Analysis of small-angle scattering data using model fitting and Bayesian regularization

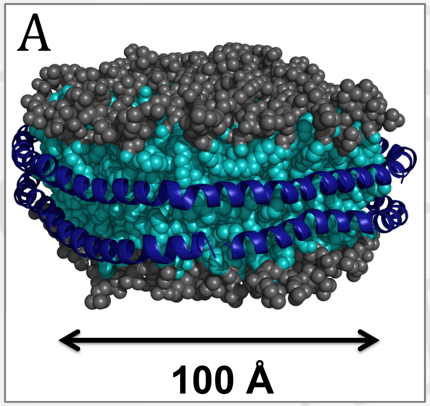
Andreas Haahr Larsen
Post Doc
Niels Bohr Institute & Dep. Biology
Bayes for Scattering,
May 6th 2019

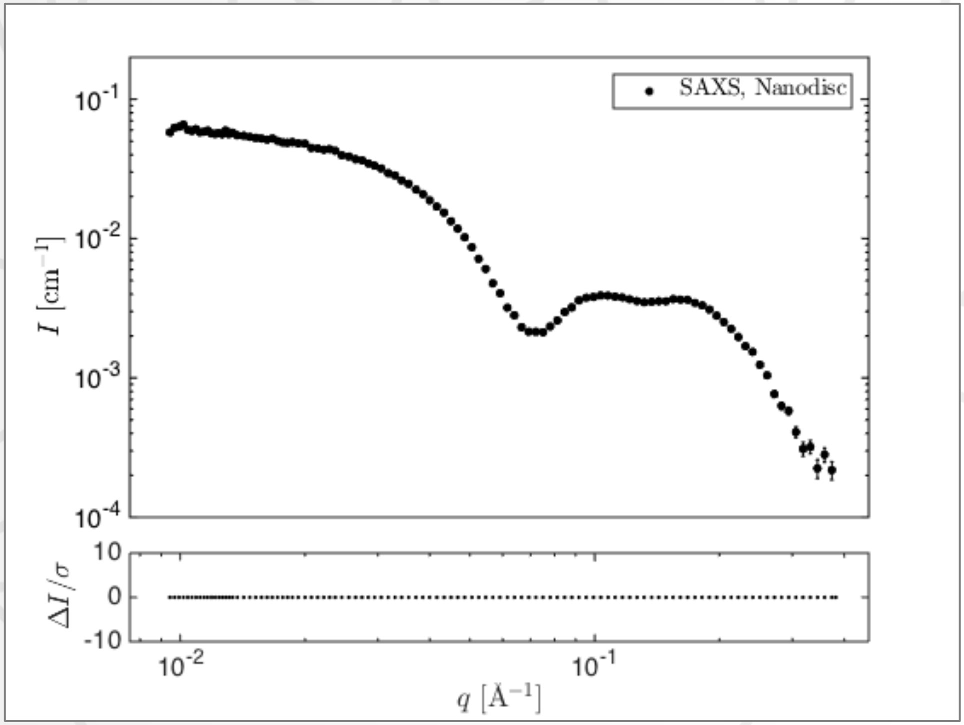
UNIVERSITY OF COPENHAGEN

Analysis of small-angle scattering data using model fitting and Bayesian regularization

Andreas Haahr Larsen
Post Doc
Niels Bohr Institute & Dep. Biology
Bayes for Scattering,
May 6th 2019

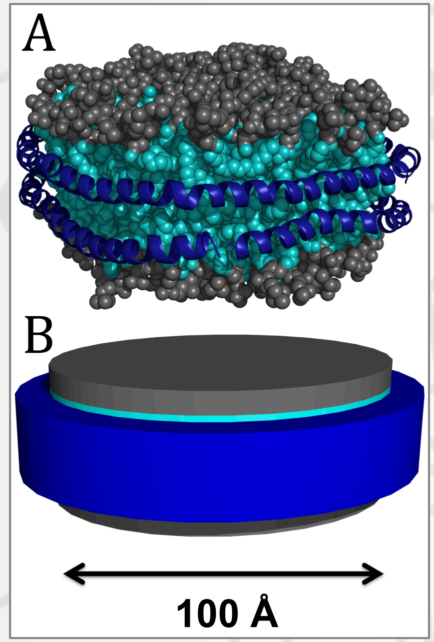
UNIVERSITY OF COPENHAGEN



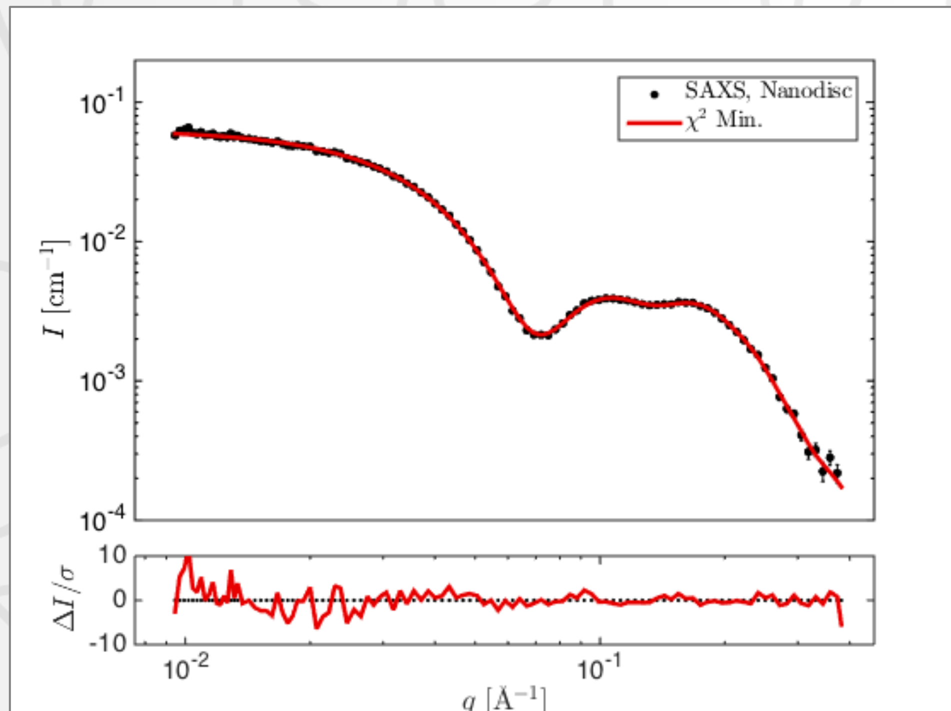


Analysis of small-angle scattering data using model fitting and Bayesian regularization

