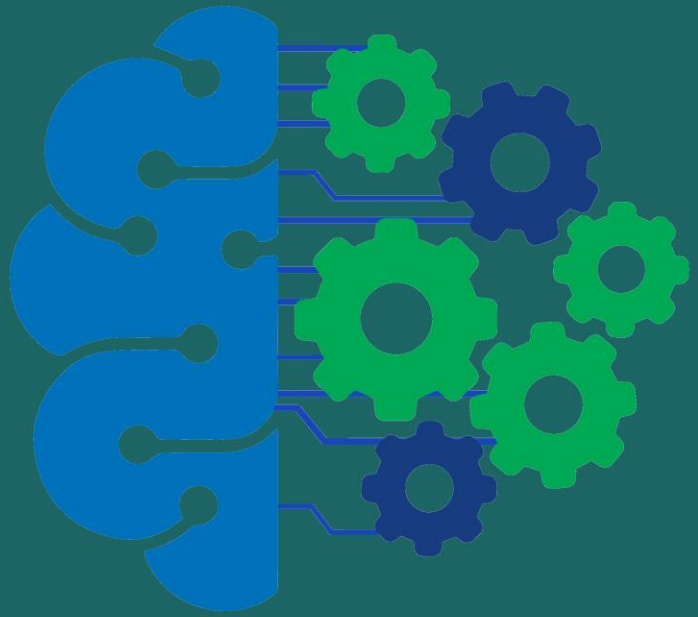
Neural networks for SAXS data analysis



Dima Molodenskiy, Al Kikhney, Dmitri Svergun
EMBL Hamburg

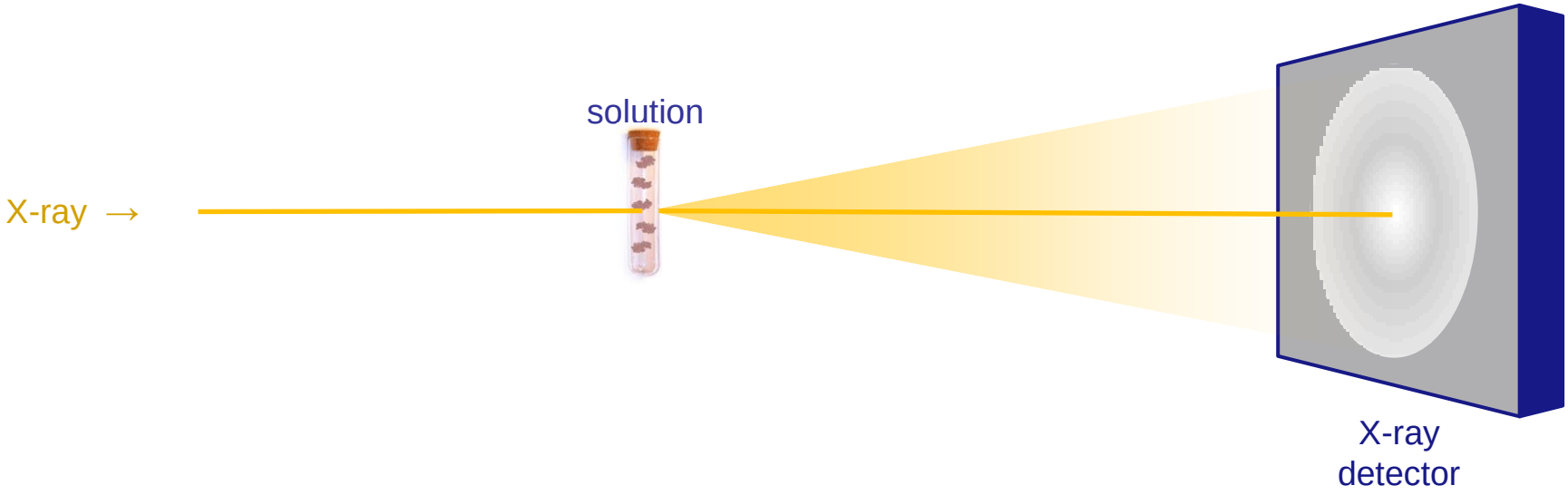
EMBL

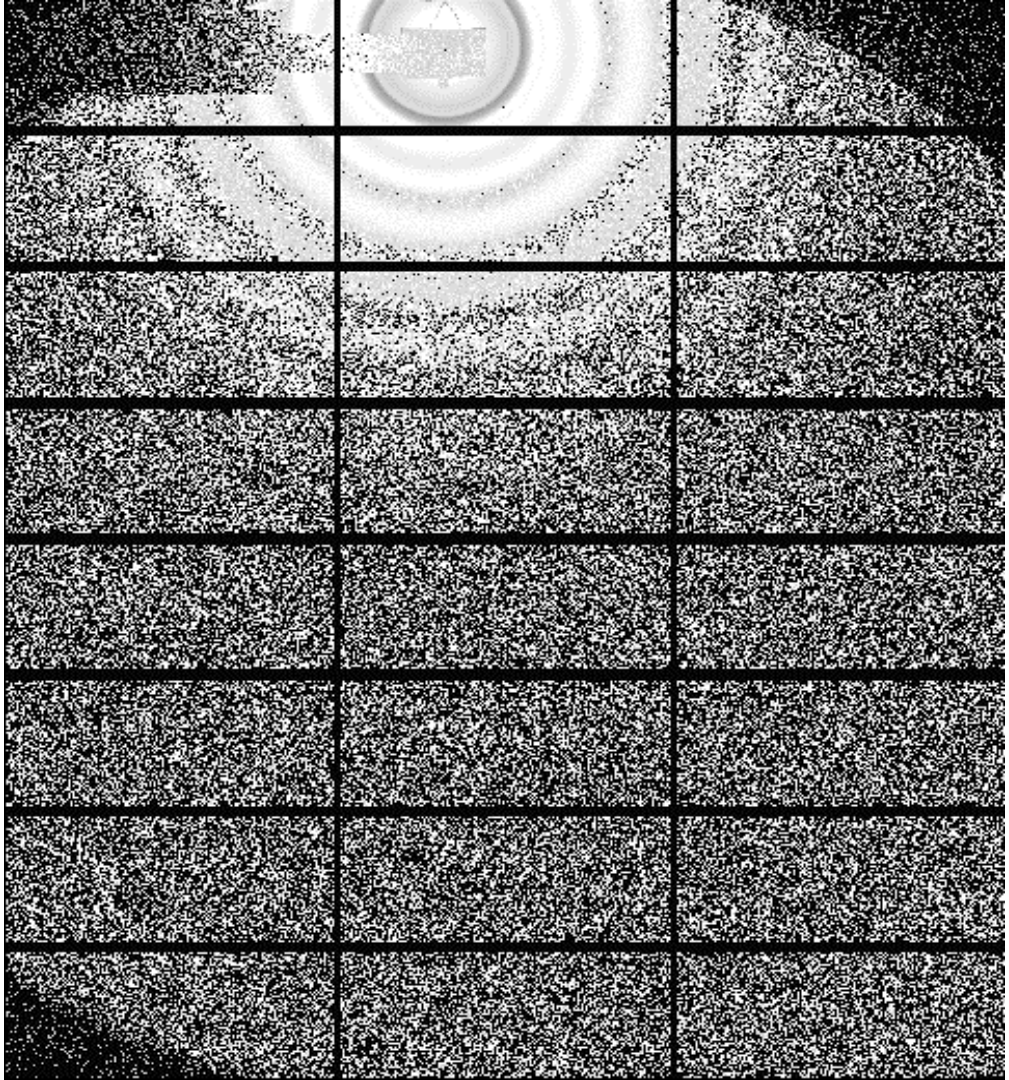
Outline

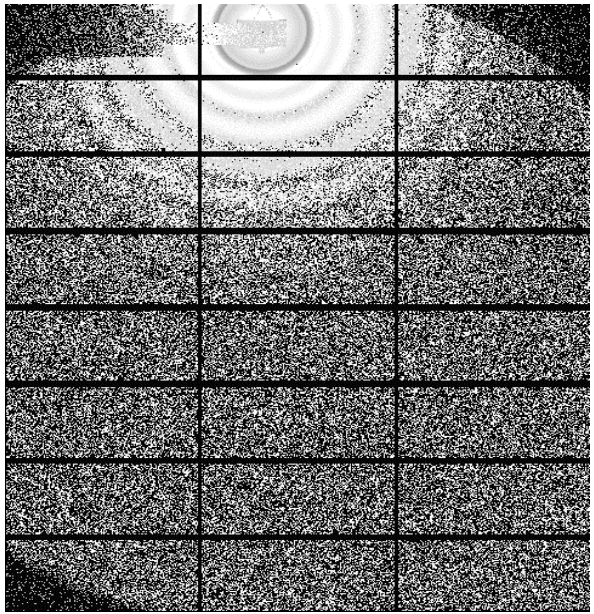
1. Small angle scattering
2. NN architecture and training set
3. Some results

