Small Angle X-ray Scattering Overview of basic data evaluation and analysis

Ralf Stehle
Biomoleculare NMR-Spektroskopie

Software

ATSAS Package

Collection of tools for data evaluation and modeling of SAXS and SANS data

https://www.embl-hamburg.de/biosaxs/software.html

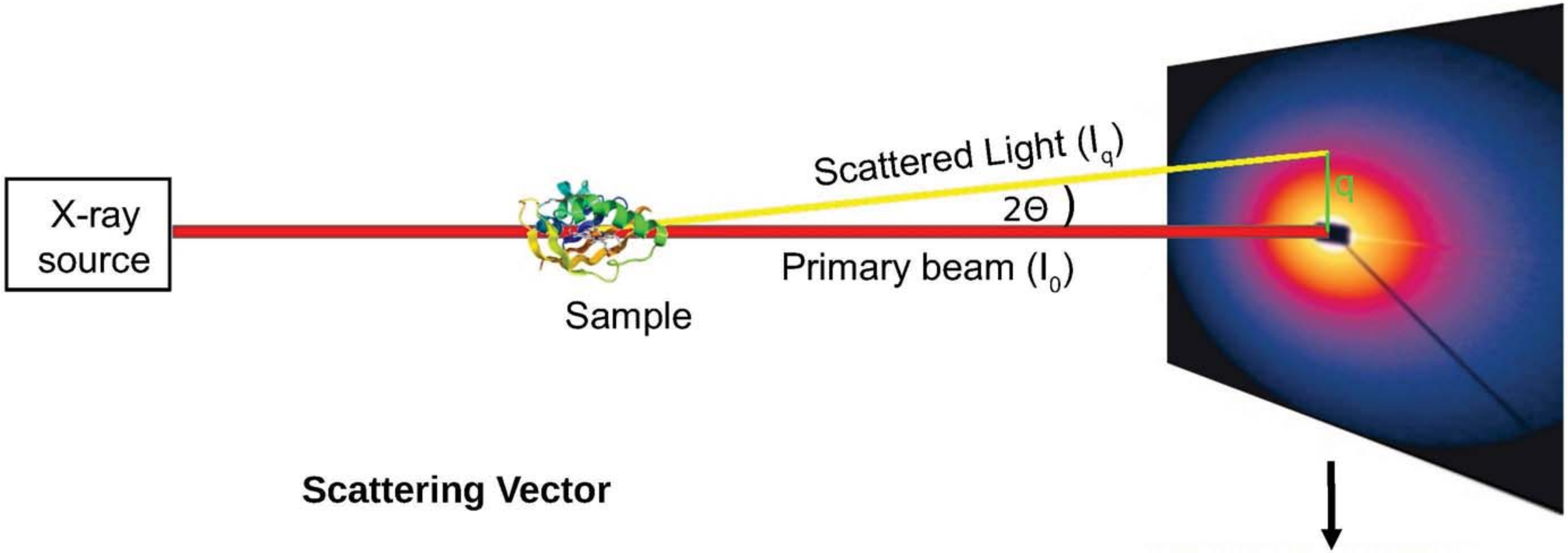
Patforms:
Linux
Windows
Mac OS

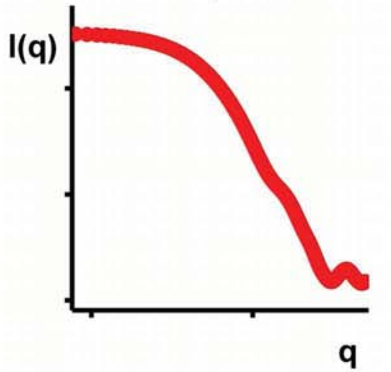
For academic use the software is free.

Download:
https://www.embl-hamburg.de/biosaxs/download.html
(Registration required)

Scattering experiment

Small Angle X-ray Scattering Experiment





Scattering Vector

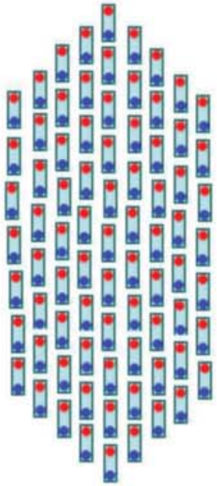
$$|\vec{q}| = \frac{4\pi}{\lambda}\sin 2\Theta \qquad \vec{q} = (\vec{k}_s - \vec{k}_0)$$

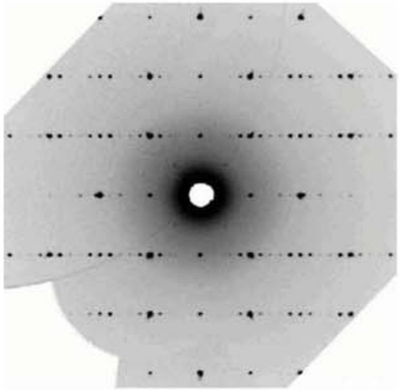
k_0, k_s Wave vectors of incident and scattered beam
λ Wavelength of incident beam

Scattering experiments

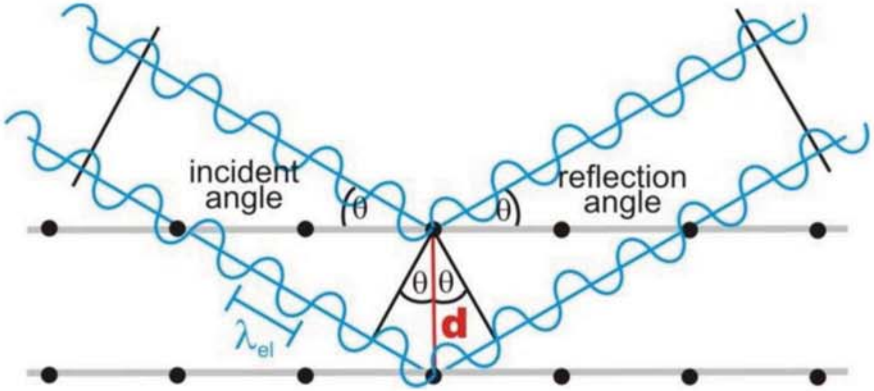
Small angle scattering and Single crystal diffraction

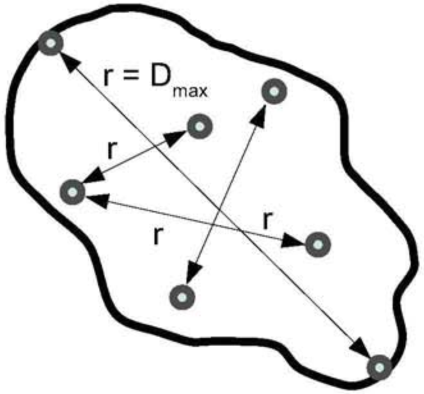
Defined orientation of molecules in a crystal



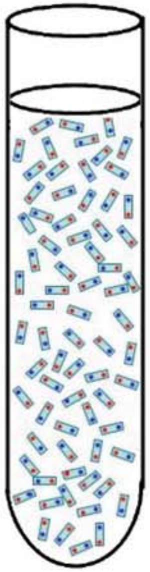


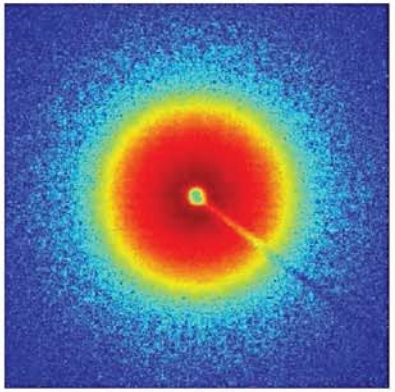
Braggs Law



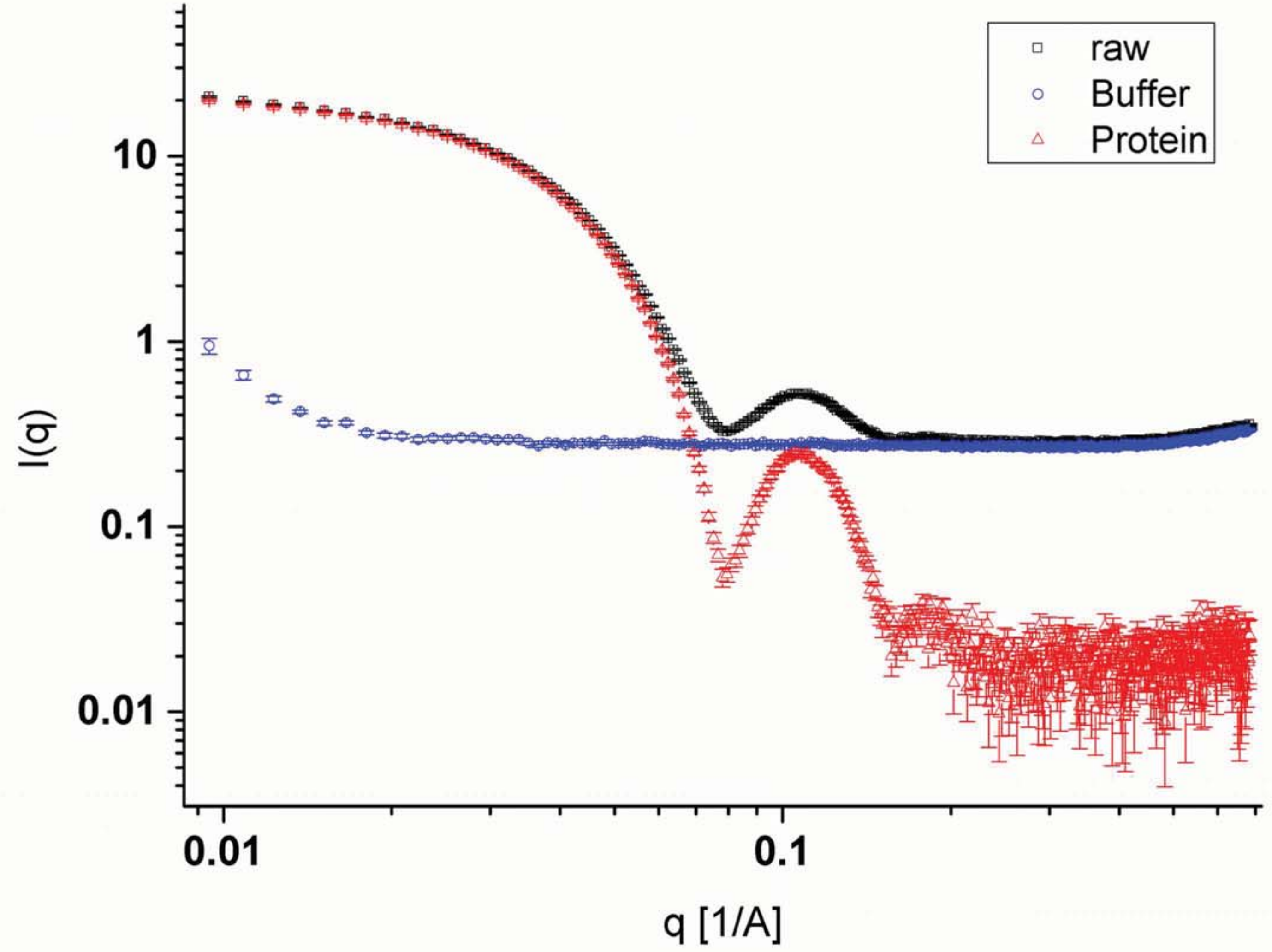


Random orientation of molecules in solution





Measurements



Direct Information

What can we learn from Small Angle X-ray Scattering Curves

What we put in = what we get out

Direct Information

What can we learn from Small Angle X-ray Scattering Curves

Garbage in = Garbage out

Direct Information

What can we learn from Small Angle X-ray Scattering Curves

More secondary information in

=

Much more information out

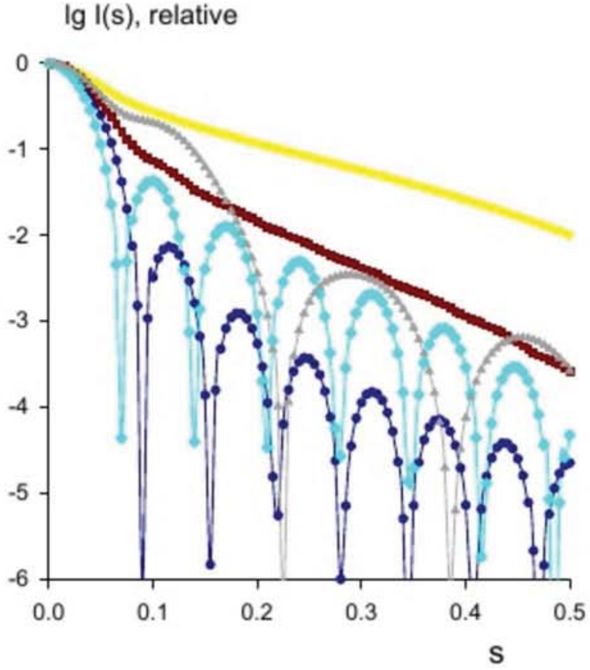
Direct Information

Direct Information

Shape

Solid sphere

Hollow sphere



Dumbbell

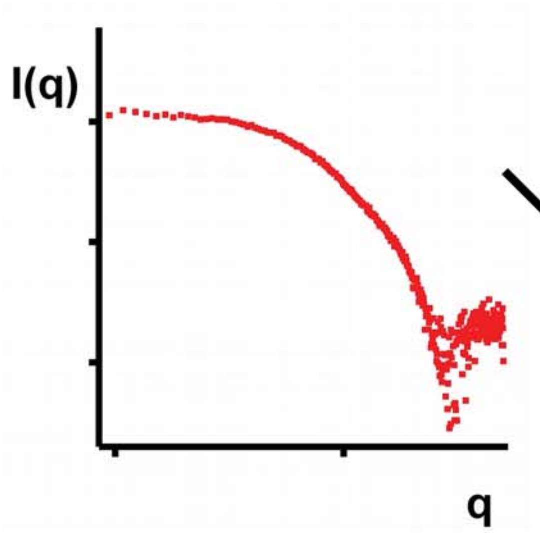
Long rod

Flat disc

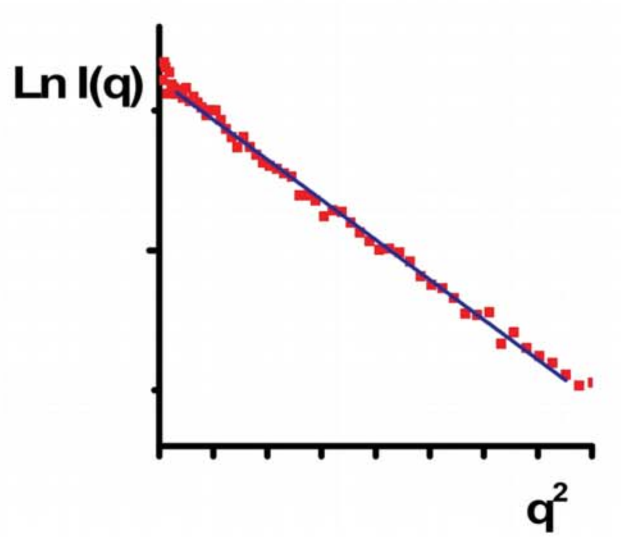
Direct Information

- **Radius of Gyration**
- **I_0 and Molecular Weight**
- **Dimerisation or ligand binding**