Andreas Haahr Larsen
Post Doc
Niels Bohr Institute & Dep. Biology
Bayes for Scattering,
May 6th 2019
UNIVERSITY OF COPENHAGEN



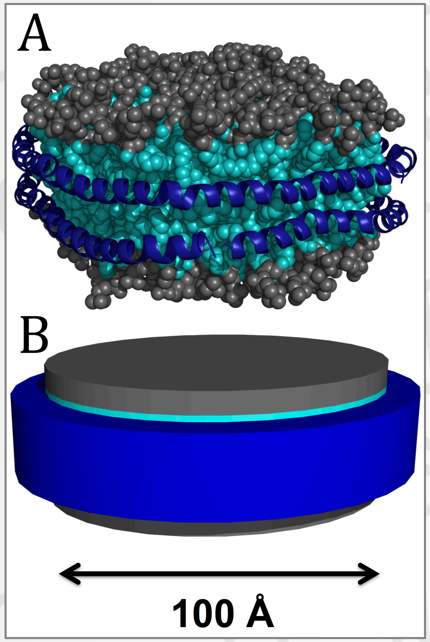
Model parameter	χ^2 minimization
N	103 (22)
ε	1.3 (4.5)
A (Å^2)	76 (19)
n_w	18 (12)
V_l (Å^3)	996 (19)
V_t (Å^3)	702 (111)
V_p ($\times 10^4$ Å^3)	5.3 (0.7)
T (Å)	11.4 (2.1)
R_g (Å)	13.8 (1.2)
σ_R (Å)	3.2 (1.0)
c (μM)	22 (19)
B ($\times 10^{-4}$ cm^{-1})	−0.3 (2.1)
Goodness of fit, χ^2_r	6.26
Goodness of fit, Cmap	$C = 10$, $N = 106$ $P(C \geq 10 \mid N) = 9.2\%$



Analysis of small-angle scattering data using model fitting and Bayesian regularization

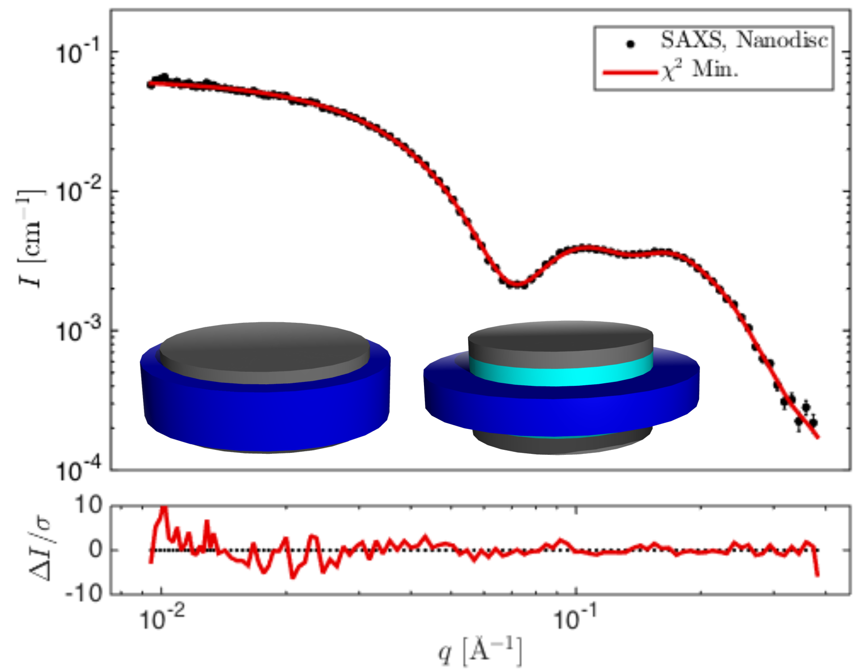
Andreas Haahr Larsen
Post Doc
Niels Bohr Institute & Dep. Biology
Bayes for Scattering,
May 6th 2019

UNIVERSITY OF COPENHAGEN



Model parameter	χ^2 minimization
N	103 (22)
ε	1.3 (4.5)
A (Å^2)	76 (19)
n_w	18 (12)
V_l (Å^3)	996 (19)
V_t (Å^3)	702 (111)
V_p ($\times 10^4$ Å^3)	5.3 (0.7)
T (Å)	11.4 (2.1)
R_g (Å)	13.8 (1.2)
σ_R (Å)	3.2 (1.0)
c (μM)	22 (19)
B ($\times 10^{-4}$ cm^{-1})	−0.3 (2.1)
Goodness of fit, χ^2_r	6.26
Goodness of fit, Cmap	$C = 10$, $N = 106$ $P(C \geq 10 \mid N) = 9.2\%$

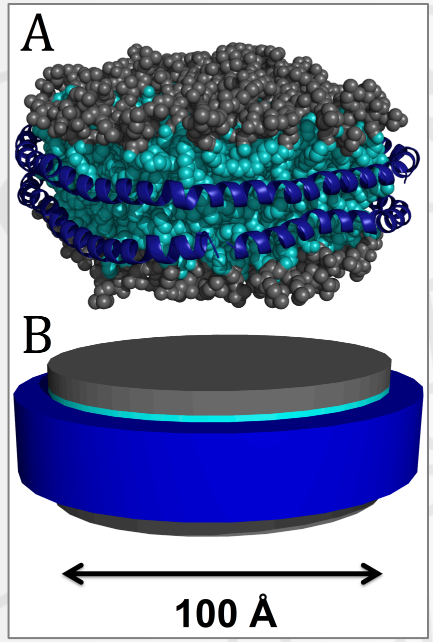
Problem:
- Several possible parameters, some unphysical
- How to select the most probable set of parameters