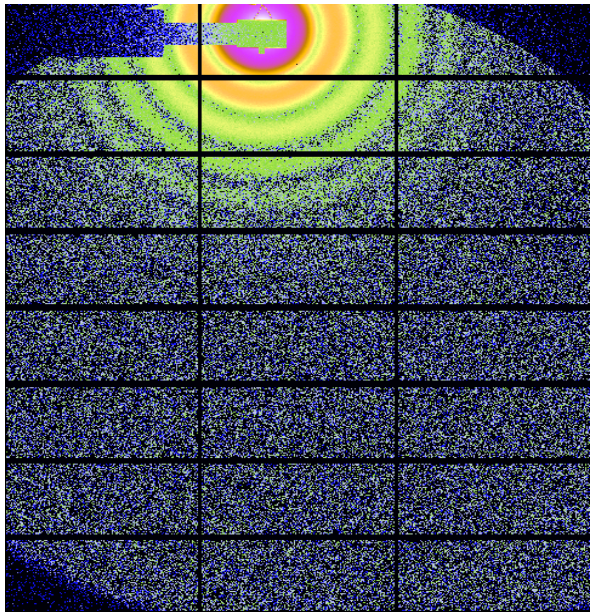
1. Small angle scattering

SAXS experiment

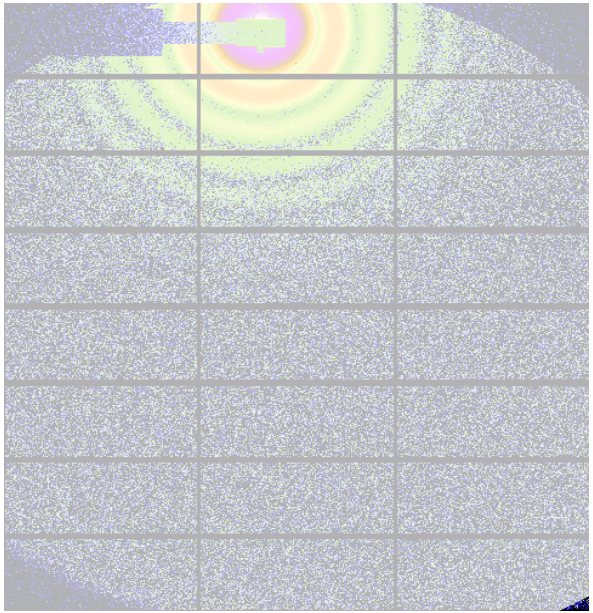


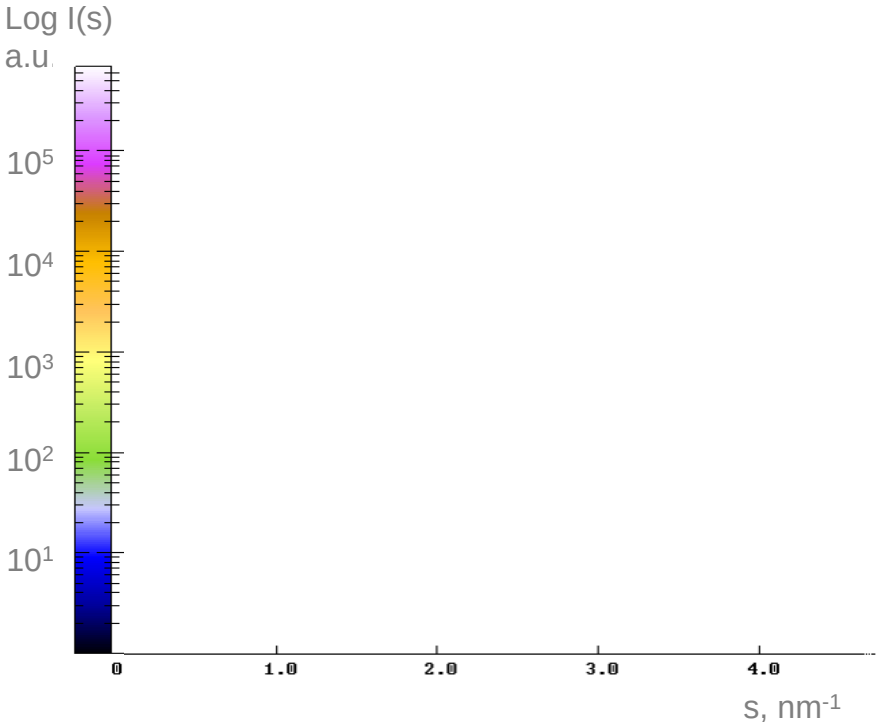




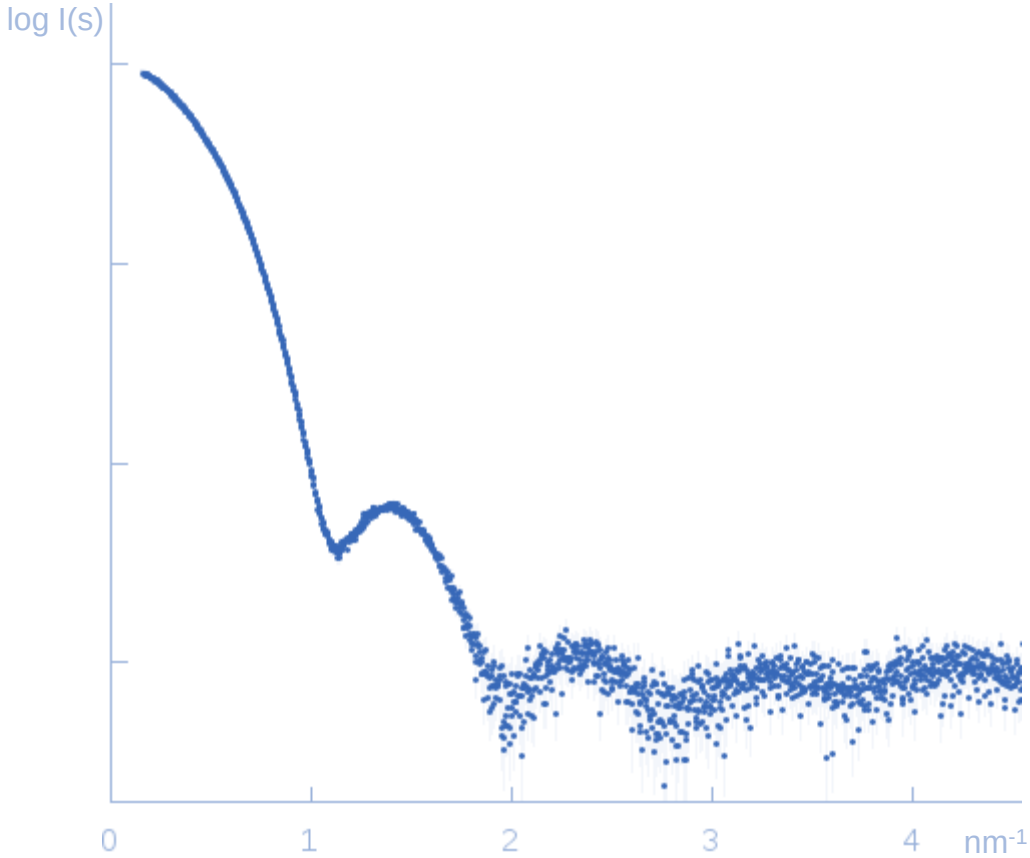

10^5
10^4
10^3
10^2
10^1