Scattering curve



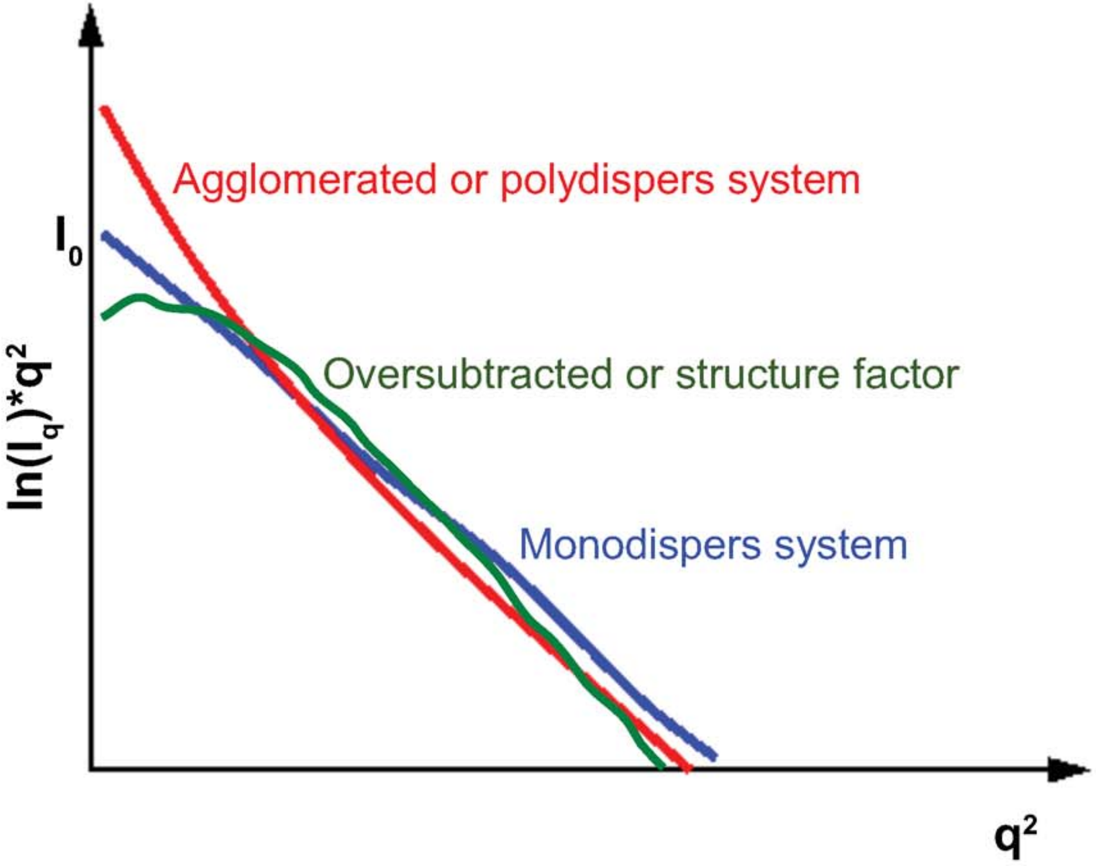
Guinier Plot



Direct Information

Guinier plot

Radius of Gyration (R_g): Average of square centers of mass distances in a particle



Objects of same mass

Guinier Approximation

$$I(q) \approx I_0 \, e^{\frac{-R_g^2}{3} q^2}$$

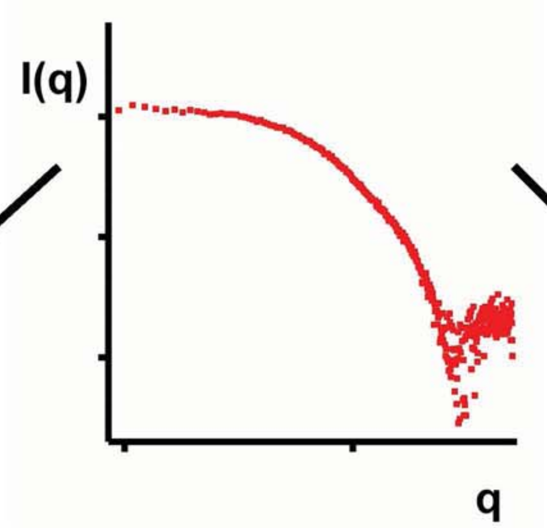
linearisation

$$\boxed{\ln(I_q) = -\frac{1}{3} R_g^2 q^2 + I_0}$$

Limit: $R_g * q < 1.3$

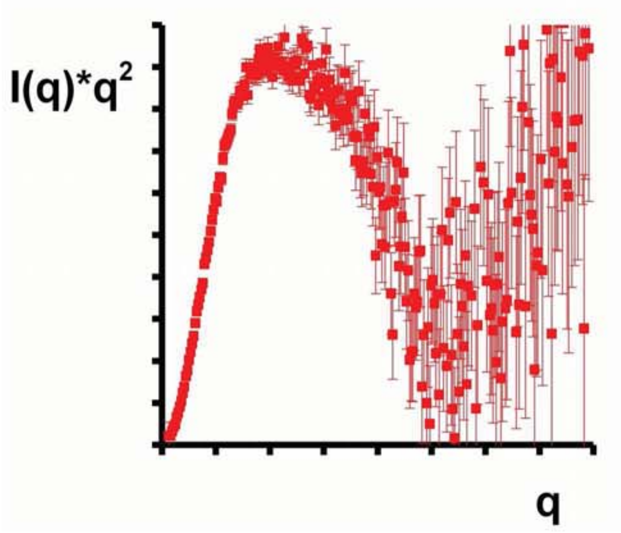
Direct Information

Scattering curve



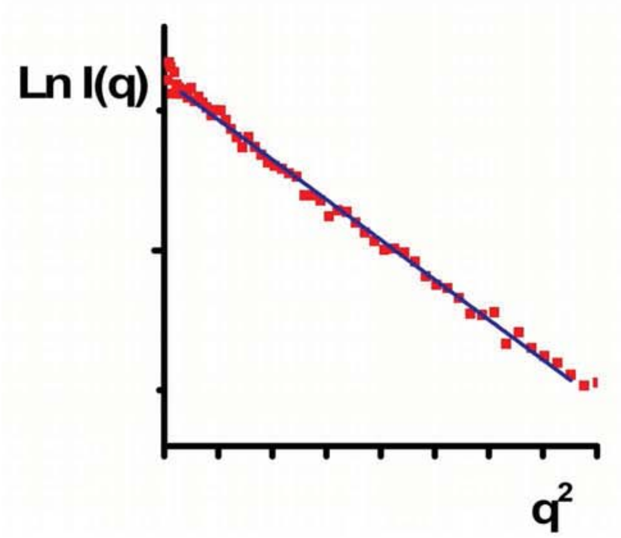
- **State of folding**
  folded
  unfolded
  partially folded

Kratky Plot



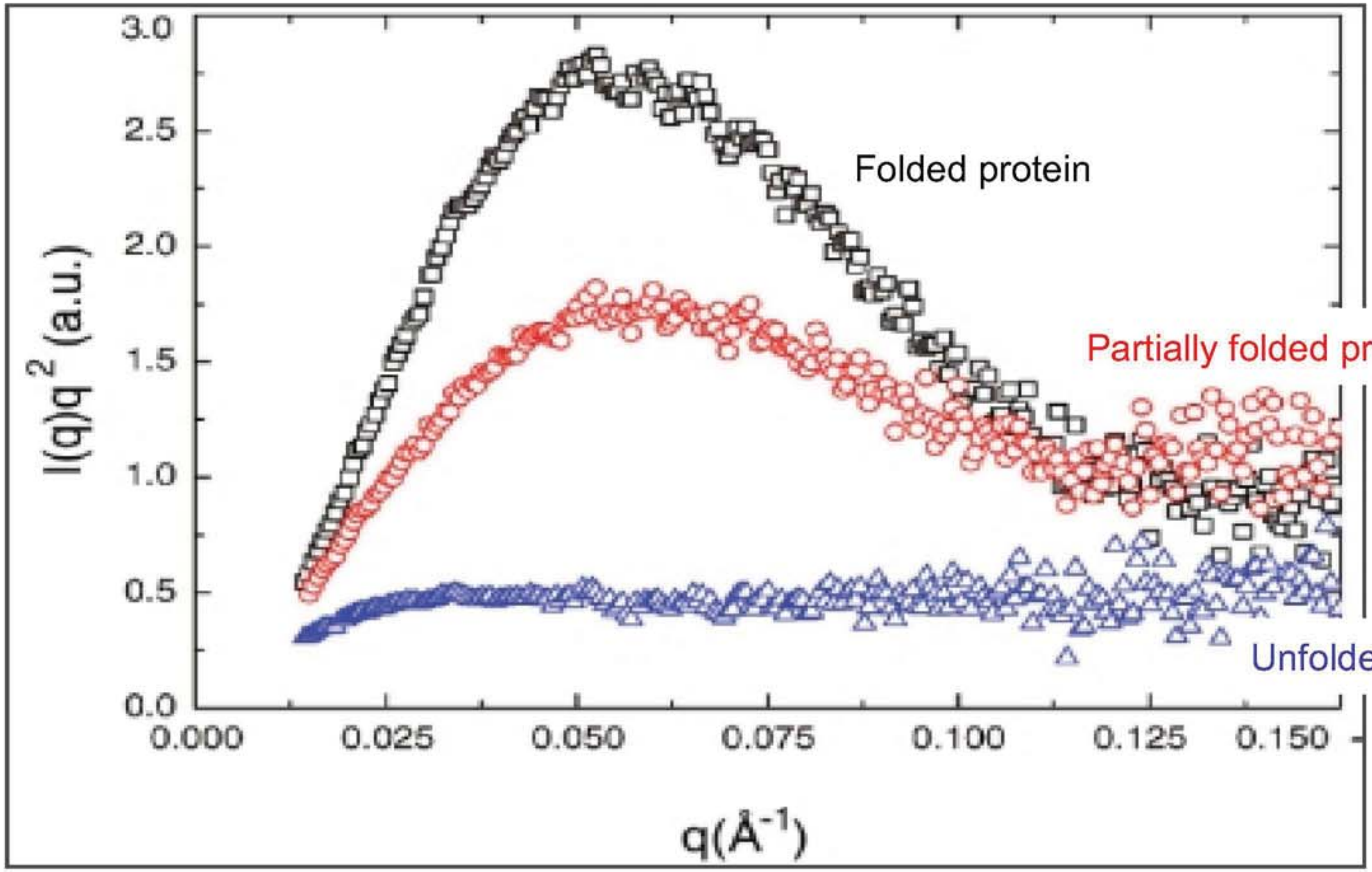
- **Radius of Gyration**
- **I_0 and Molecular Weight**
- **Dimerisation or ligand binding**

Guinier Plot



Direct Information

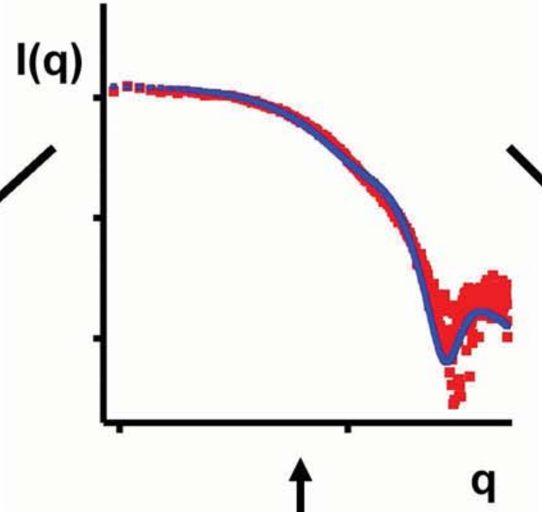
Kratky plot



The Kratky plot shows the state of folding of a protein (BSA with increasing concentrations of urea)

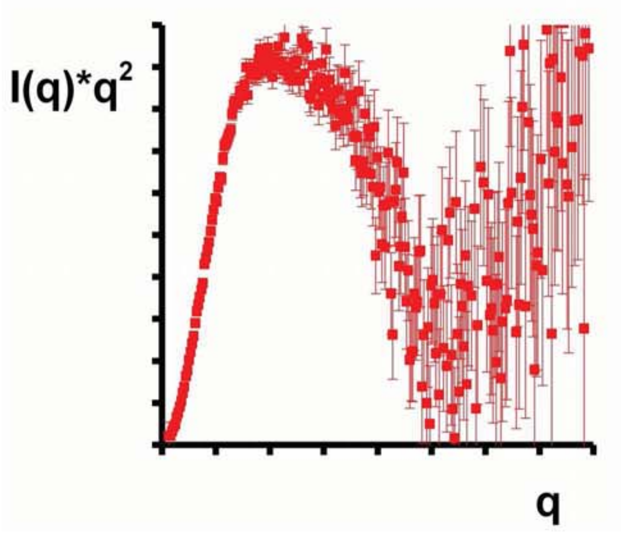
Direct Information

Fit of crystal structure to SAXS data



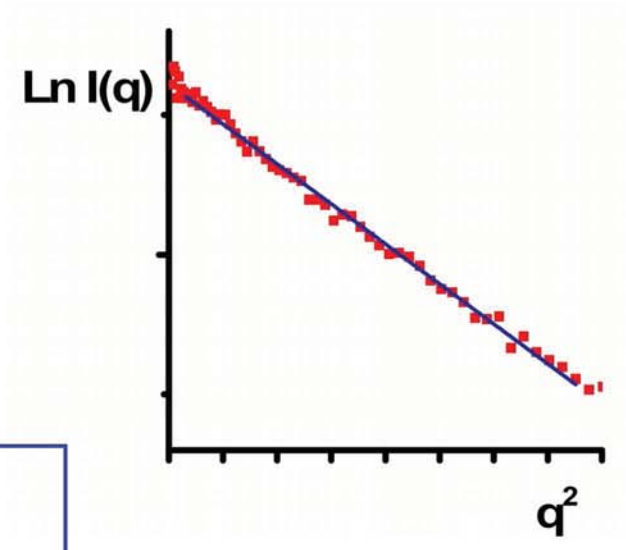
- State of folding
  folded
  unfolded
  partially folded