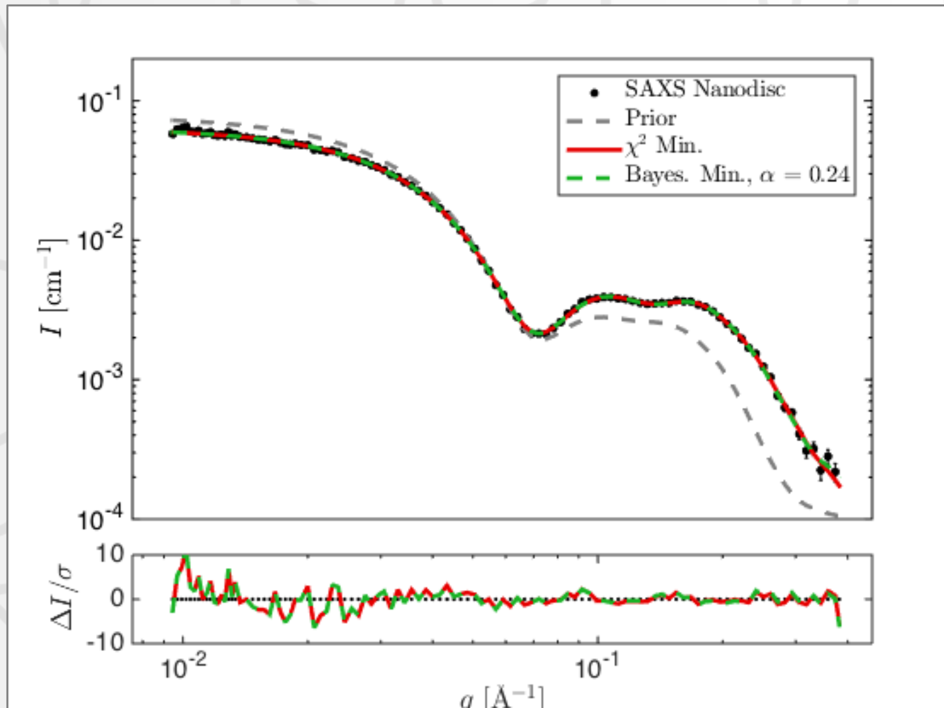
Analysis of small-angle scattering data using model fitting and Bayesian regularization

Andreas Haahr Larsen
Niels Bohr Institute
Bayes for Scattering,
May 6th 2019

UNIVERSITY OF COPENHAGEN



Model parameter	χ^2 minimization
N	103 (22)
ε	1.3 (4.5)
A (Å^2)	76 (19)
n_w	18 (12)
V_1 (Å^3)	996 (19)
V_t (Å^3)	702 (111)
V_p ($\times 10^4$ Å^3)	5.3 (0.7)
T (Å)	11.4 (2.1)
R_g (Å)	13.8 (1.2)
σ_R (Å)	3.2 (1.0)
c (μM)	22 (19)
B ($\times 10^{-4}$ cm^{-1})	−0.3 (2.1)
Goodness of fit, χ^2_r	6.26
Goodness of fit, Cmap	$C = 10$, $N = 106$ $P(C \geq 10 \mid N) = 9.2\%$

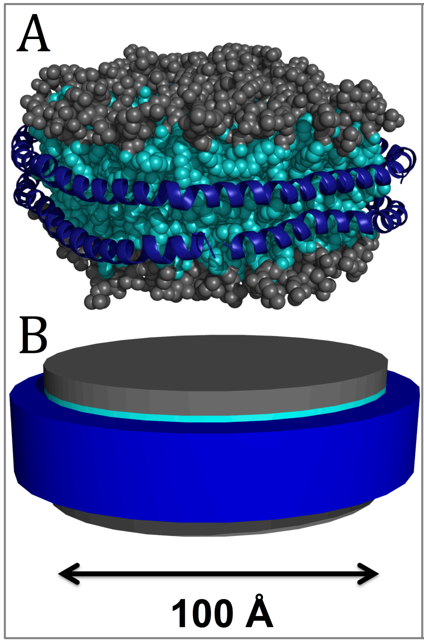


Problem:
- Several possible parameters, some unphysical
- How to select the most probable set of parameters

Solution:
Analysis with Bayesian regularization

Strategies to select the most probable set of parameters

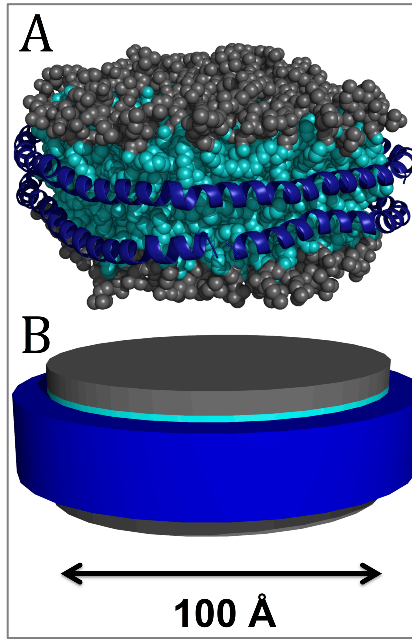
1. **Molecular constraints**
   - **reparametrisation**



Model parameter	χ^2 minimization
N	103 (22)
ε	1.3 (4.5)
A (Å^2)	76 (19)
n_w	18 (12)
V_l (Å^3)	996 (19)
V_t (Å^3)	702 (111)
V_p ($\times 10^4$ Å^3)	5.3 (0.7)
T (Å)	11.4 (2.1)
R_g (Å)	13.8 (1.2)
σ_R (Å)	3.2 (1.0)
c (μM)	22 (19)
B ($\times 10^{-4}$ cm^{-1})	−0.3 (2.1)
Goodness of fit, χ^2_r	6.26
Goodness of fit, Cmap	$C = 10$, $N = 106$ $P(C \geq 10 \mid N) = 9.2\%$

Strategies to select the most probable set of parameters

1. **Molecular constraints**
   - **reparametrisation**
2. **Fix parameters**



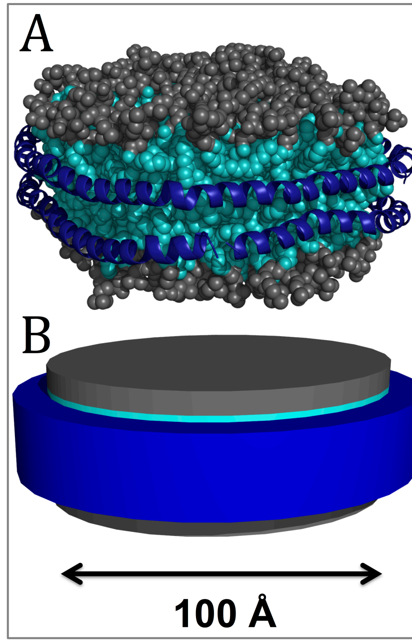
Model parameter	χ^2 minimization
N	103 (22)
ε	1.3 (4.5)
A (Å^2)	76 (19)
n_w	18 (12)
V_1 (Å^3)	996 (19)
V_t (Å^3)	702 (111)
V_p ($\times 10^4$ Å^3)	5.3 (0.7)
T (Å)	11.4 (2.1)
R_g (Å)	13.8 (1.2)
σ_R (Å)	3.2 (1.0)
c (μM)	22 (19)
B ($\times 10^{-4}$ cm^{-1})	−0.3 (2.1)
Goodness of fit, χ_r^2	6.26
Goodness of fit, Cmap	$C = 10, N = 106$ $P(C \geq 10 \mid N) = 9.2\%$

Strategies to select the most probable set of parameters

1. **Molecular constraints**
   - **reparametrisation**
2. **Fix parameters**
3. **Limit allowed parameter interval**

Uniform prior

$$p(\kappa) = \begin{cases} s & \text{if} \quad a \leq \kappa \leq b, \\ 0 & \text{otherwise,} \end{cases}$$



Model parameter	χ^2 minimization
N	103 (22)
ε	1.3 (4.5)
A (Å^2)	76 (19)
n_w	18 (12)
V_l (Å^3)	996 (19)
V_t (Å^3)	702 (111)
V_p ($\times 10^4$ Å^3)	5.3 (0.7)
T (Å)	11.4 (2.1)
R_g (Å)	13.8 (1.2)
σ_R (Å)	3.2 (1.0)
c (μM)	22 (19)
B ($\times 10^{-4}$ cm^{-1})	−0.3 (2.1)
Goodness of fit, χ^2_r	6.26
Goodness of fit, Cmap	$C = 10$, $N = 106$ $P(C \geq 10 \mid N) = 9.2\%$

SasView

Will It Fit?