2D → 1D



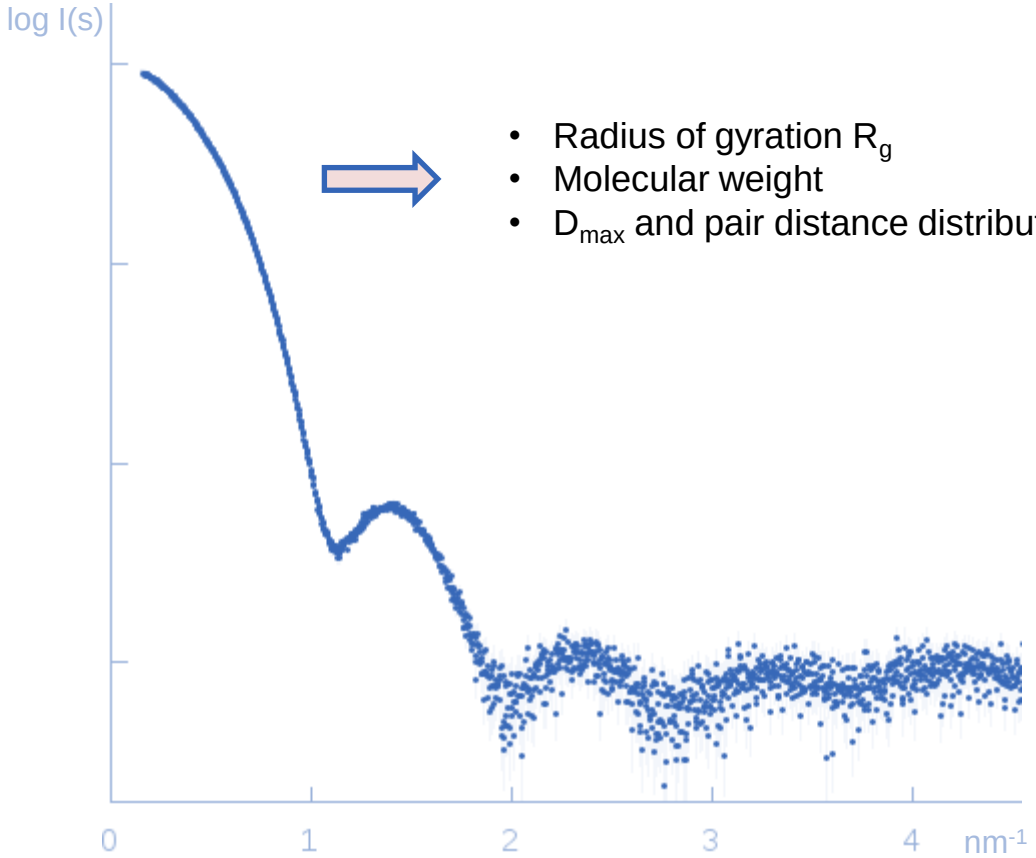


SA(X)S data



experimental SAXS pattern

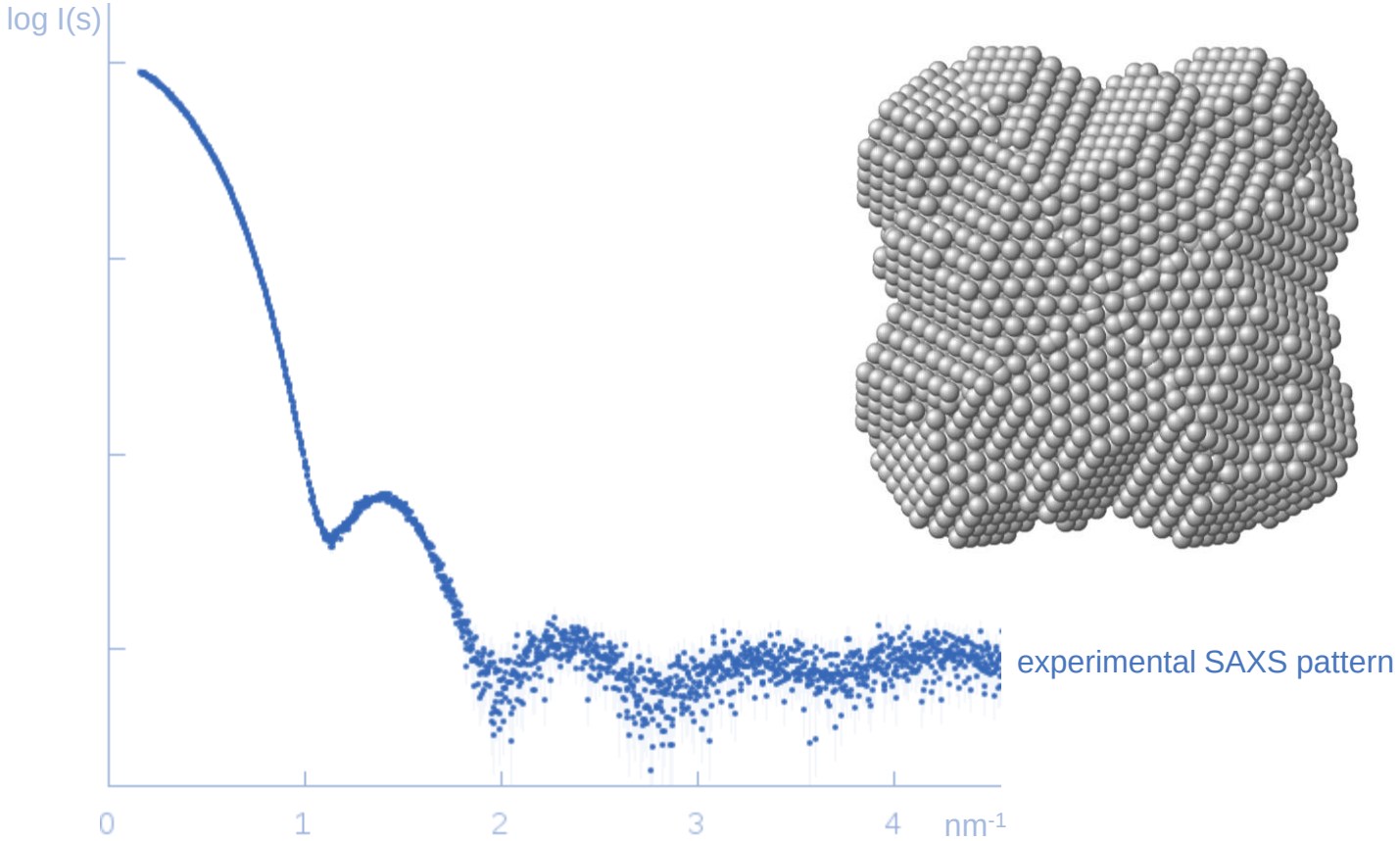
SA(X)S data



- Radius of gyration R_g
- Molecular weight
- D_{max} and pair distance distribution function