Kratky Plot

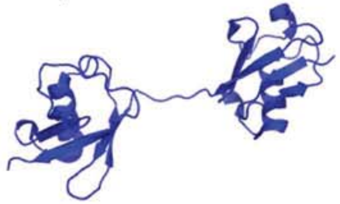


- Radius of Gyration
- I_0 and Molecular Weight
- Dimerisation or ligand binding

Guinier Plot



Crystal Structure



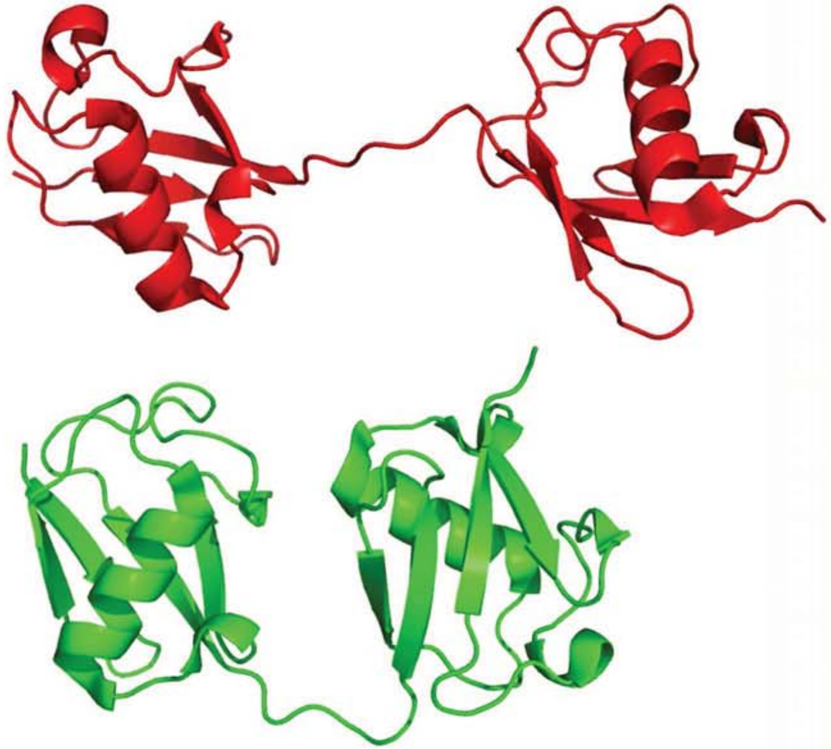
- Difference between crystal structure and structure in solution
- Dimerisation

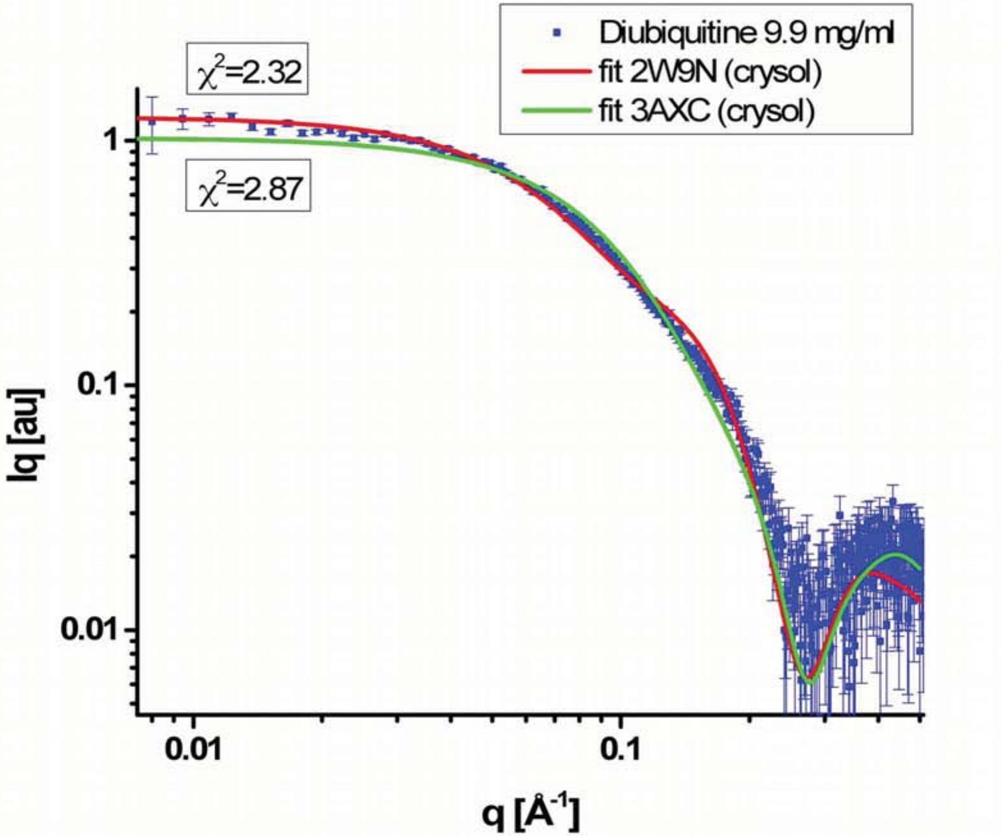
Direct Information

Combination of high resolution models with SAXS

CRYSOL

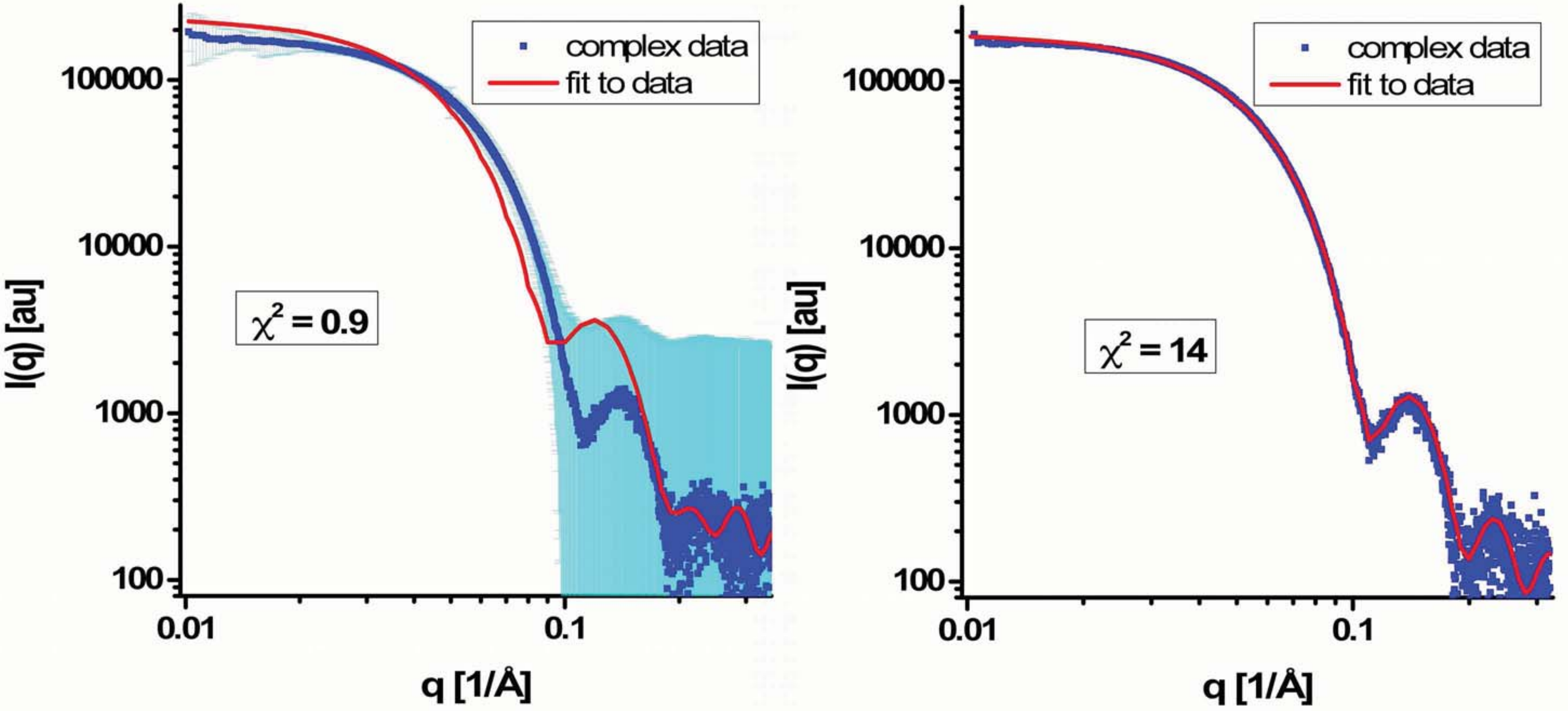
crysol calculates a scattering profile from a high resolution structure and compares to scattering data