CRYSOL

FoXS

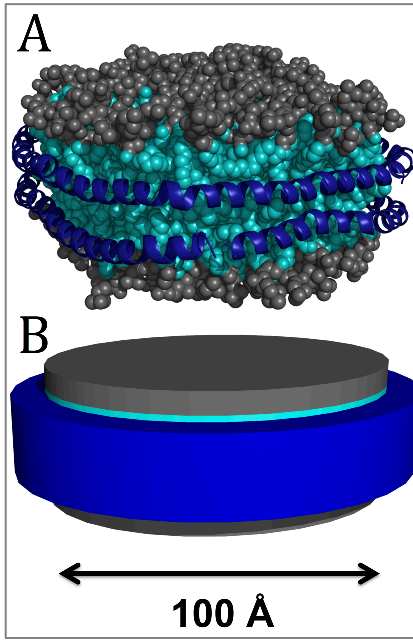
Strategies to select the most probable set of parameters

1. **Molecular constraints**
   - **reparametrisation**
2. **Fix parameters**
3. **Limit allowed parameter interval**

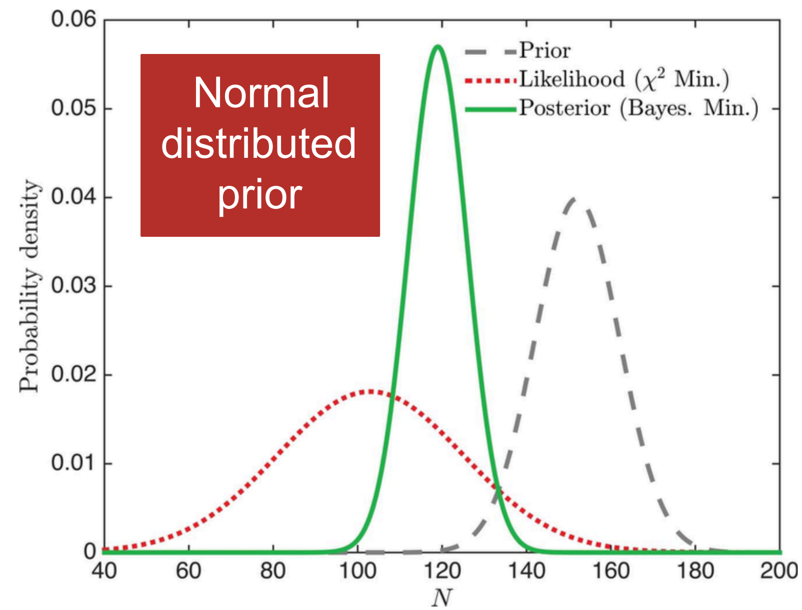
Uniform prior

$$p(\kappa) = \begin{cases} s & \text{if} \quad a \leq \kappa \leq b, \\ 0 & \text{otherwise,} \end{cases}$$

4. **Bayesian regularisation**



Model parameter	χ^2 minimization
N	103 (22)
ε	1.3 (4.5)
A (Å^2)	76 (19)
n_w	18 (12)
V_l (Å^3)	996 (19)
V_t (Å^3)	702 (111)
V_p ($\times 10^4$ Å^3)	5.3 (0.7)
T (Å)	11.4 (2.1)
R_g (Å)	13.8 (1.2)
σ_R (Å)	3.2 (1.0)
c (μM)	22 (19)
B ($\times 10^{-4}$ cm^{-1})	−0.3 (2.1)
Goodness of fit, χ_r^2	6.26
Goodness of fit, Cmap	$C = 10$, $N = 106$ $P(C \geq 10 \mid N) = 9.2\%$



Strategies to select the most probable set of parameters

1. **Molecular constraints**
   - **reparametrisation**
2. **Fix parameters**
3. **Limit allowed parameter interval**

Uniform prior

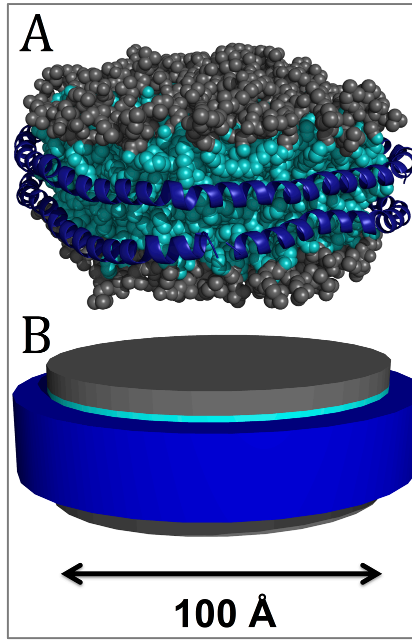
$$p(\kappa) = \begin{cases} s & \text{if} \quad a \leq \kappa \leq b, \\ 0 & \text{otherwise,} \end{cases}$$

4. **Bayesian regularisation**

not regularised: $$\chi^2(\mathbf{p}) = \sum_{i=1}^{N} \frac{\left[I_i^{\text{exp}} - I_i^{\text{th}}(\mathbf{p})\right]^2}{\sigma_i^2}$$