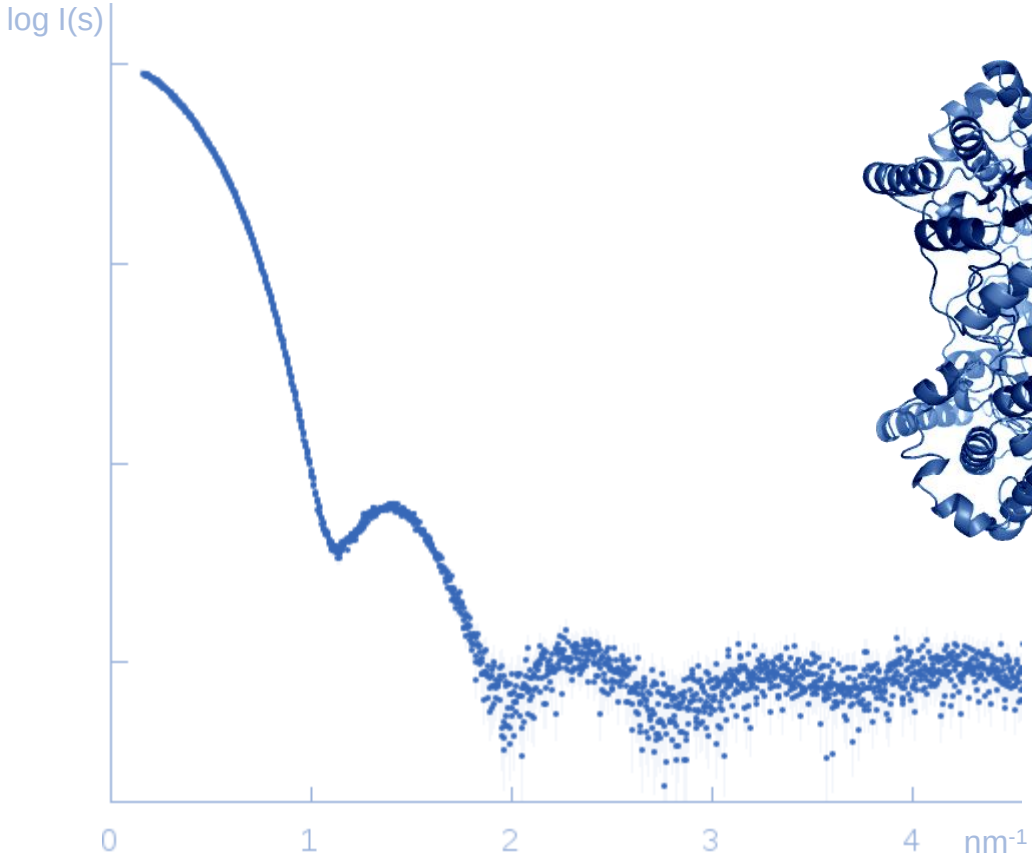
experimental SAXS pattern

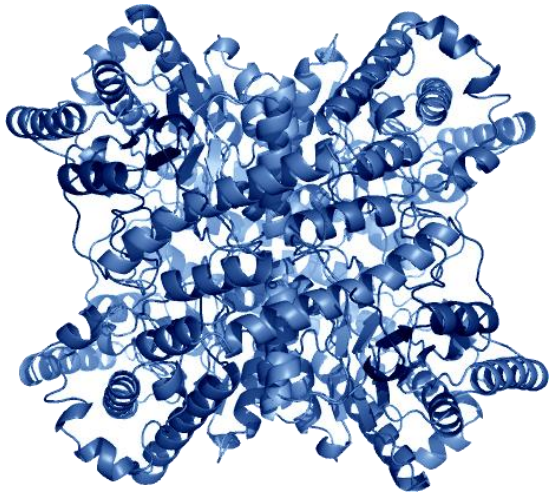
SA(X)S data



experimental SAXS pattern

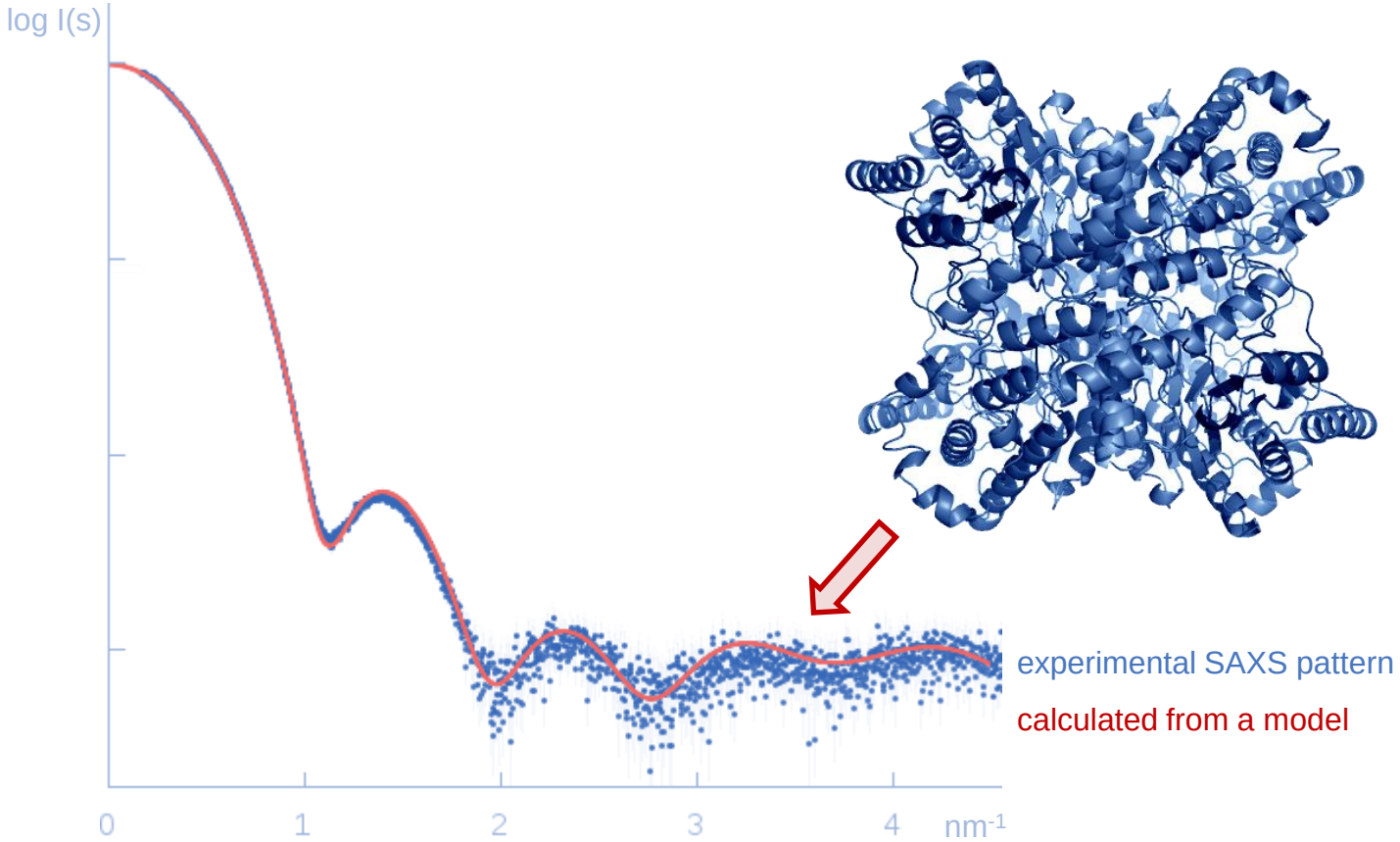
SA(X)S data





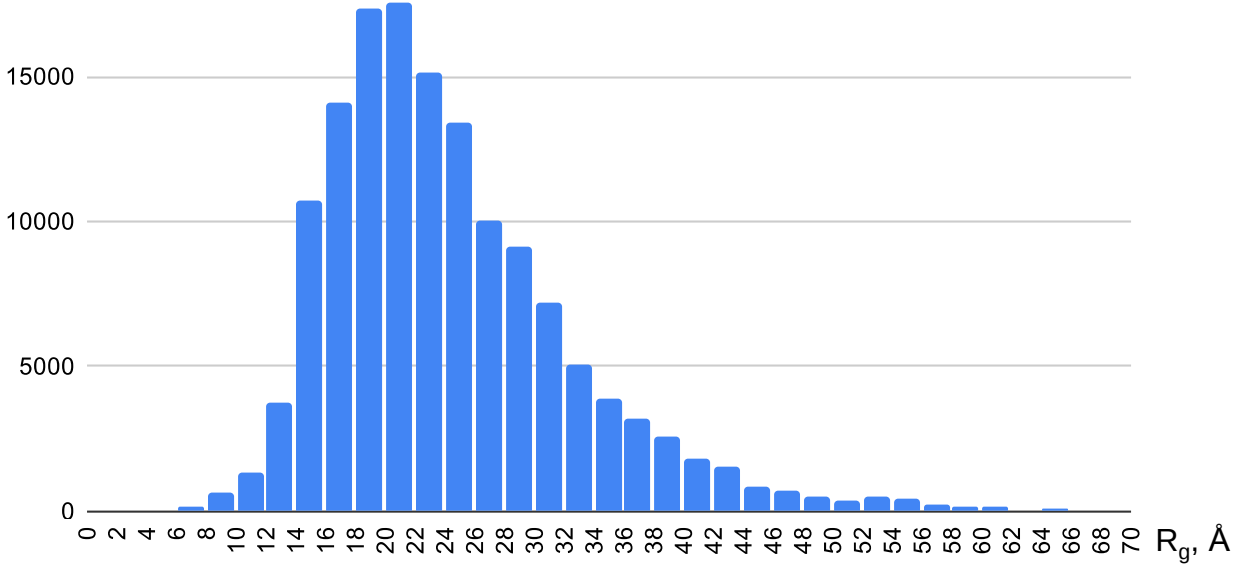
experimental SAXS pattern

SA(X)S data

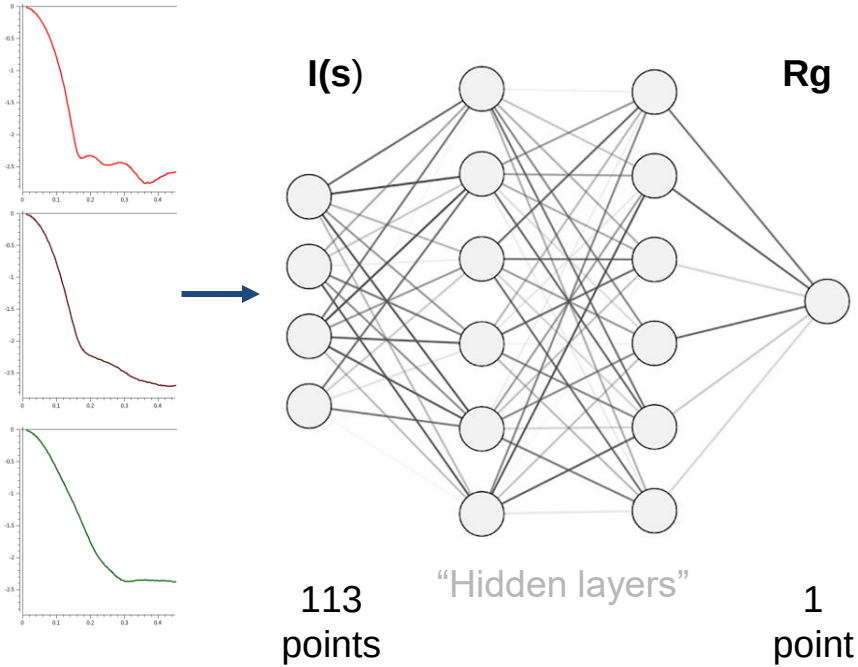

log I(s)
0
1
2
3
4
nm-1
experimental SAXS pattern
calculated from a model

2. Training set and NN architecture

Training set: > 150 000 proteins

