichroism calculations on user-supplied structures. European Biophysics Journal. 2023 Jul;52(4):225-32.*
- *Brookes, E., Rocco, M., Vachette, P. and Trewhella, J., 2023. AlphaFold-predicted protein structures and small-angle X-ray scattering: insights from an extended examination of selected data in the Small-Angle Scattering Biological Data Bank. Journal of Applied Crystallography, 56(4), pp.910-926.*