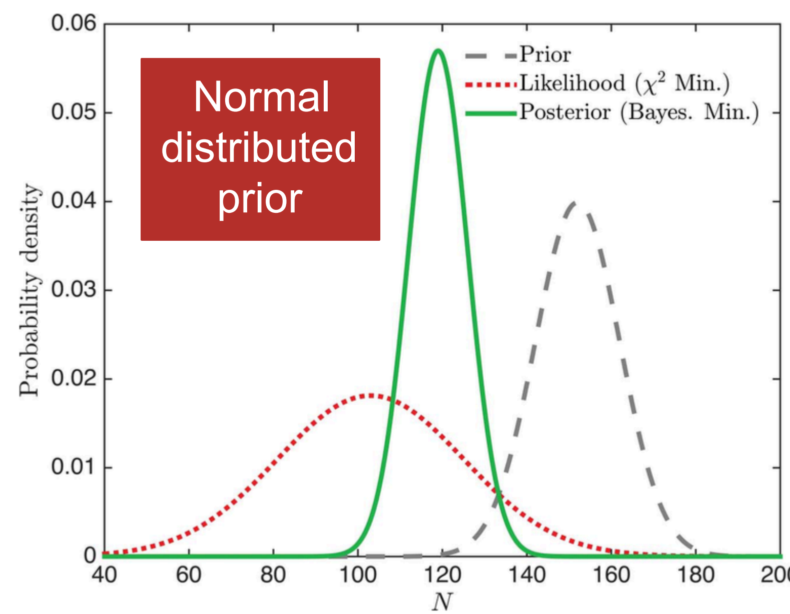
regularised: $$Q(\mathbf{p}) = \chi^2(\mathbf{p}) + \alpha S(\mathbf{p})$$

prior: $$S(\mathbf{p}) = \sum_{k=1}^{K} \frac{(p_k - \mu_k)^2}{\delta p_k^2}$$



Model parameter	χ^2 minimization
N	103 (22)
ε	1.3 (4.5)
A (Å^2)	76 (19)
n_w	18 (12)
V_l (Å^3)	996 (19)
V_t (Å^3)	702 (111)
V_p ($\times10^4$ Å^3)	5.3 (0.7)
T (Å)	11.4 (2.1)
R_g (Å)	13.8 (1.2)
σ_R (Å)	3.2 (1.0)
c (μM)	22 (19)
B ($\times10^{-4}$ cm^{-1})	−0.3 (2.1)
Goodness of fit, χ_r^2	6.26
Goodness of fit, Cmap	$C = 10$, $N = 106$ $P(C \geq 10 \mid N) = 9.2\%$

Normal distributed prior



Challenges in Bayesian regularization

***P**robability*

1. **Determining α:** $-2\log[P(D,\alpha \mid H)] = Q(\mathbf{p}) + \log(\Gamma) + 2\log(\alpha)$

evidence

Challenges in Bayesian regularization

Probability $\chi^2(\mathbf{p}) + \alpha S(\mathbf{p})$

1. **Determining α:** $-2\log[P(D, \alpha \mid H)] = Q(\mathbf{p}) + \log(\Gamma) + 2\log(\alpha)$

evidence

Challenges in Bayesian regularization

Probability $\chi^2(\mathbf{p}) + \alpha S(\mathbf{p})$

1. Determining α: $-2\log[P(D, \alpha \mid H)] = Q(\mathbf{p}) + \log(\Gamma) + 2\log(\alpha)$

evidence *Occam term*

Occam's razor:
Choose the simplest explanation

MacKay, D. J. C. (1992). Adv. Neural Inf. Process. Syst. 4, 839–846

Challenges in Bayesian regularization

***P**robability* $\chi^2(\mathbf{p}) + \alpha S(\mathbf{p})$ *prior for α*

1. Determining α:

$$-2\log[P(D,\alpha \mid H)] = Q(\mathbf{p}) + \log(\Gamma) + 2\log(\alpha)$$

evidence *Occam term*

Occam's razor:
Choose the simplest explanation

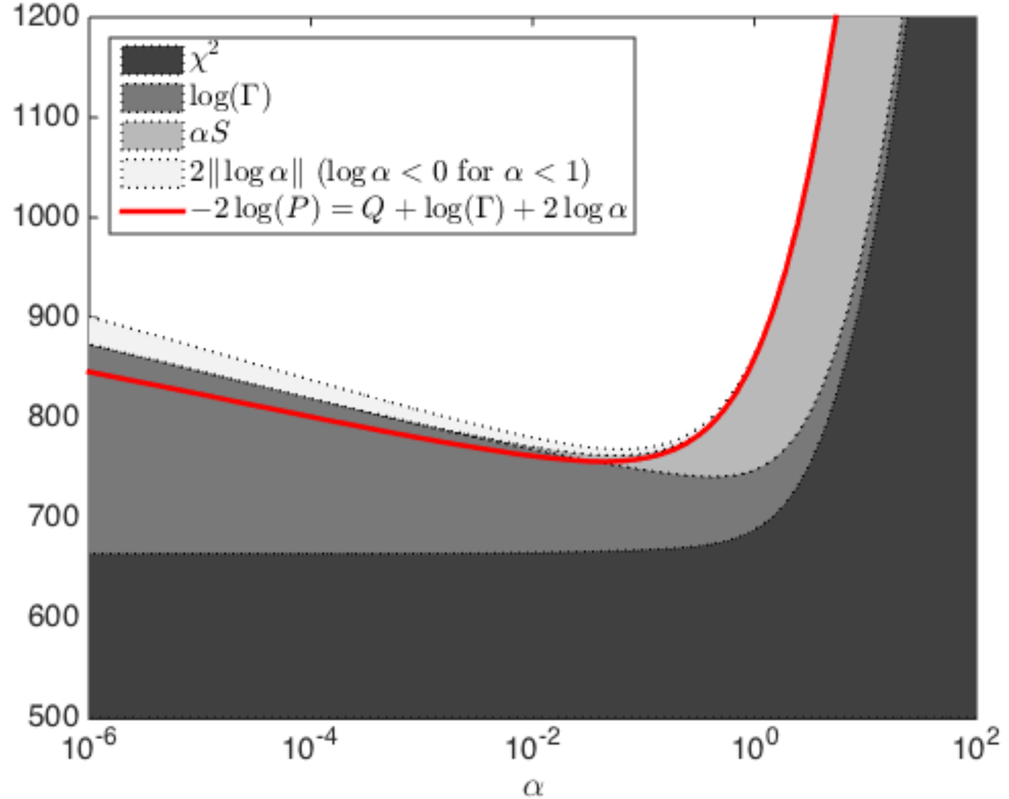
Challenges in Bayesian regularization

Probability $\chi^2(\mathbf{p}) + \alpha S(\mathbf{p})$ *prior for α*

1. Determining α: $-2\log[P(D,\alpha \mid H)] = Q(\mathbf{p}) + \log(\Gamma) + 2\log(\alpha)$

evidence *Occam term*

Occam's razor:
Choose the simplest explanation
Here: same as prior (large α)