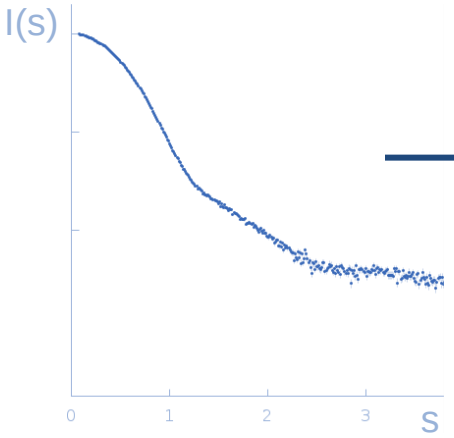


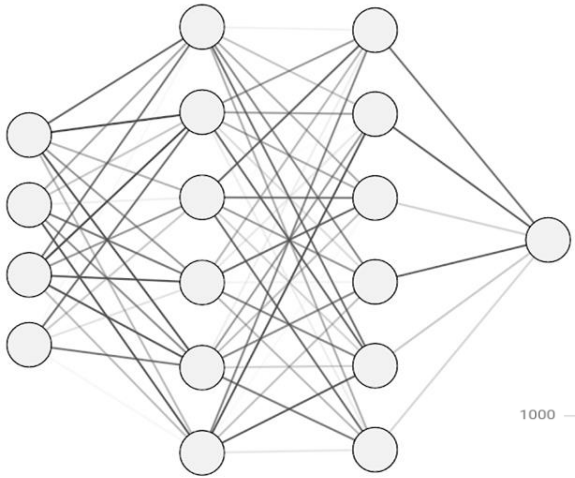
Supervised learning: thousands of I(s)-Rg pairs



Prediction:
previously unseen I(s)

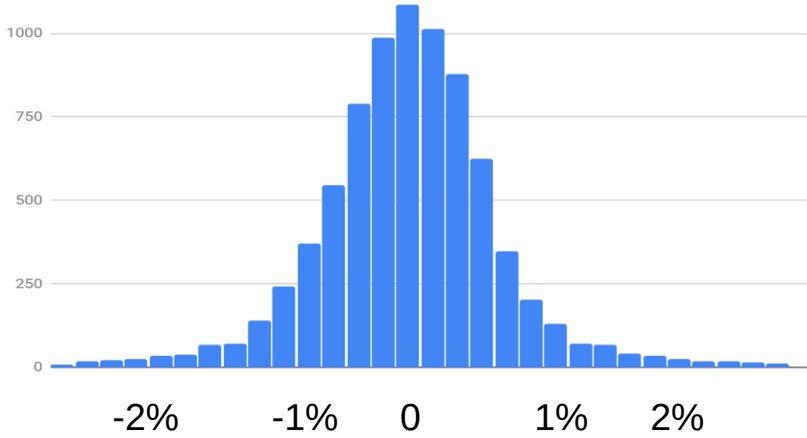
Determine R_g from SAXS data





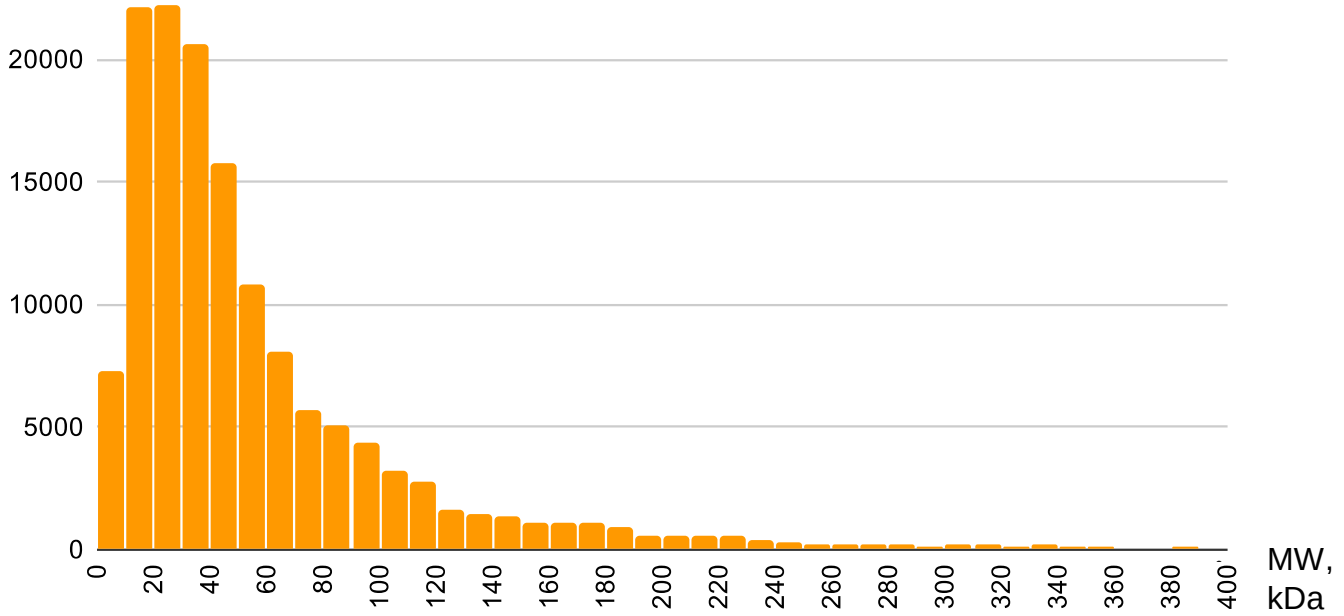
R_g

$$1 - \frac{R_g^{\,predicted}}{R_g^{\,ground\ truth}}$$

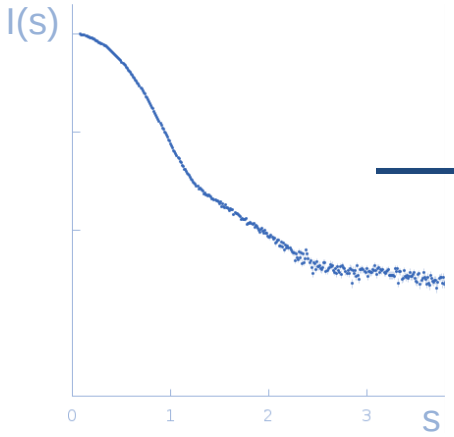


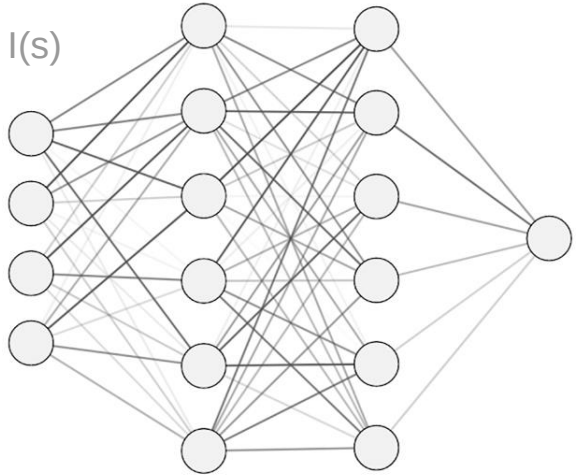
Tested on data simulated from 8192 proteins