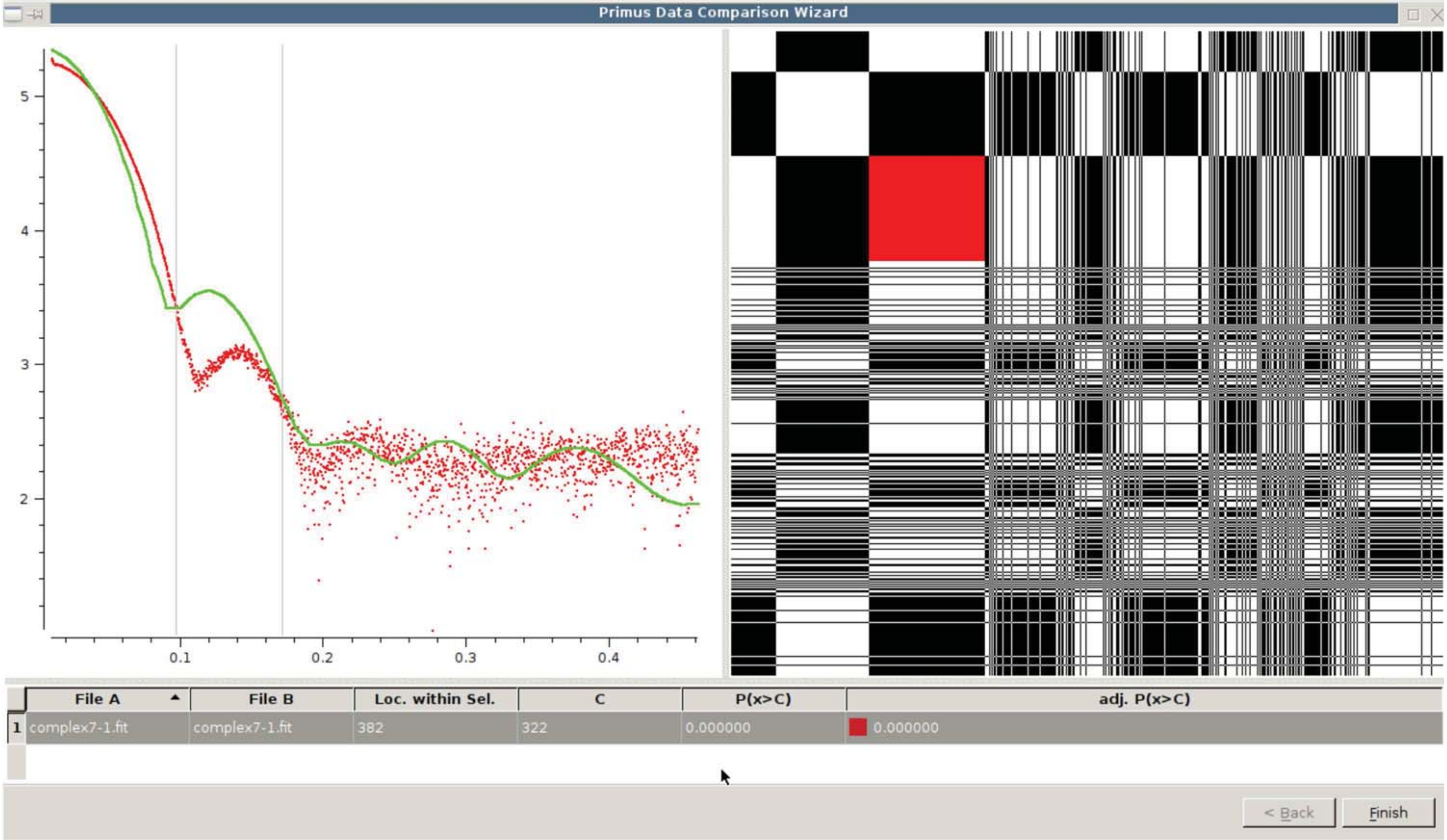
Direct Information

CRYSOL and χ^2 values



Direct Information

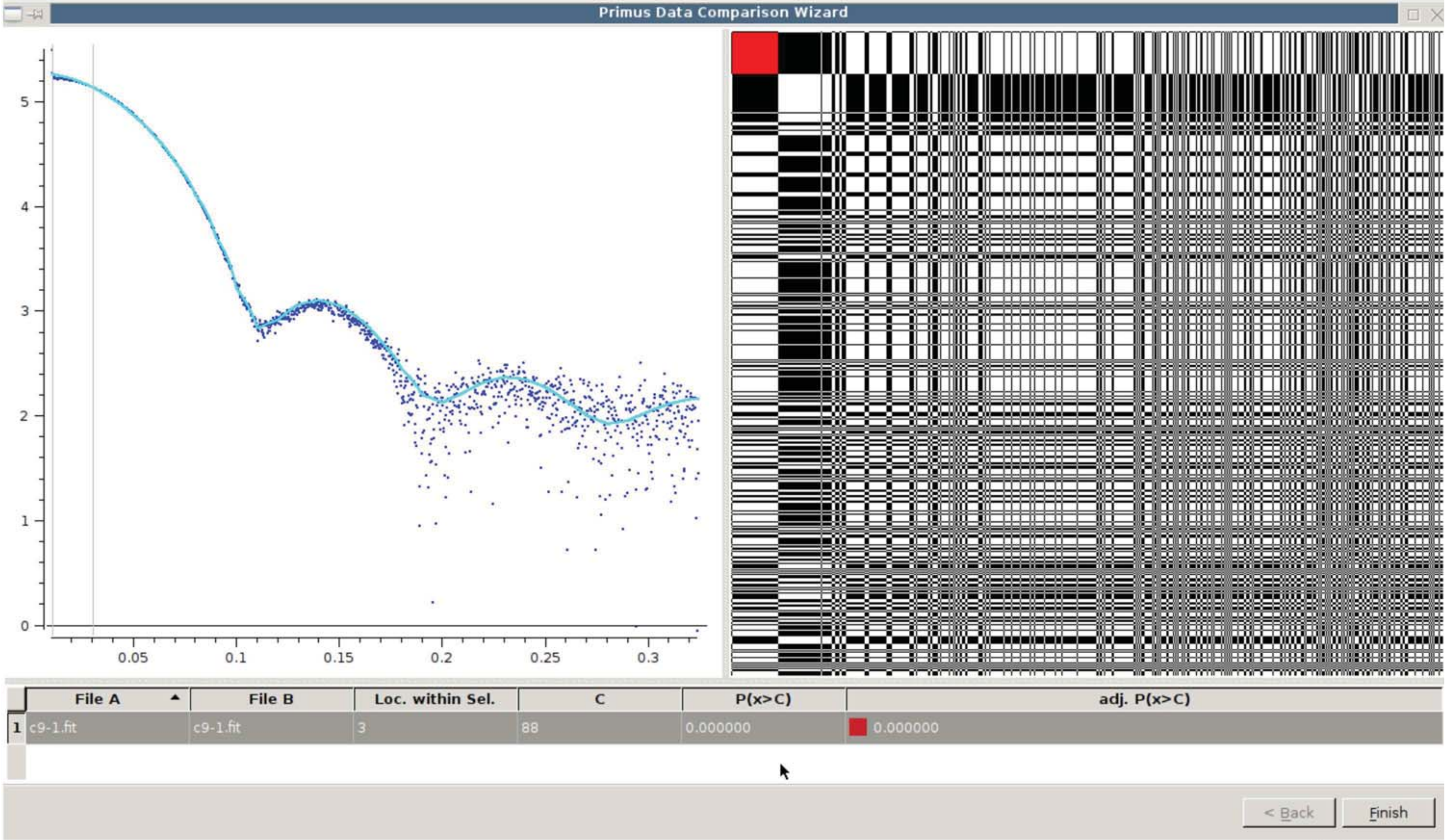
Correlation map primus



	File A ▲	File B	Loc. within Sel.	C	P(x>C)	adj. P(x>C)
1	complex7-1.fit	complex7-1.fit	382	322	0.000000	🟥 0.000000

Direct Information

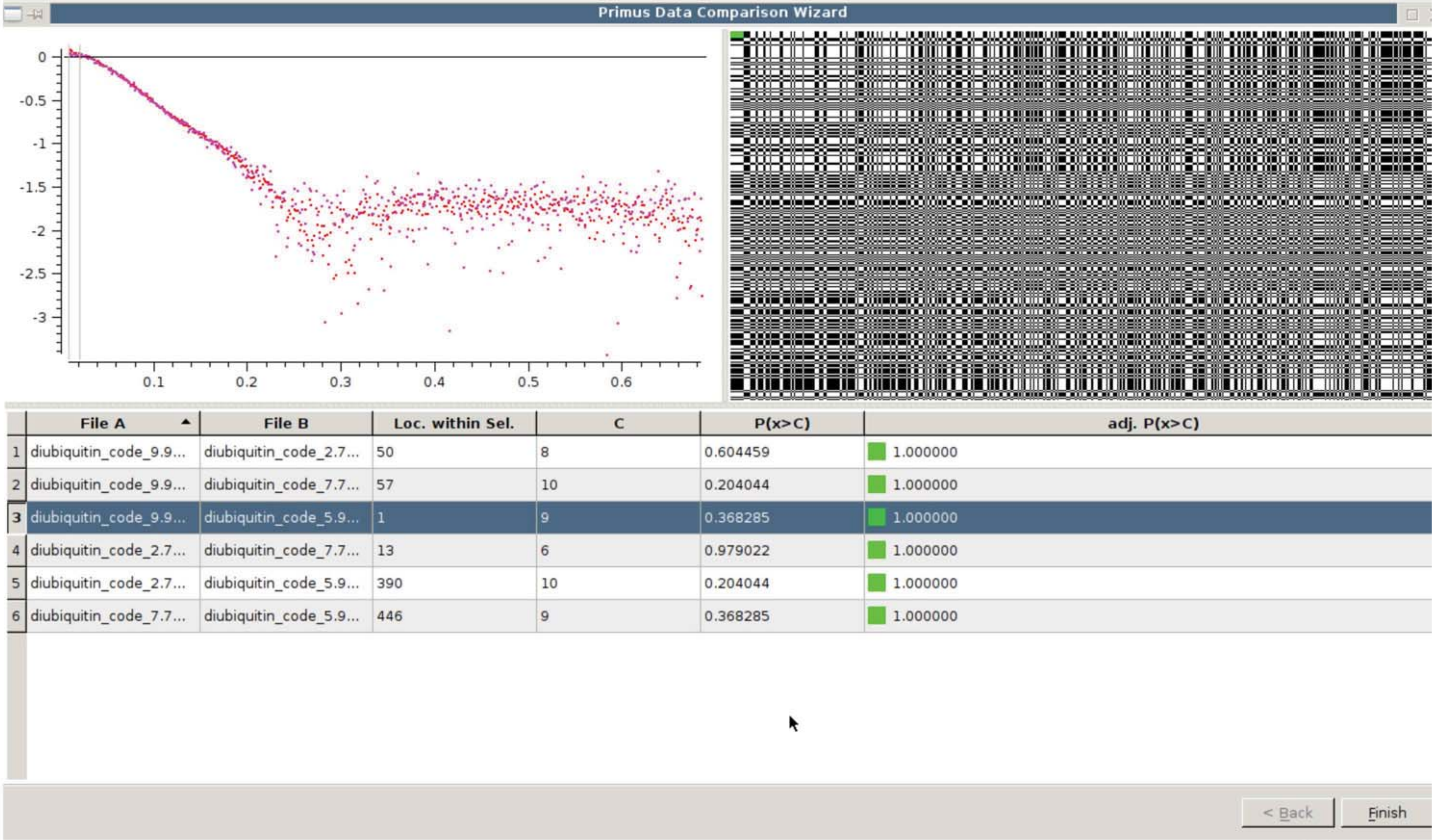
Correlation map primus



	File A ▲	File B	Loc. within Sel.	C	P(x>C)	adj. P(x>C)
1	c9-1.fit	c9-1.fit	3	88	0.000000	🟥 0.000000

Direct Information

Correlation map primus

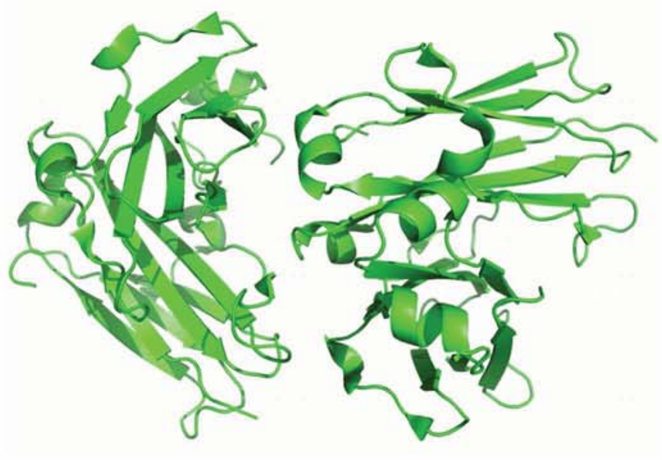


	File A	File B	Loc. within Sel.	C	P(x>C)	adj. P(x>C)
1	diubiquitin_code_9.9...	diubiquitin_code_2.7...	50	8	0.604459	🟩 1.000000
2	diubiquitin_code_9.9...	diubiquitin_code_7.7...	57	10	0.204044	🟩 1.000000
3	diubiquitin_code_9.9...	diubiquitin_code_5.9...	1	9	0.368285	🟩 1.000000
4	diubiquitin_code_2.7...	diubiquitin_code_7.7...	13	6	0.979022	🟩 1.000000
5	diubiquitin_code_2.7...	diubiquitin_code_5.9...	390	10	0.204044	🟩 1.000000
6	diubiquitin_code_7.7...	diubiquitin_code_5.9...	446	9	0.368285	🟩 1.000000