Challenges in Bayesian regularization

Probability $\chi^2(\mathbf{p}) + \alpha S(\mathbf{p})$ *prior for α*

1. Determining α:

$$-2\log[P(D,\alpha \mid H)] = Q(\mathbf{p}) + \log(\Gamma) + 2\log(\alpha)$$

evidence *Occam term*

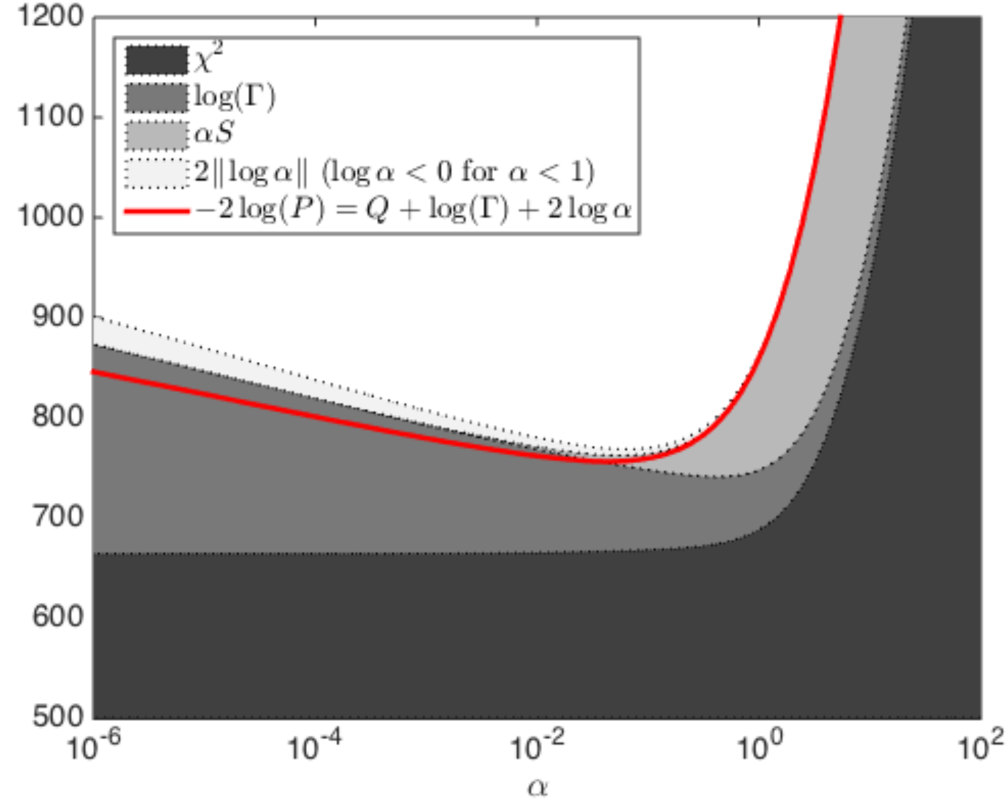
Occam's razor:
Choose the simplest explanation
Here: same as prior (large α)