MW distribution



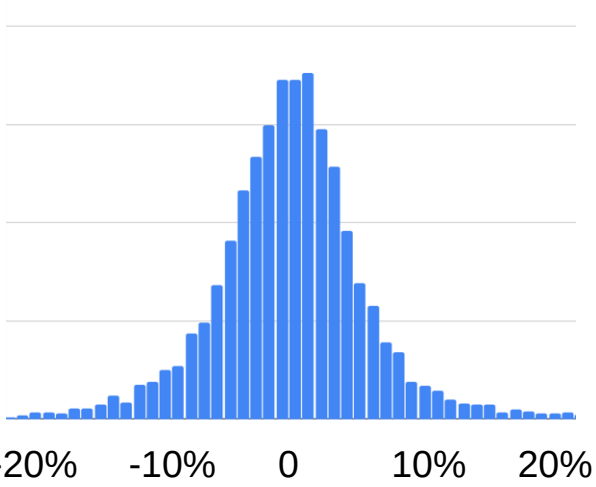
Determine ***molecular weight*** from SAXS data





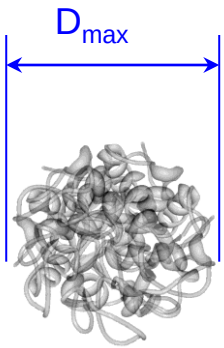
MW

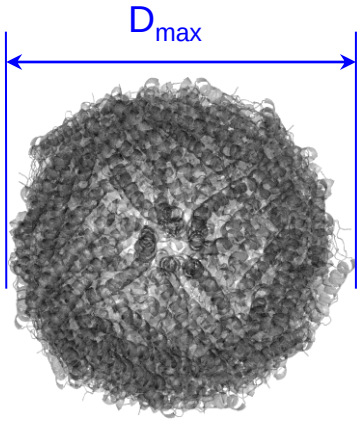
$$1 - \frac{MW^{predicted}}{MW^{ground\ truth}}$$



Tested on data simulated from 8192 proteins

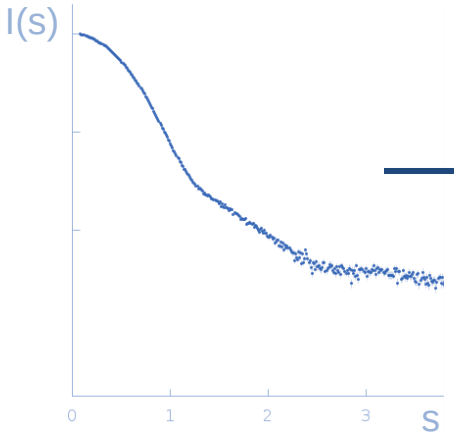
Maximum intra-particle distance of the structure (D_{max})

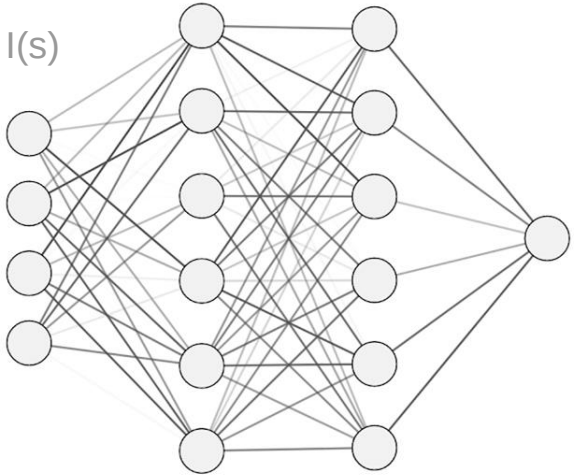




No direct mathematical way to obtain!

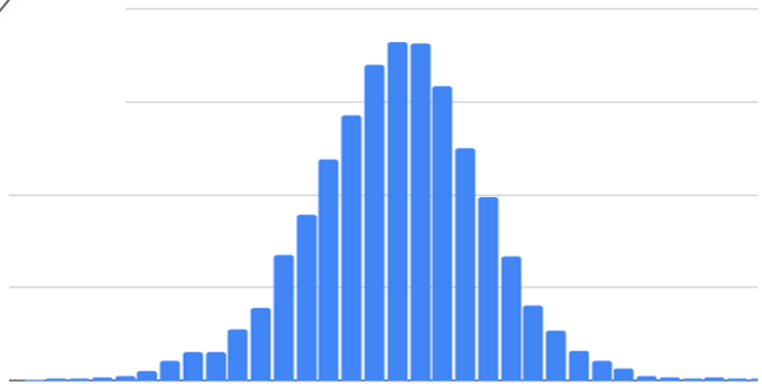
Determine D_{max} from SAXS data





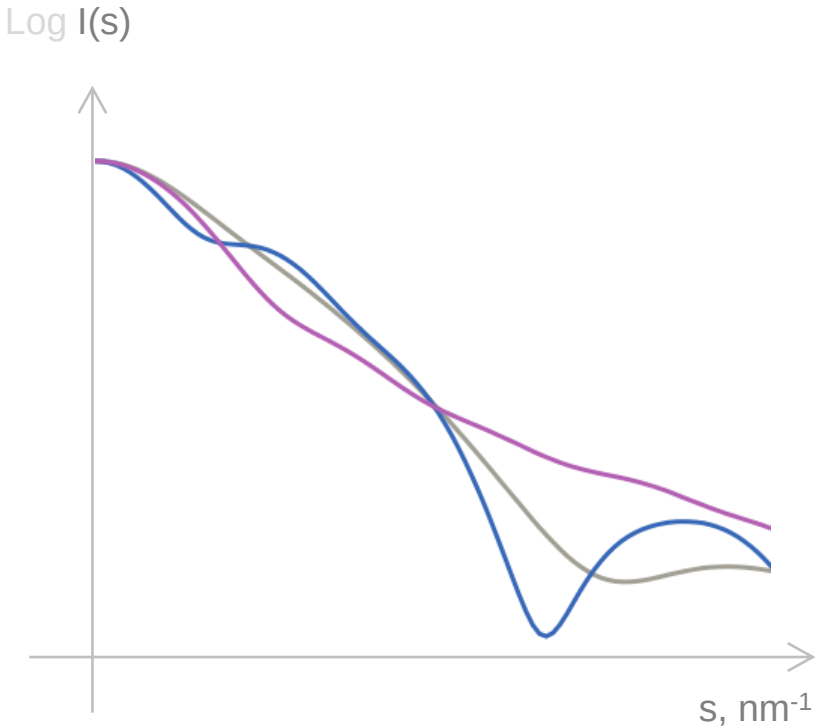
$$1 - \frac{D_{max}^{predicted}}{D_{max}^{ground\ truth}}$$