Direct Information

Dimerisation oligomer

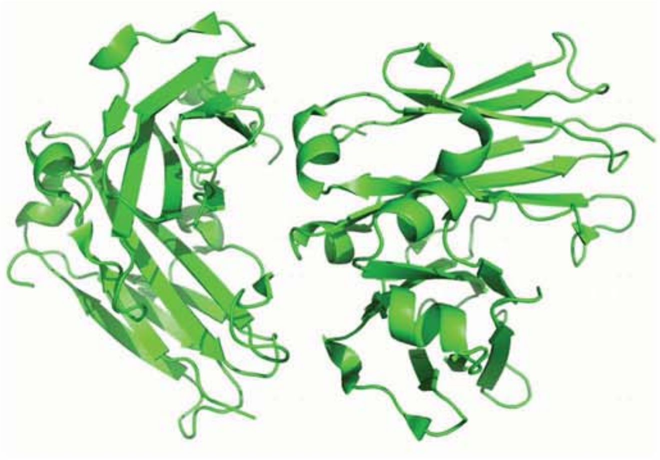
IGMCH4 Dimer

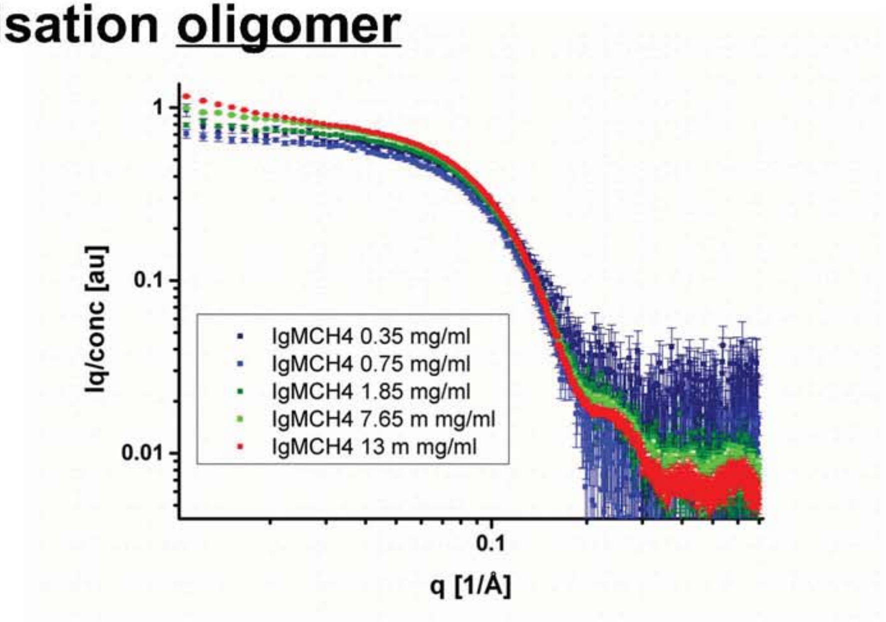


Direct Information

Dimerisation oligomer

IGMCH4 Dimer

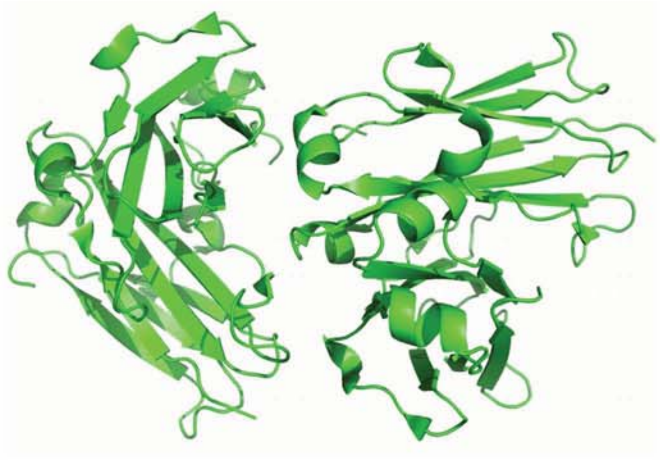


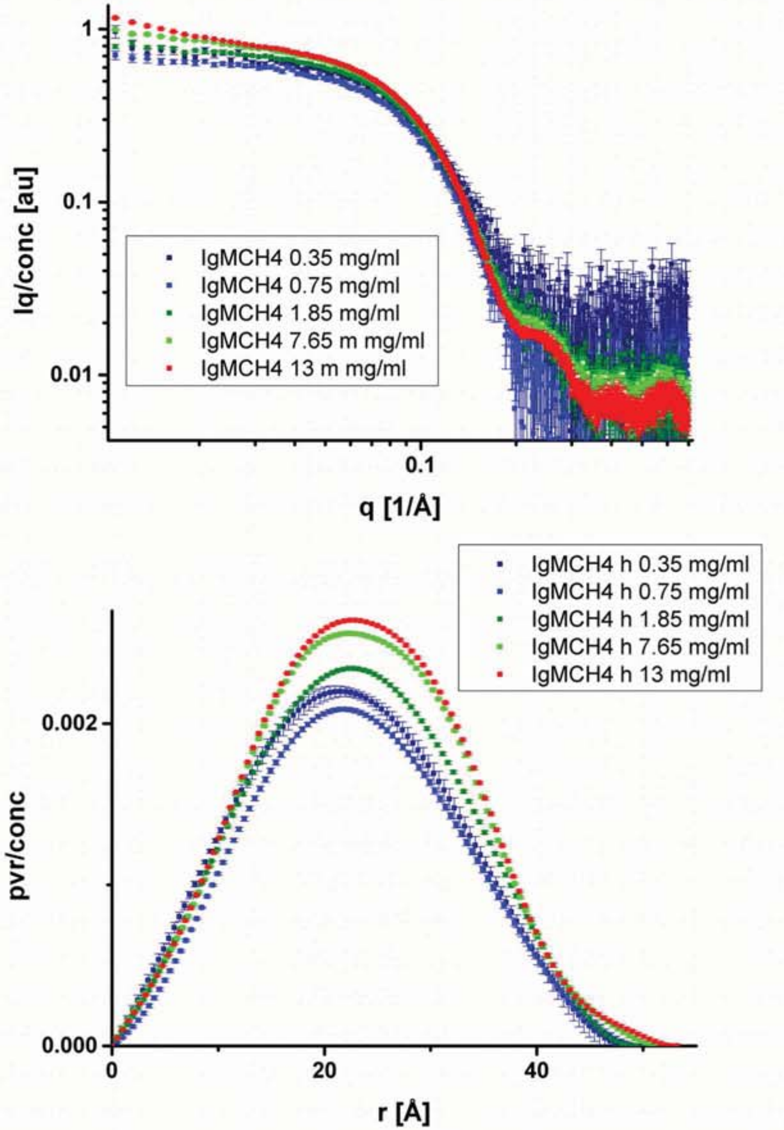


Direct Information

Dimerisation oligomer

IGMCH4 Dimer

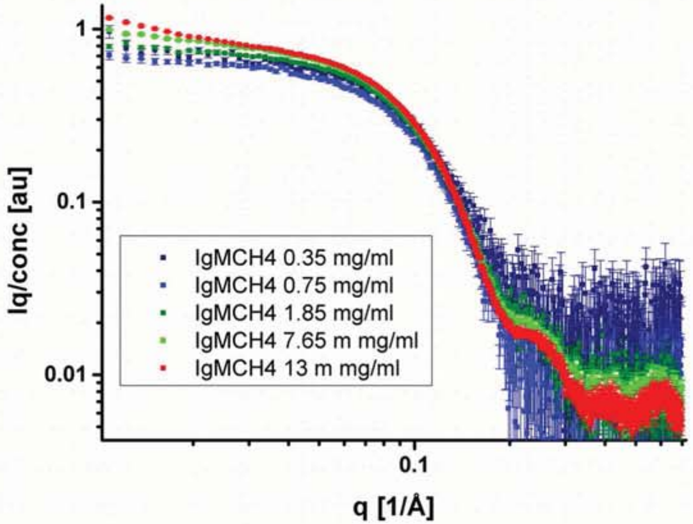


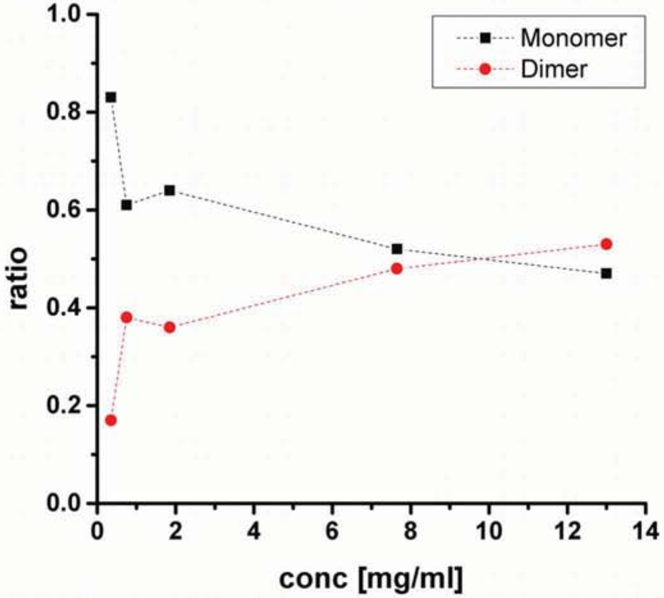


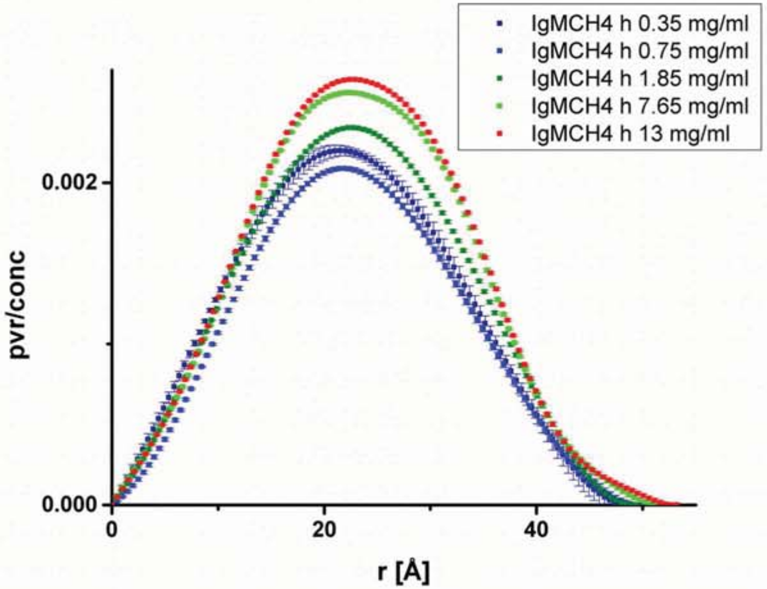
Direct Information

Dimerisation oligomer

IGMCH4 Dimer







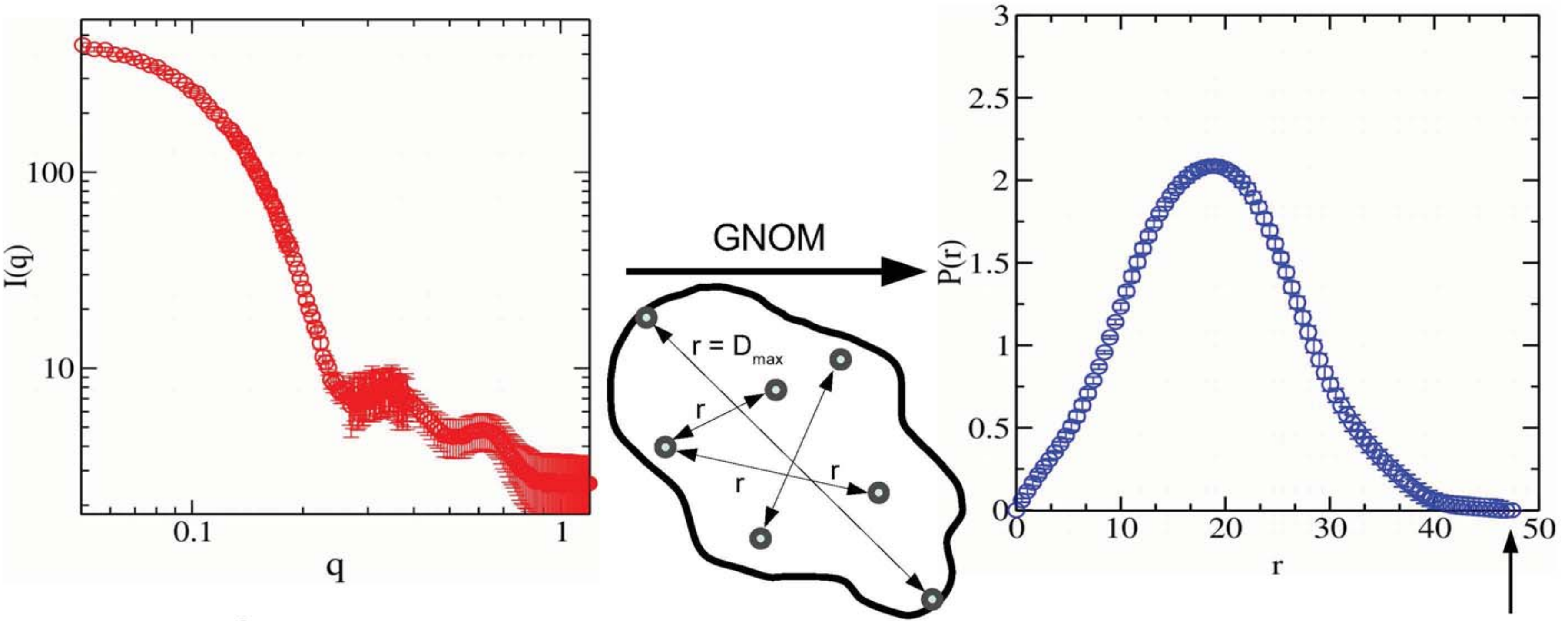
Modeling

Modeling

Transformation of scattering curves to real space

Transformation of the scattering curve from reciprocal space to real space with the programm GNOM to get the

pair distance distribution function p(r) (PDDF)