2. Choose prior

Case 1: From experiment

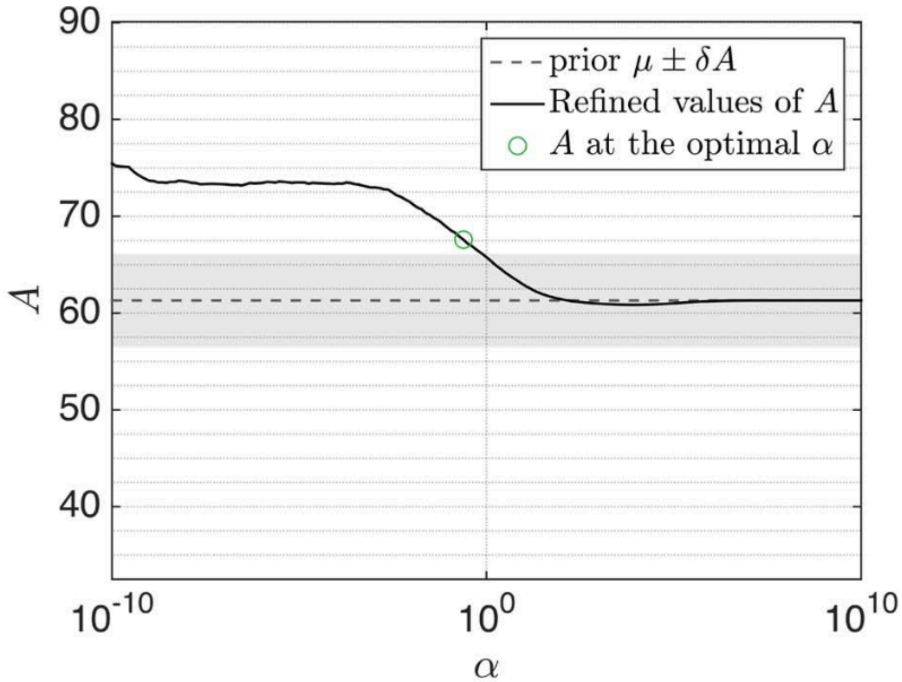
Case 2: General knowledge

Case 3: Unknown



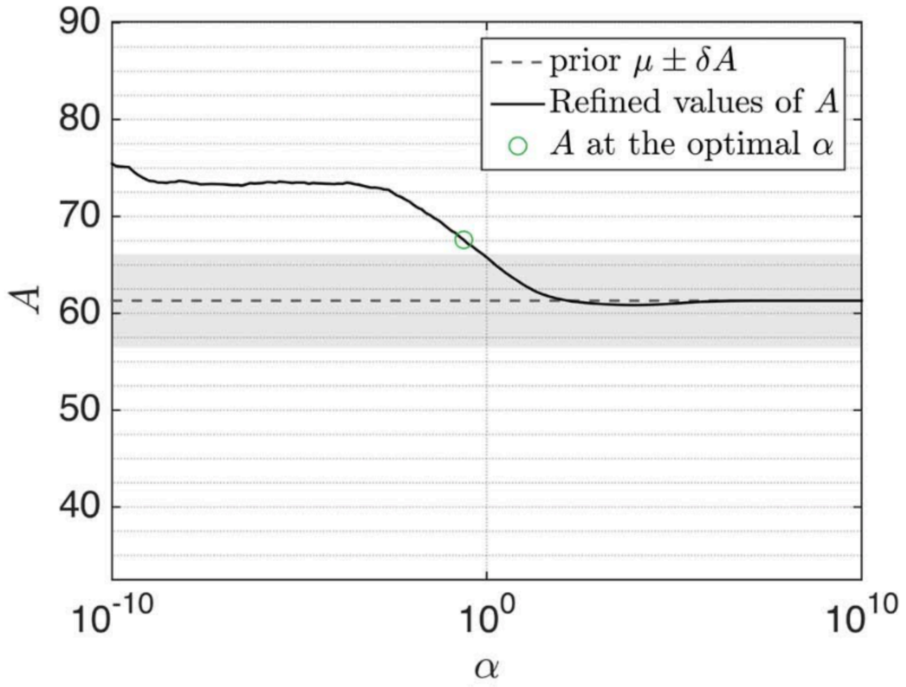
Benefits of Bayesian regularization

1. **Most probable solution found**



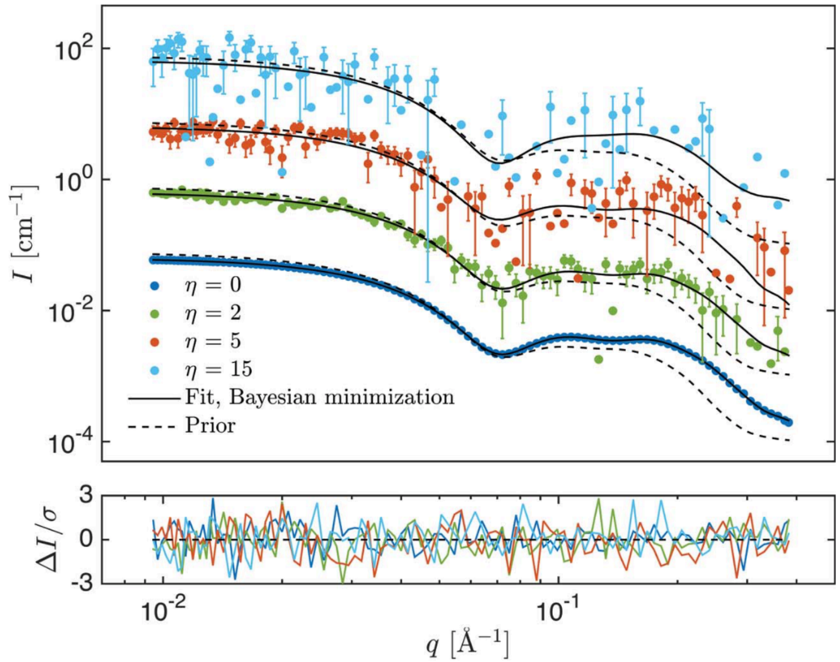
Benefits of Bayesian regularization

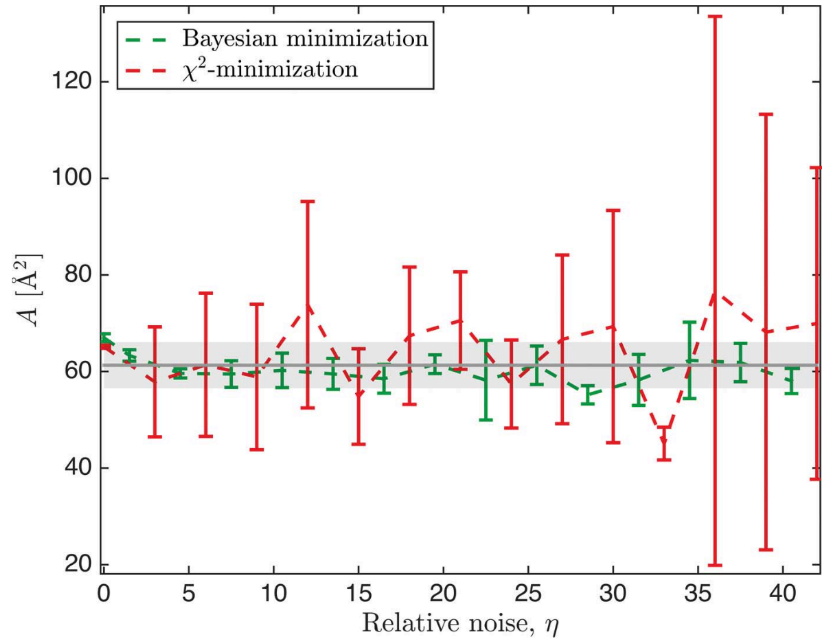
1. **Most probable solution found**
2. **Better error estimation**



Benefits of Bayesian regularization

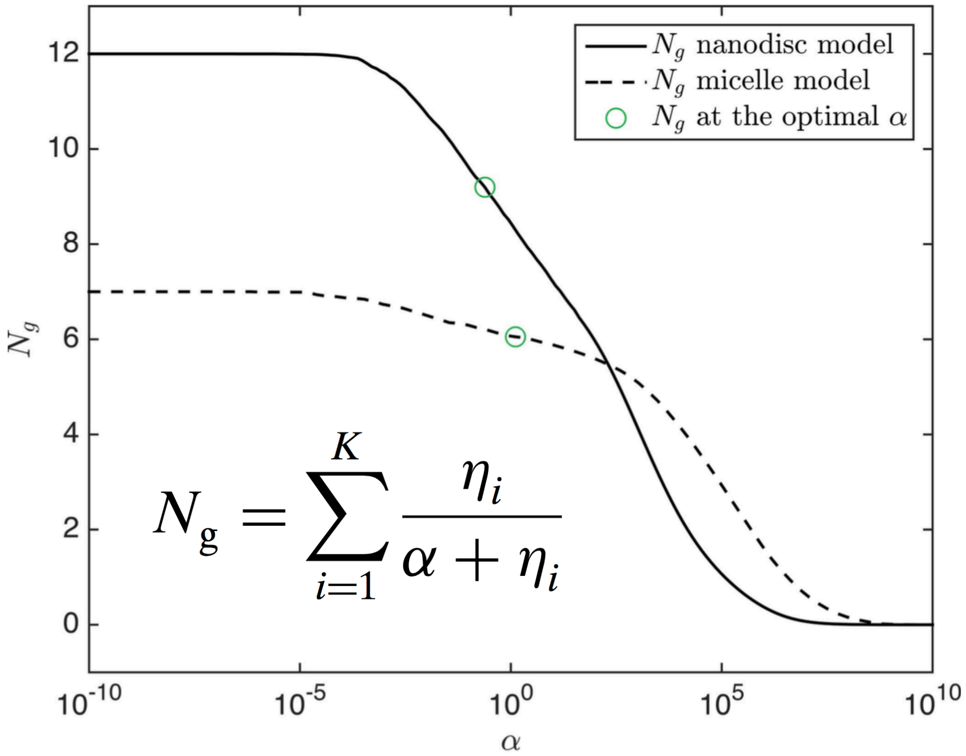
1. **Most probable solution found**
2. **Better error estimation**
3. **Stable solution for noisy data**