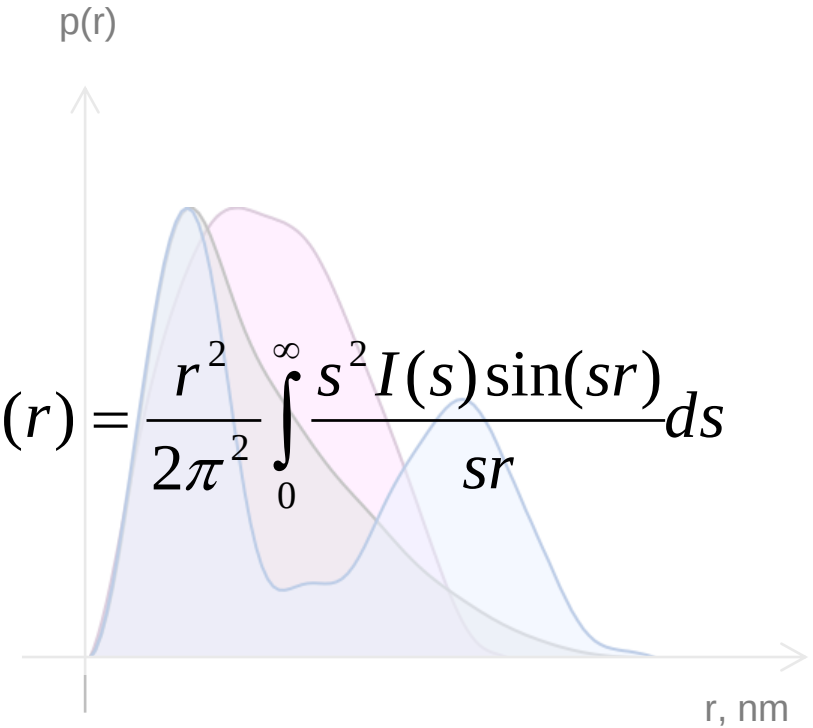
D_{max}



Tested on data simulated from 8192 proteins

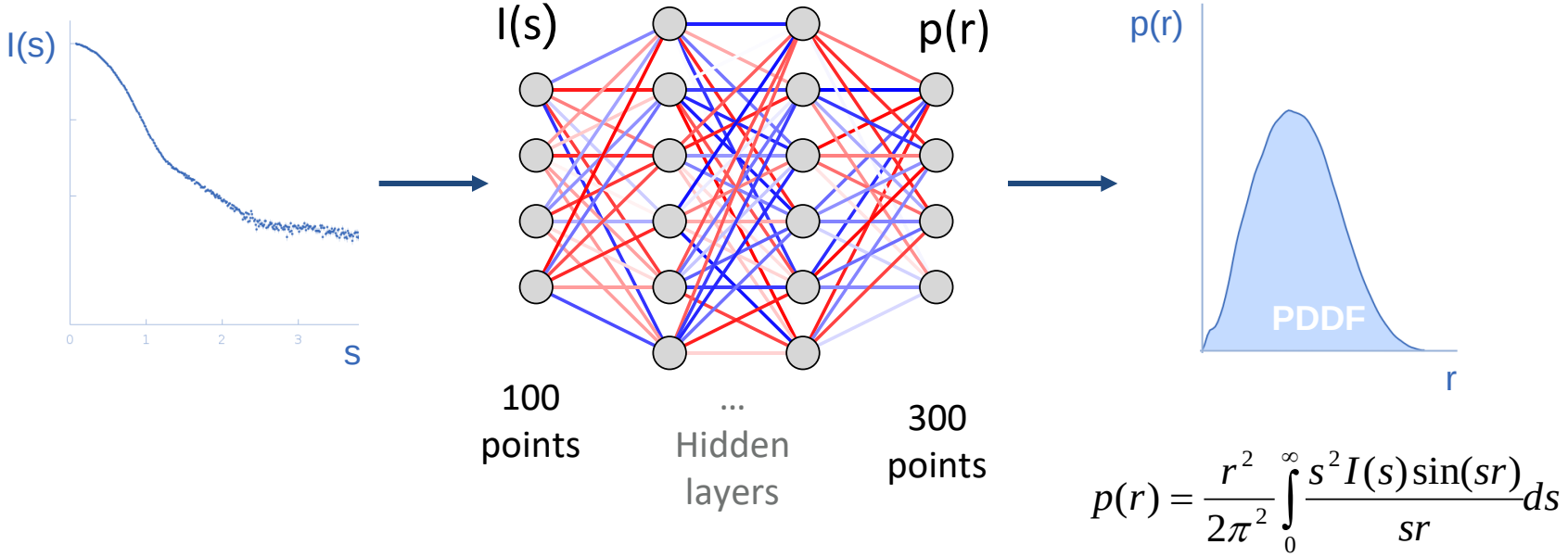
Pair-distance distribution function (PDDF)





$$p(r) = \frac{r^2}{2\pi^2} \int_0^\infty \frac{s^2 I(s) \sin(sr)}{sr} ds$$

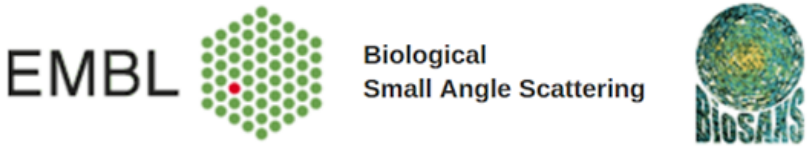
Determine ***PDDF*** from SAXS data



$$p(r) = \frac{r^2}{2\pi^2} \int_0^\infty \frac{s^2 I(s) \sin(sr)}{sr} ds$$

3. Where are we now?

https://dara.embl-hamburg.de/gnnom.php



GNNOM

Fit a PDDF p(r) to experimental data

Experimental data (*.dat) Choose File SASDF99.dat

Angular units (optional) ▼ $s = 4\pi\sin(\theta)/\lambda$

Macromolecule type protein ▼

Submit

EMBL Biological Small Angle Scattering BioSAXS

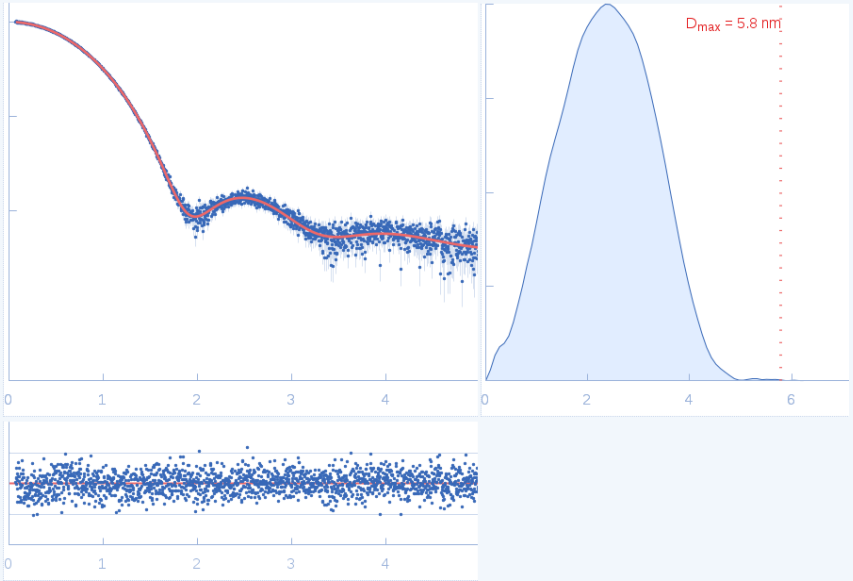
Home > Web services > DARA > SASDFP8.dat

GNNOM

SASDFP8

Predicted R_g	1.8 nm
p(r) R_g	1.8 nm
autorg R_g	1.8 nm
Predicted D_{max}	5.8 nm
p(r) D_{max}	6.2 nm
Predicted MW	23.1 kDa
χ^2	1.0

SASDFP8 – Carbonic anhydrase



Download SASDFP8.fit
Download pddf-predicted.dat

EMBL Biological Small Angle Scattering BioSAXS

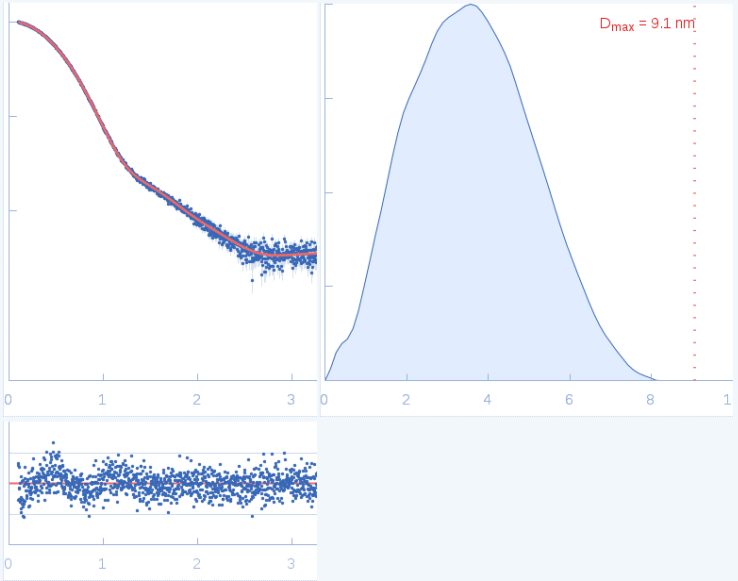
Home > Web services > DARA > SASDFQ8.dat

GNNOM

SASDFQ8

Predicted R_g	2.8 nm
p(r) R_g	2.8 nm
autorg R_g	2.7 nm
Predicted D_{max}	9.1 nm
p(r) D_{max}	8.2 nm
Predicted MW	56.7 kDa
χ^2	1.1