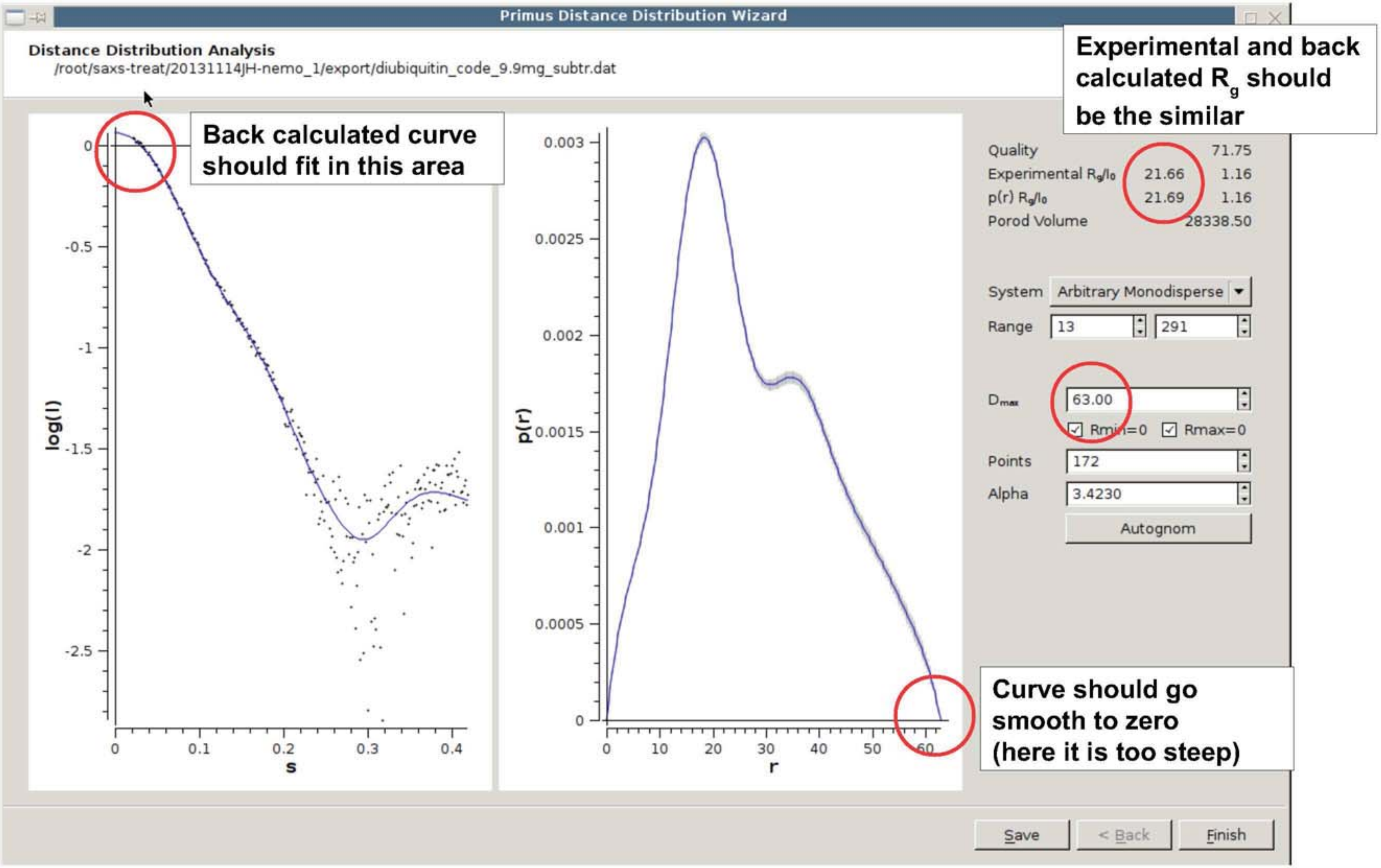


$$p(r)=\frac{r^2}{2\pi^2}\int_0^\infty q^2 I(q)\frac{\sin qr}{qr}\,dq$$

$$I(0)=4\pi\int_0^{D\max} p(r)\,dr$$

Modeling

Transformation of scattering curves to real space with GNOM



Primus Distance Distribution Wizard

Distance Distribution Analysis

/root/saxs-treat/20131114JH-nemo_1/export/diubiquitin_code_9.9mg_subtr.dat

Back calculated curve should fit in this area

Experimental and back calculated R_g should be the similar

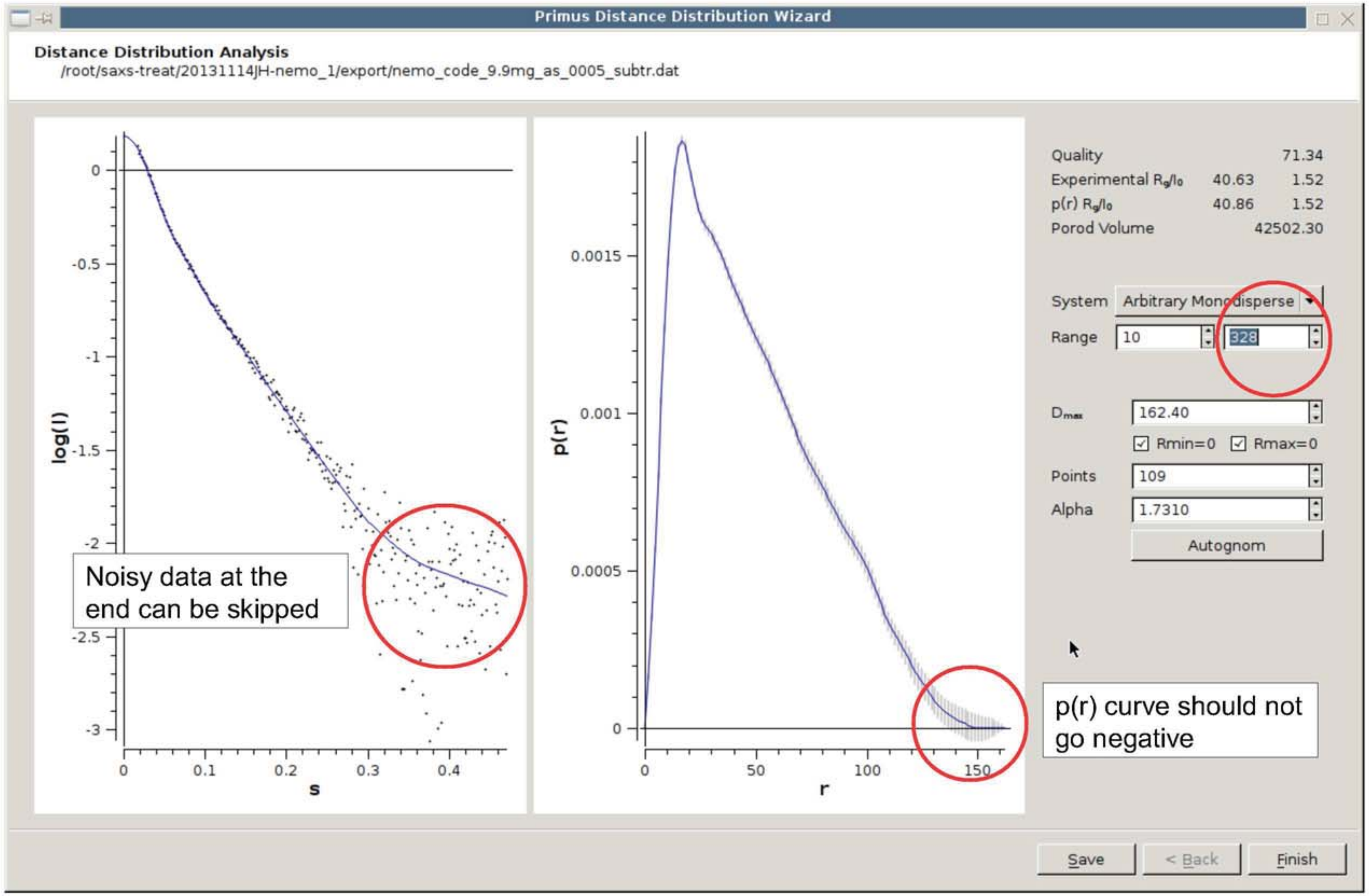
Quality 71.75
Experimental R_g/I_0 21.66 1.16
p(r) R_g/I_0 21.69 1.16
Porod Volume 28338.50

System Arbitrary Monodisperse
Range 13 291
D_{max} 63.00
☑ Rmin=0 ☑ Rmax=0
Points 172
Alpha 3.4230

Curve should go smooth to zero (here it is too steep)

Modeling

Transformation of scattering curves to real space with GNOM



Primus Distance Distribution Wizard

Distance Distribution Analysis
/root/saxs-treat/20131114JH-nemo_1/export/nemo_code_9.9mg_as_0005_subtr.dat

Quality 71.34
Experimental R_g/I_0 40.63 1.52
p(r) R_g/I_0 40.86 1.52
Porod Volume 42502.30

System Arbitrary Monodisperse
Range 10 328

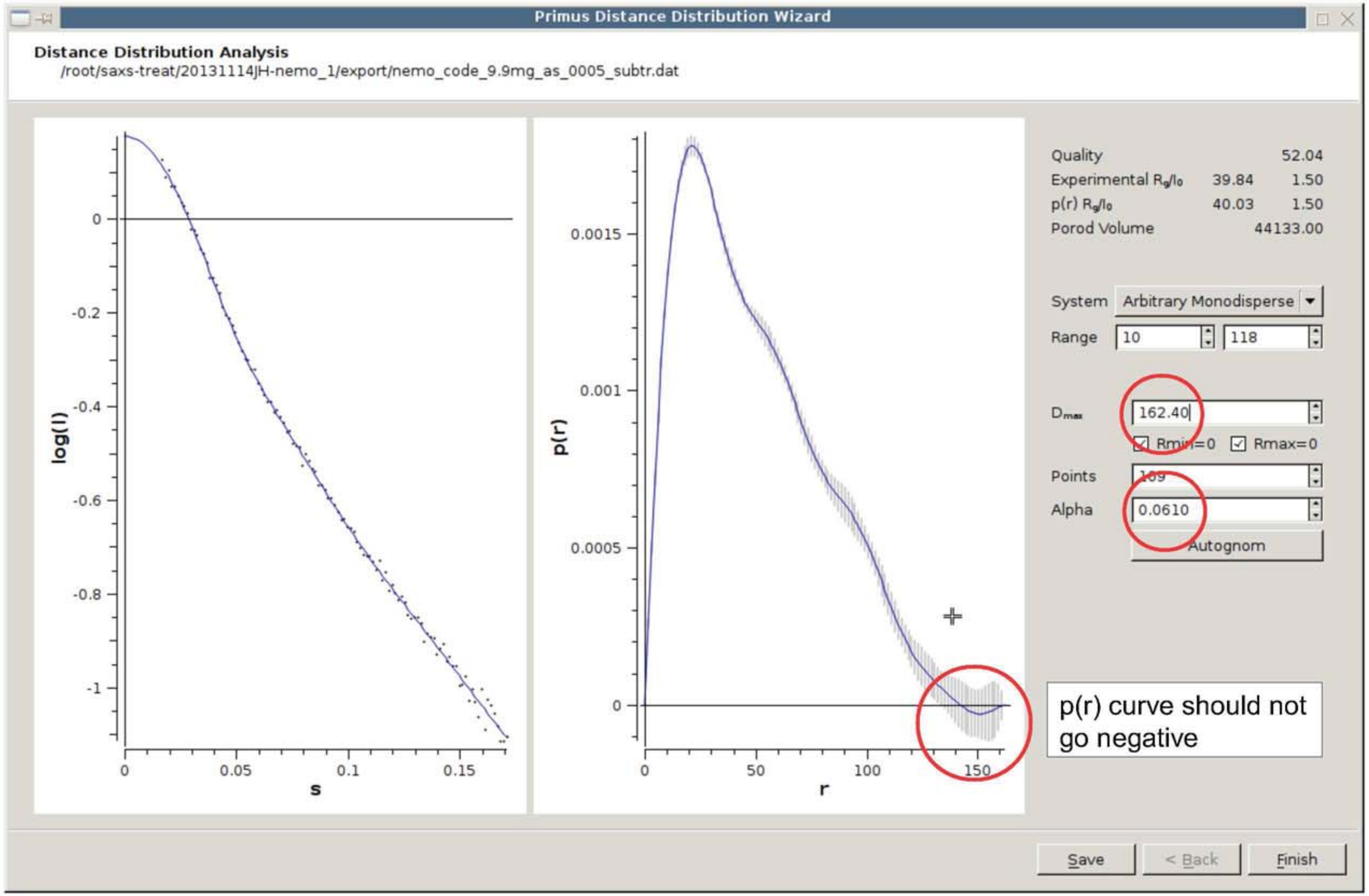
D_{max} 162.40
☑ Rmin=0 ☑ Rmax=0
Points 109
Alpha 1.7310

Noisy data at the end can be skipped

p(r) curve should not go negative

Modeling

Transformation of scattering curves to real space with GNOM



Primus Distance Distribution Wizard

Distance Distribution Analysis
/root/saxs-treat/20131114JH-nemo_1/export/nemo_code_9.9mg_as_0005_subtr.dat

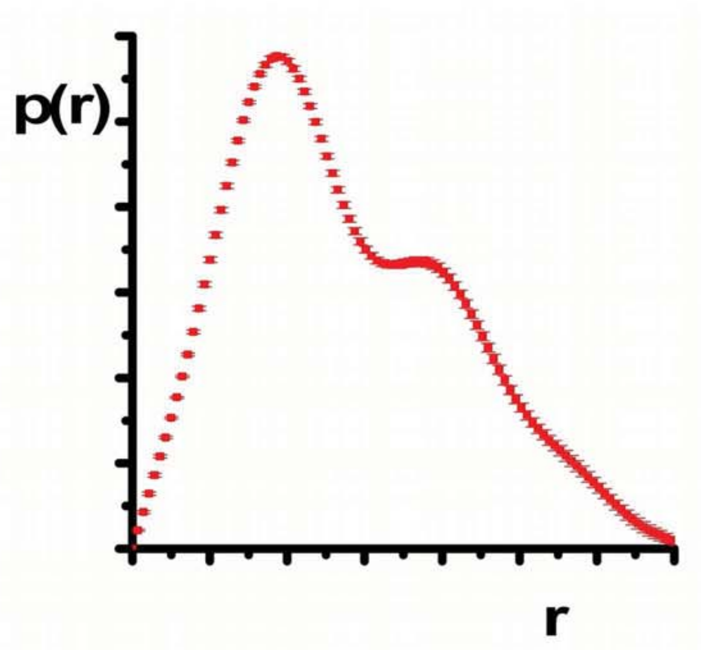
Quality		52.04
Experimental R_g/I_0	39.84	1.50
p(r) R_g/I_0	40.03	1.50
Porod Volume		44133.00

System Arbitrary Monodisperse
Range 10 118
D_{max} 162.40
☑ Rmin=0 ☑ Rmax=0
Points 109
Alpha 0.0610

p(r) curve should not go negative

Modeling

Calculating ab initio models

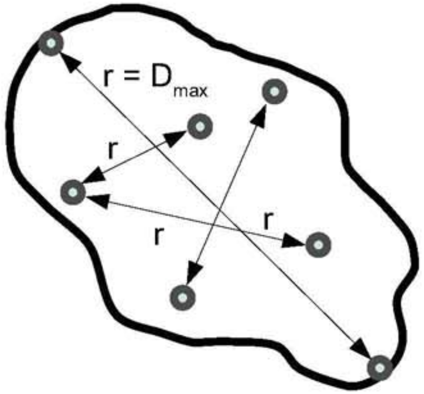


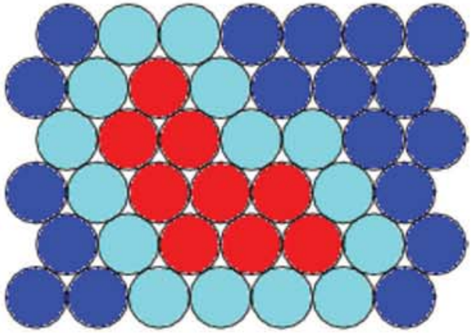
Dammin initially 2000 beads fixed Volume. Each bead can be flipped. Lots of symmetries allowed.

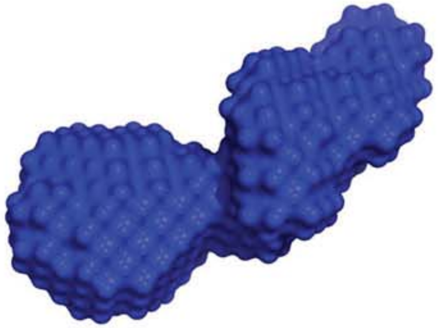
Dammif flexible Volume. Only beads at the phase border can flip. Few symmetries.

Both have problems with flat disks. To avoid this force symmetry (P5, P6)

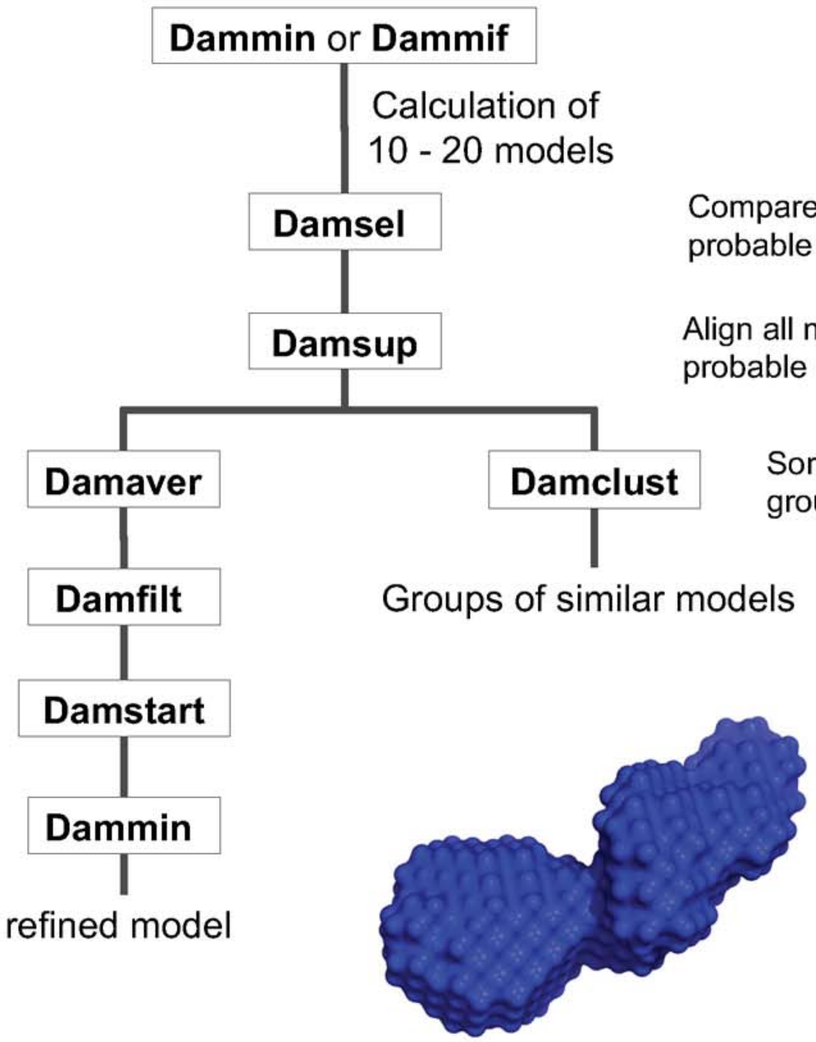
10-20 models are calculated normally. They are superimposed merged and the result is filtered to the excluded volume of the particle.







Modeling



Dammin or **Dammif**

Ab initio shape determination by simulated annealing using a single phase dummy atom model

Calculation of 10 - 20 models

Damsel

Compare all models, find most probable one and outliers

Damsup

Align all models with the most probable one

Damaver

Average aligned models and compute probability map

Damclust

Sorts similar models into groups

Groups of similar models

Damfilt

Filters the averaged model at a given cut-off volume

Damstart

Generates an input file from the averaged model with fixed core for dammin

Dammin

refined model

Rigid body modeling

Rigid body modeling

Fit of partial structures to SAXS data

sasref fit of complexes. Input multiple chains. Symetry and distance constrains possible.

coral fit of multi domain proteins. Lines up domains to one chain. Flexible linkers are taken from a database. Symetry and distance constrains possible.