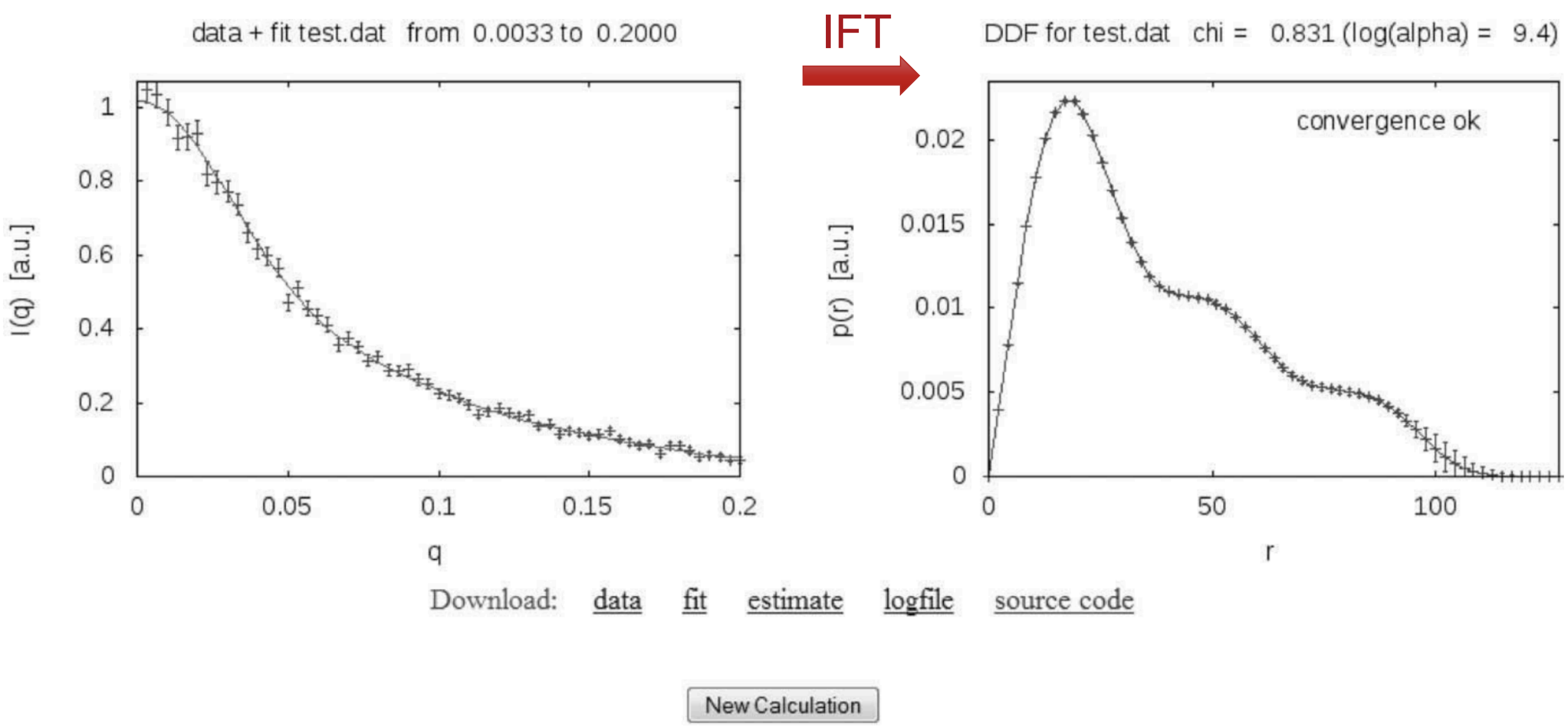


Benefits of Bayesian regularization

1. **Most probable solution found**
2. **Better error determination**
3. **Stable solution for noisy data**
4. **Measure for information**
   - include prior knowledge
   - depends on model



Bayes for Indirect Fourier Transformation



$$Q(\mathbf{p}) = \chi^2(\mathbf{p}) \underbrace{+ \alpha S(\mathbf{p})}_{\text{smoothness constraint}}$$

smoothness
constraint

www.bayesapp.org
Hansen, 2000, J. Appl. Cryst., 33, 1415-1421
Hansen, 2014, J. Appl. Cryst., 33, 1469-1471

Perspectives

- **Include in SasView, WillItFit and similar programs**
- **Combining SAXS and SANS**
- **Other techniques, e.g. reflectometry**

SasView

Will It Fit?

See more:
Larsen, Arleth & Hansen, 2018, *J. Appl. Cryst*. 51, 1151-1161.

Acknowledgements

Co-authors/ supervisors

Lise Arleth

Steen Hansen

Test data:

Pie Huda

Nicholas Skar Gislinge

Søren Roi Midtgaard

Discussions and valuable input:

Per Hedegård

Martin Cramer Pedersen

Reviewers

Funding:

DanScatt

CoNeXT

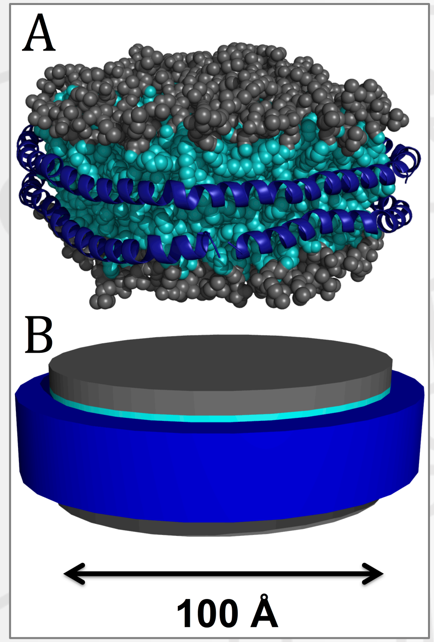
Structural Biophysics at University of Copenhagen (Arleth Group):

Thank you for your attention!

Analysis of small-angle scattering data using model fitting and Bayesian regularization

Andreas Haahr Larsen
Niels Bohr Institute
Bayes for Scattering,
May 6th 2019

UNIVERSITY OF COPENHAGEN



Model parameter	χ^2 minimization
N	103 (22)
ε	1.3 (4.5)
A (Å^2)	76 (19)
n_w	18 (12)
V_l (Å^3)	996 (19)
V_t (Å^3)	702 (111)
V_p ($\times 10^4$ Å^3)	5.3 (0.7)
T (Å)	11.4 (2.1)
R_g (Å)	13.8 (1.2)
σ_R (Å)	3.2 (1.0)
c (μM)	22 (19)
B ($\times 10^{-4}$ cm^{-1})	-0.3 (2.1)
Goodness of fit, χ^2_r	6.26
Goodness of fit, Cmap	$C = 10$, $N = 106$ $P(C \geq 10 \mid N) = 9.2\%$

Problem:
- Several possible solutions, some unphysical
- How to select the most probable solution