SASDFQ8 – BSA standard



Download SASDFQ8.fit
Download pddf-predicted.dat

EMBL Biological Small Angle Scattering BioSAXS

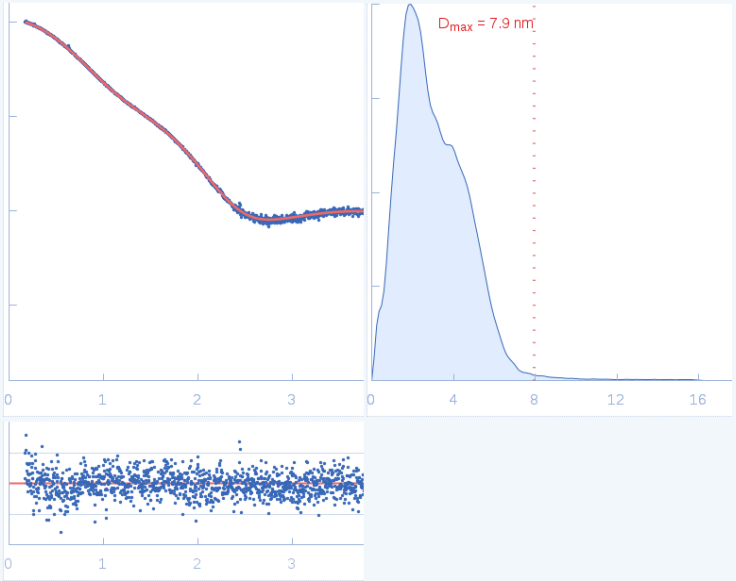
Home > Web services > DARA > SEC-LIG-Jan16-clo.dat

GNNOM

SEC-LIG-Jan16-clo

Predicted R_g	24.5 Å
p(r) R_g	24.9 Å
autorg R_g	23.9 Å
Predicted D_{max}	79.3 Å
p(r) D_{max}	16.5 Å
Predicted MW	19.3 kDa
χ^2	3.5

PDZ domain (extended)



Download SEC-LIG-Jan16-clo.fit
Download pddf-predicted.dat

EMBL Biological Small Angle Scattering BioSAXS

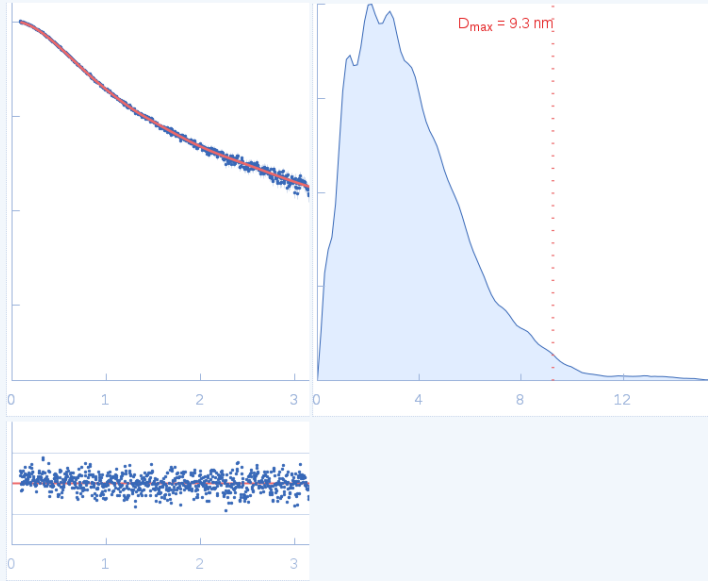
Home > Web services > DARA > Colicin-N.dat

GNNOM

Colicin-N

Predicted R_g	3.0 nm
p(r) R_g	3.0 nm
autorg R_g	2.8 nm
Predicted D_{max}	9.3 nm
p(r) D_{max}	15.8 nm
Predicted MW	19.0 kDa
χ^2	0.7

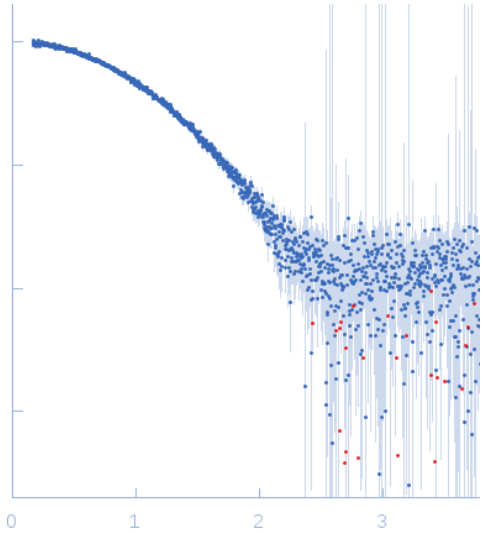
SASDC53 – Colicin N (IDP)

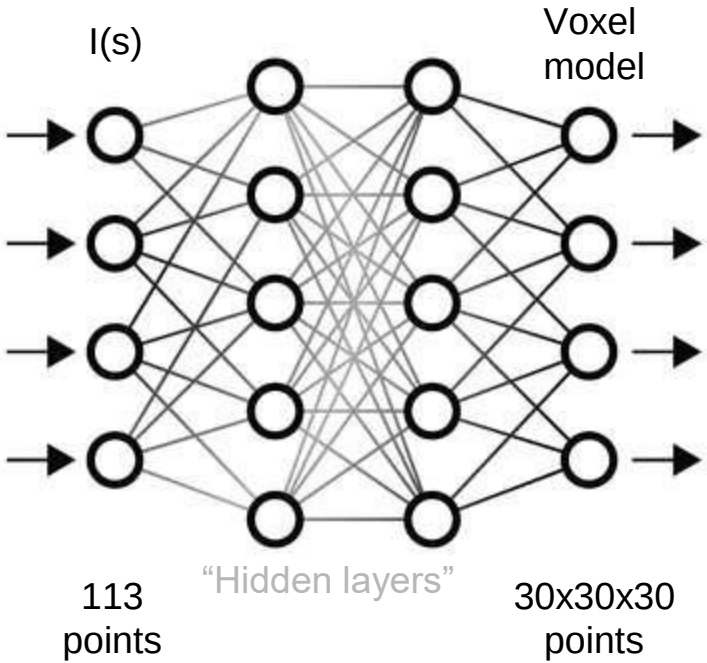


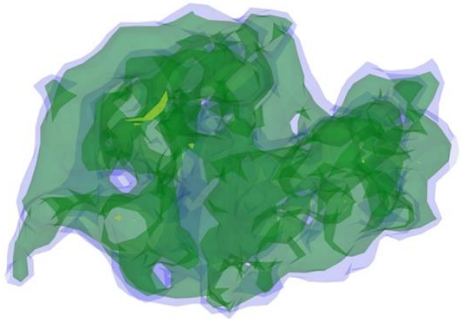
Download Colicin-N.fit
Download pddf-predicted.dat

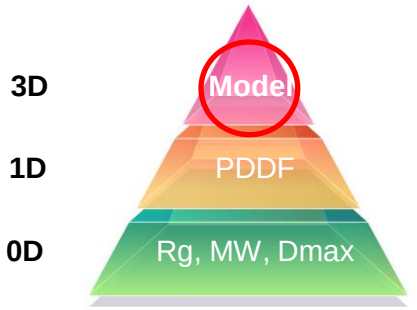
What's next?

Modelling! (3D)



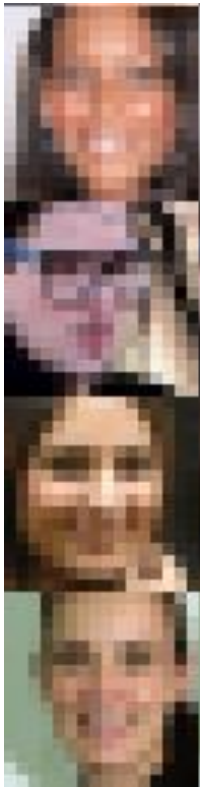






Input

Input



https://github.com/david-gpu/srez

Group members

Dima Molodenskiy
Al Kikhney
Dmitri Svergun

Clément Blanchet
Cy Jeffries
Daniel Franke
Melissa Gräwert
Haydyn Mertens
Clemente Borges
Stefano Da Vela
Tobias Gräwert
Andrey Gruzinov
Martin Schroer
Karen Manalastas
Taja Cheremnykh

EMBL

DESY.

BioSAXS

Keras

TensorFlow

python™

Thank you!

https://dara.embl-hamburg.de/gnnom.php

dmolodenskiy@embl-hamburg.de