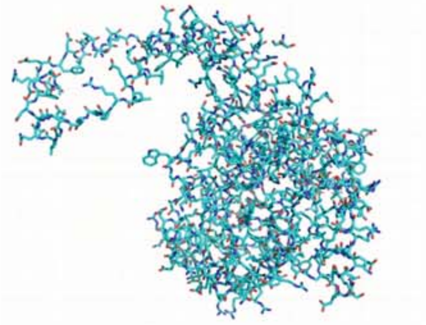
bunch fit of multi domain proteins. Needs a sequence including missing and flexible parts. Domains are rotated around residues of flexible parts. Distance constrains possible.

eom ensemble calculation. Produces 10000 structures clusters them to “chromosomes” and fits them with an evolutionary approach to the experimental data. Lines up domains to one chain. Flexible linkers have to be defined. Symetry and distance constrains possible in a limited way.

Rigid body modeling

Fit of partial structures to SAXS data: sasref

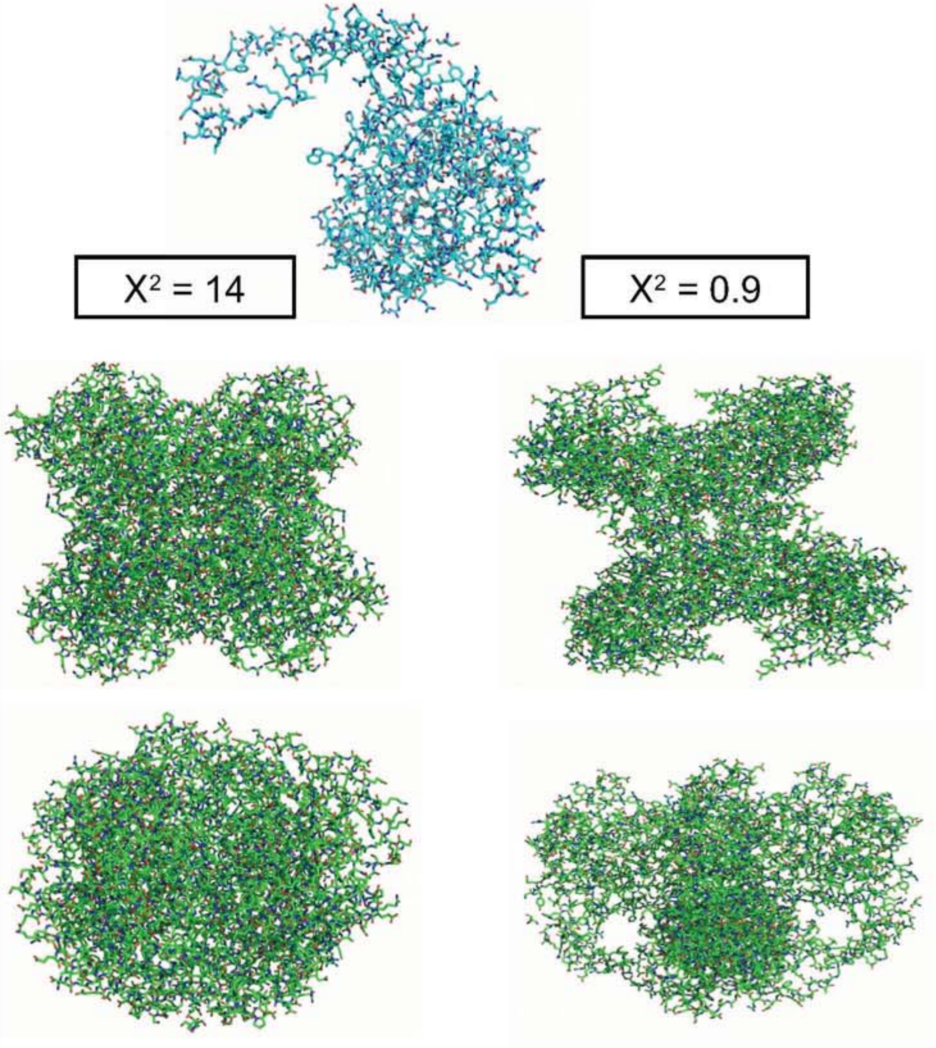
```
$ sasref
  === SASREF06  started at          21-May-2015   10:42:49
 Computation mode ....................................... : User
 Project identificator .................................. : c9
 Project decsription .................................... : c9
 Random sequence initialized from ...................... : 104249989
 Total number of subunits ............................... : 1
 Point symmetry of the particle ......................... : P222
 Number of equivalent positions ........................ : 4
 Total number of scattering curves ...................... : 1
 1-st construct, of the first and the last subunits: 1,1
 File name, 1-st experimental data ...................... : sub_004.dat
 Number of experimental points found ................... : 1370
 Angular units multiplied by ........................... : 0.1000
 Experimental radius of gyration ....................... : 248.0
 Number of points in the Guinier Plot .................. : 5
 Amplitudes, 1-st subunit ............................... : mon00.alm
 Number of points in partial ampliudes ................. : 51
 Maximum order of harmonics ............................ : 15
 Evaluated from the file ............................... : mon.pdb
 Current subunit:   3089 atoms read, center at      0.03   -0.02    0.11
 Spatial step in angstroems ............................ : 5.000
 Angular step in degrees ............................... : 20.00
 Cross value ........................................... : 68.20
 Dock value ............................................ : 0.0
 Discontiguity value ................................... : 0.0
 Cross penalty weight .................................. : 10.00
 Disconnectivity penalty weight ........................ : 10.00
 Docking penalty weight ................................ : 0.5000
 Expected particle anisometry .......................... : Unknown
 .
 .
```



Rigid body modeling

Fit of partial structures to SAXS data: <u>sasref</u>

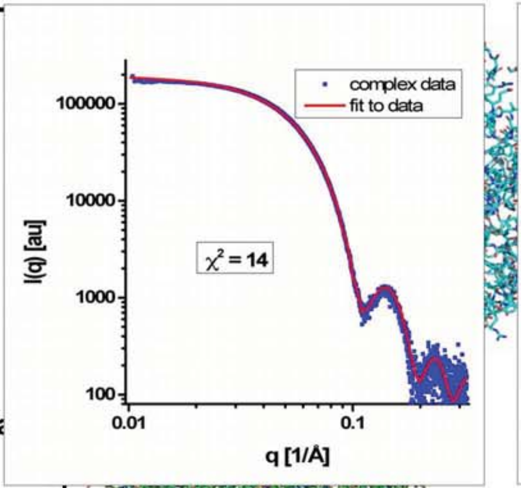
```
$ sasref
  === SASREF06  started at          21-May-2015   10:42:49
 Computation mode ...................................... : User
 Project identificator ................................. : c9
 Project decsription ................................... : c9
 Random sequence initialized from ...................... : 104249989
 Total number of subunits .............................. : 1
 Point symmetry of the particle ........................ : P222
 Number of equivalent positions ........................ : 4
 Total number of scattering curves ..................... : 1
 1-st construct, of the first and the last subunits: 1,1
 File name, 1-st experimental data ..................... : sub_004.dat
 Number of experimental points found ................... : 1370
 Angular units multiplied by ........................... : 0.1000
 Experimental radius of gyration ....................... : 248.0
 Number of points in the Guinier Plot .................. : 5
 Amplitudes, 1-st subunit .............................. : mon00.alm
 Number of points in partial ampliudes ................. : 51
 Maximum order of harmonics ............................ : 15
 Evaluated from the file ............................... : mon.pdb
 Current subunit:   3089 atoms read, center at      0.03   -0.02    0.11
 Spatial step in angstroems ............................ : 5.000
 Angular step in degrees ............................... : 20.00
 Cross value ........................................... : 68.20
 Dock value ............................................ : 0.0
 Discontiguity value ................................... : 0.0
 Cross penalty weight .................................. : 10.00
 Disconnectivity penalty weight ........................ : 10.00
 Docking penalty weight ................................ : 0.5000
 Expected particle anisometry .......................... : Unknown
.
```

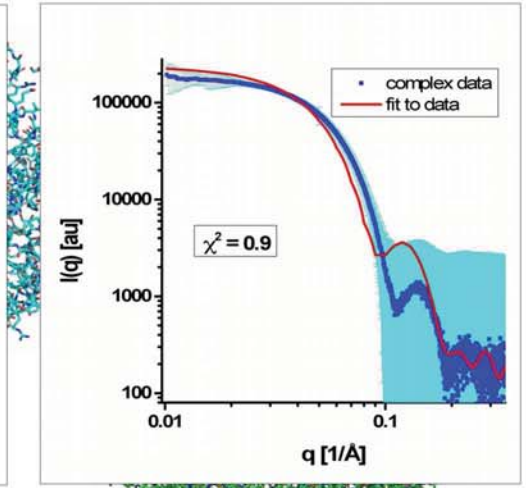


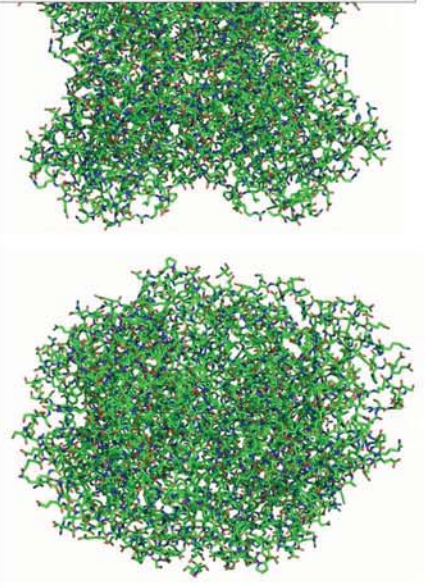
Rigid body modeling

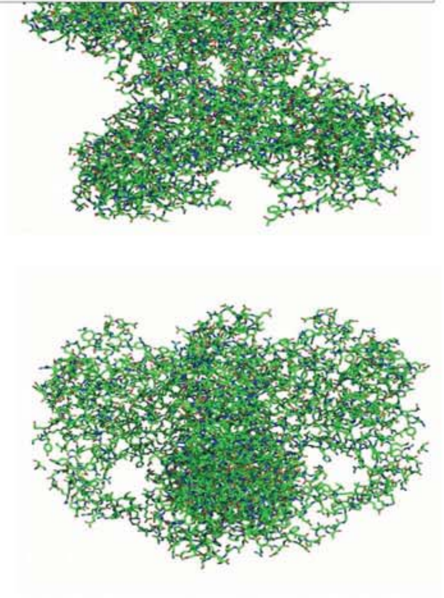
Fit of partial structures to SAXS data: sasref

```
$ sasref
  === SASREF06  started at          21-May-2015   10:42:49
 Computation mode ...................................... : User
 Project identificator ................................. : c9
 Project decsription ................................... : c9
 Random sequence initialized from ....................... : 104249989
 Total number of subunits ............................... : 1
 Point symmetry of the particle ......................... : P222
 Number of equivalent positions ......................... : 4
 Total number of scattering curves ...................... : 1
 1-st construct, of the first and the last subunits: 1,1
 File name, 1-st experimental data ...................... : sub_004.da
 Number of experimental points found .................... : 1370
 Angular units multiplied by ............................ : 0.1000
 Experimental radius of gyration ........................ : 248.0
 Number of points in the Guinier Plot ................... : 5
 Amplitudes, 1-st subunit .............................. : mon00.alm
 Number of points in partial ampliudes .................. : 51
 Maximum order of harmonics ............................. : 15
 Evaluated from the file ................................ : mon.pdb
 Current subunit:   3089 atoms read, center at     0.03   -0.02    0.11
 Spatial step in angstroems ............................. : 5.000
 Angular step in degrees ................................ : 20.00
 Cross value ............................................ : 68.20
 Dock value ............................................. : 0.0
 Discontiguity value .................................... : 0.0
 Cross penalty weight ................................... : 10.00
 Disconnectivity penalty weight ......................... : 10.00
 Docking penalty weight ................................. : 0.5000
 Expected particle anisometry ........................... : Unknown
.
```





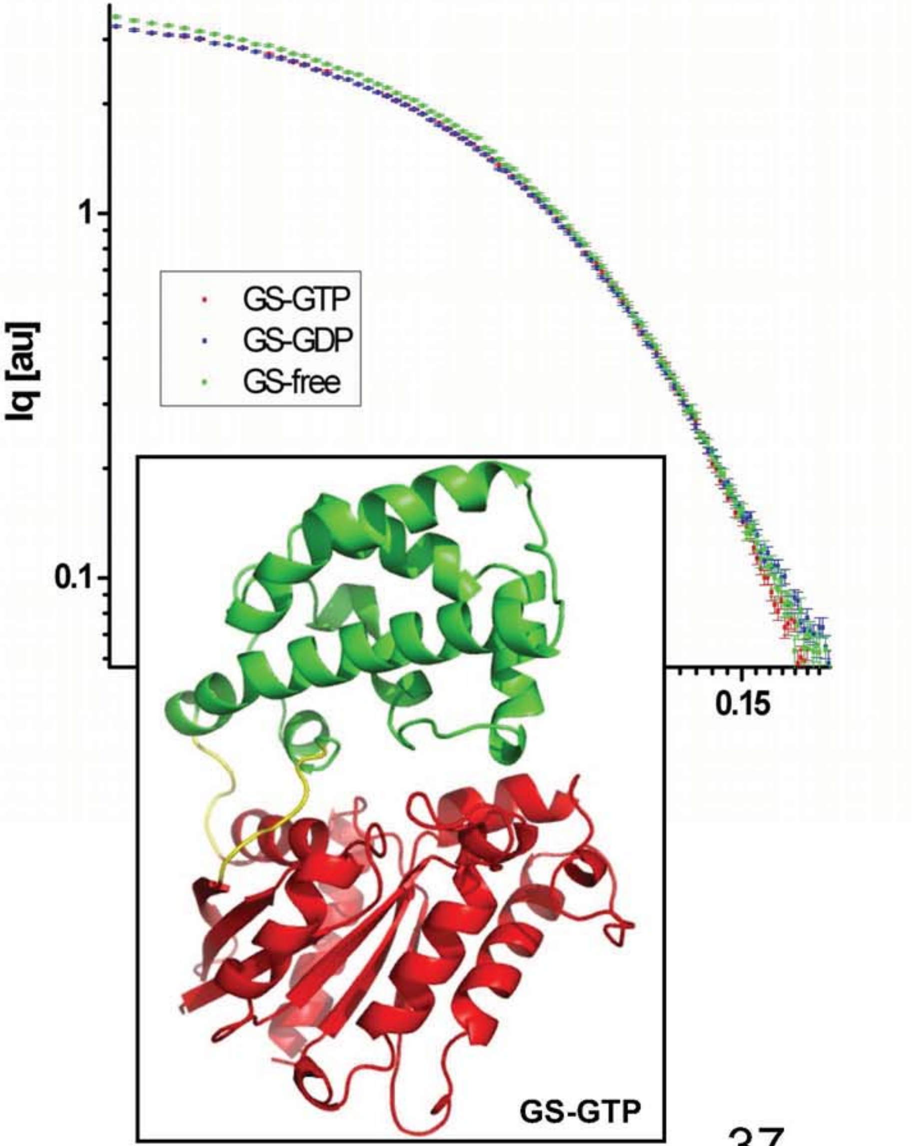




Rigid body modeling

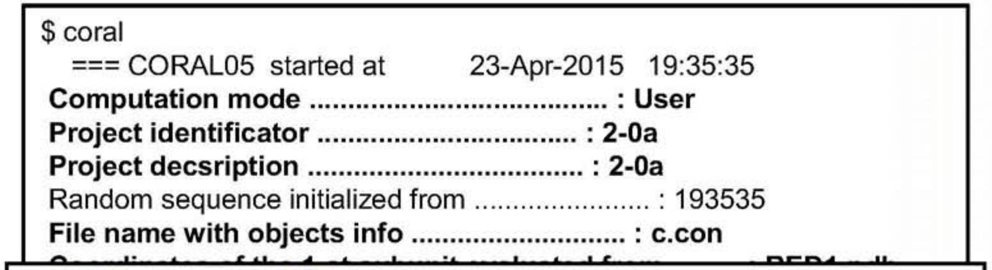
Fit of partial structures to SAXS data: coral

```
$ coral
   === CORAL05  started at         23-Apr-2015   19:35:35
 Computation mode ..................................... : User
 Project identificator ................................. : 2-0a
 Project decsription ................................... : 2-0a
 Random sequence initialized from ...................... : 193535
 File name with objects info ........................... : c.con
 Coordinates of the 1-st subunit evaluated from ......... : RED1.pdb
   198 atoms read, center at    26.89  -26.80    8.09
 Subunit was fixed
 Coordinates of the 2-nd subunit evaluated from ......... : GREEN.pdb
   911 atoms read, center at    52.99  -19.12   12.51
 Coordinates of the 3-rd subunit evaluated from ......... : RED2.pdb
  1386 atoms read, center at    31.86  -35.37    1.03
 Subunit was fixed
 Number of backbone atoms generated .................... : 316
Averaged formfactors of DRs used
 DR formfactor multiplier .............................. : 1.000
 Point symmetry of the particle ........................ : P1
 Number of equivalent positions ........................ : 1
 Cross penalty .......................................... : 0.3349
 Cross penalty weight .................................. : 100.0
 Shift penalty ......................................... : 2.376
 Shift penalty weight .................................. : 1.000
 Total number of scattering curves ..................... : 1
 1-st construct, the first and the last residues: 1   and   316
 File name, 1-st experimental data ..................... : GS-GTP-10mg.dat
 Number of experimental points found ................... : 481
 Experimental radius of gyration ....................... : 22.20
 .
 .
 .
```

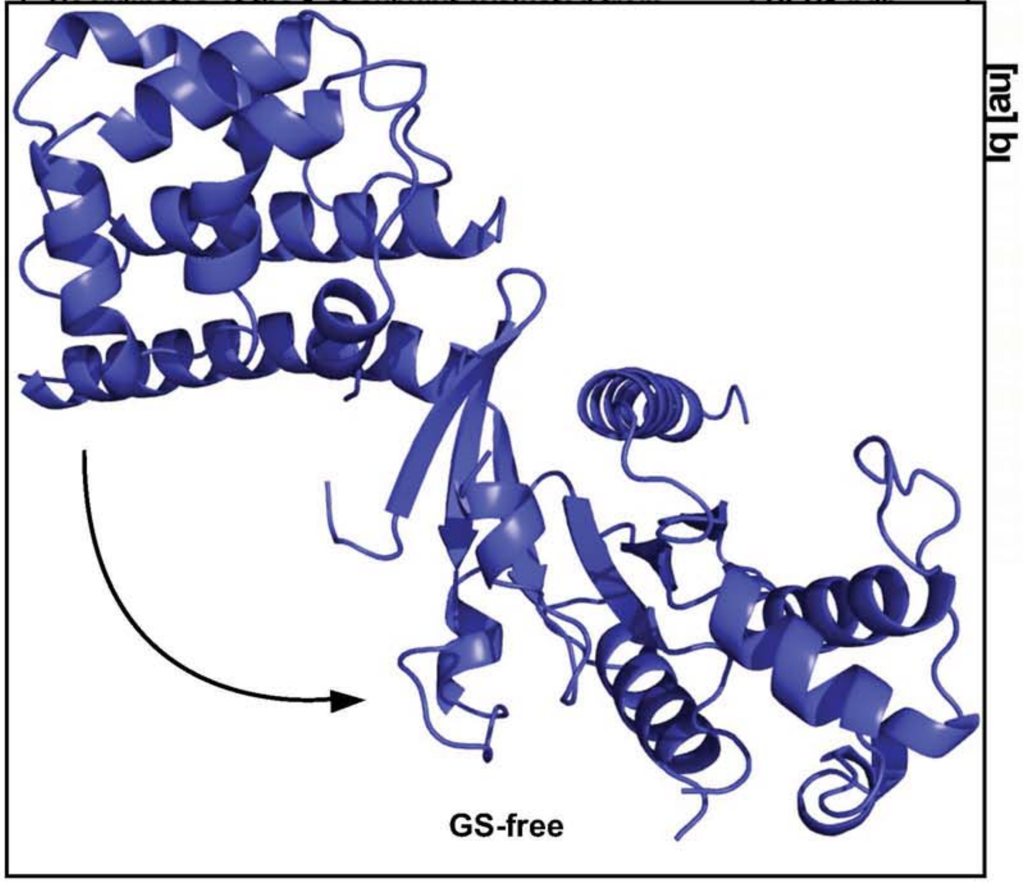


Rigid body modeling

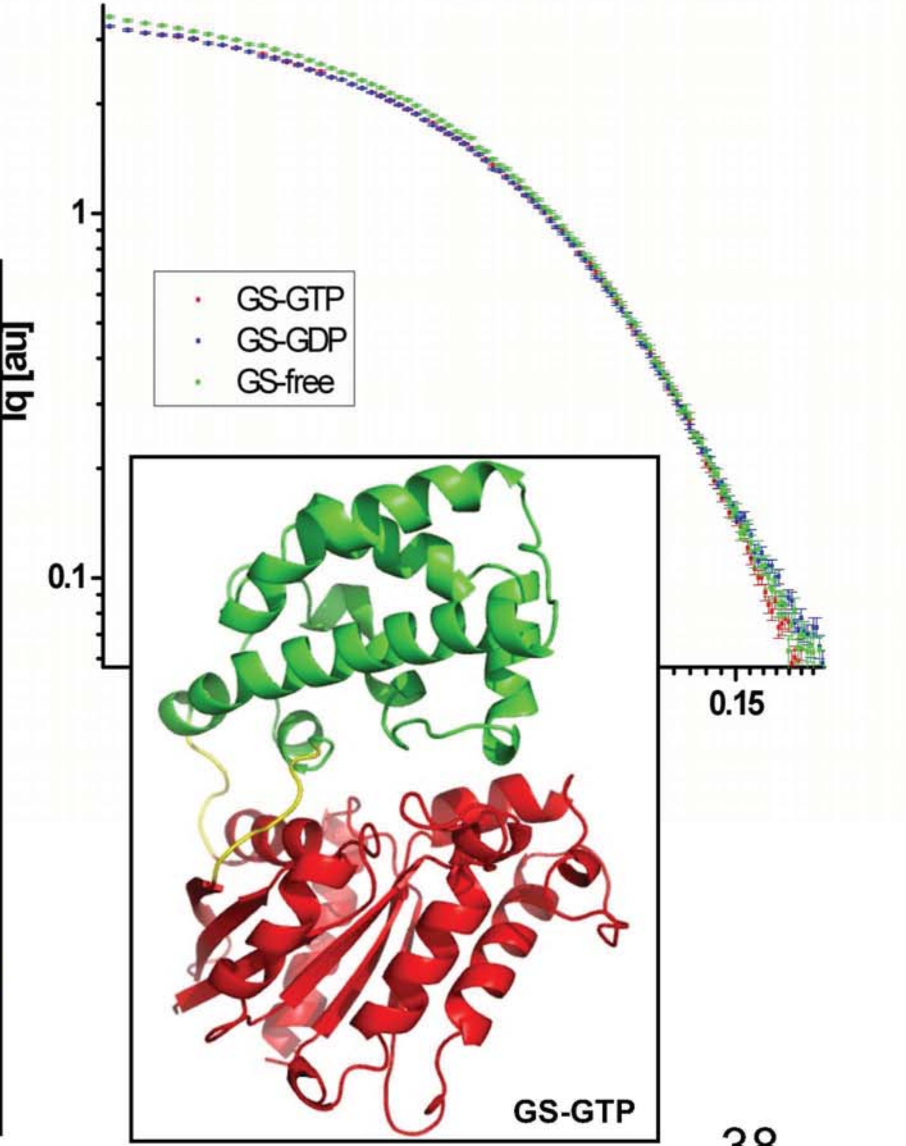
Fit of partial structures to SAXS data: coral



```
$ coral
   === CORAL05  started at        23-Apr-2015   19:35:35
 Computation mode ...................................... : User
 Project identificator .................................. : 2-0a
 Project decsription .................................... : 2-0a
 Random sequence initialized from ....................... : 193535
 File name with objects info ........................... : c.con
```



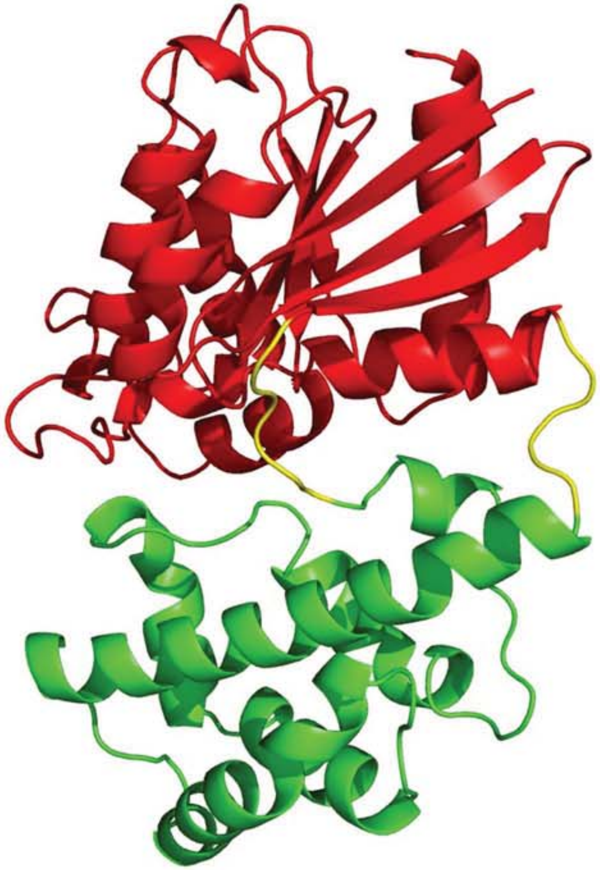
GS-free



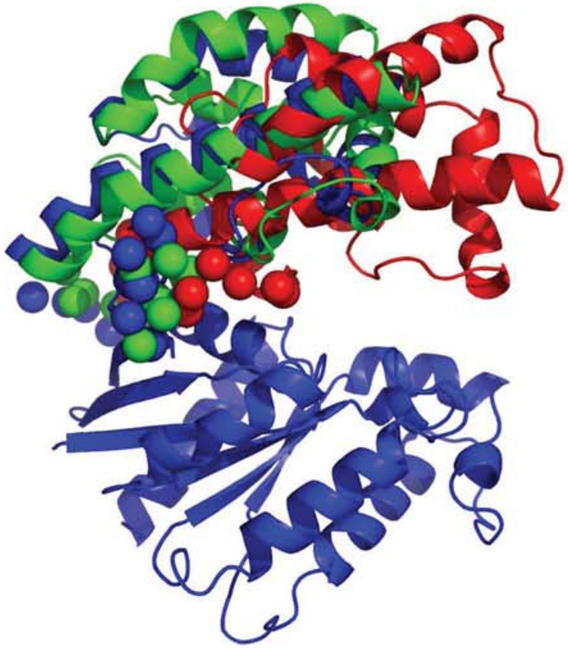
GS-GTP

Rigid body modeling

Fit of partial structures to SAXS data: coral

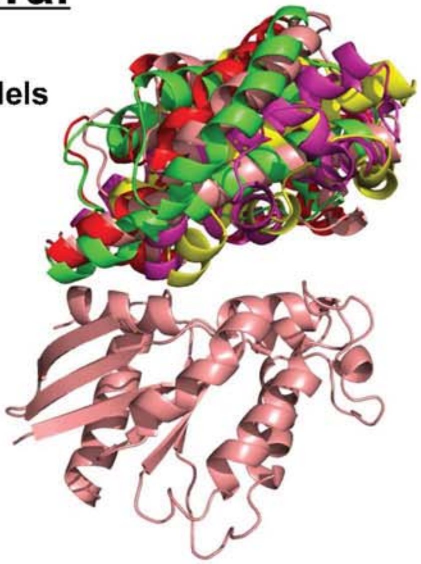


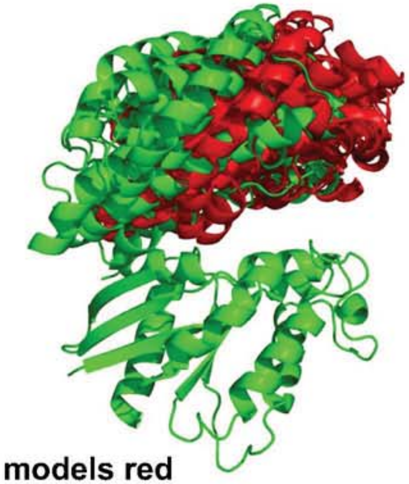
GS-GTP crystal



GS-GTP red
GS-GDP green
GS-free blue

GS-GTP 5 models





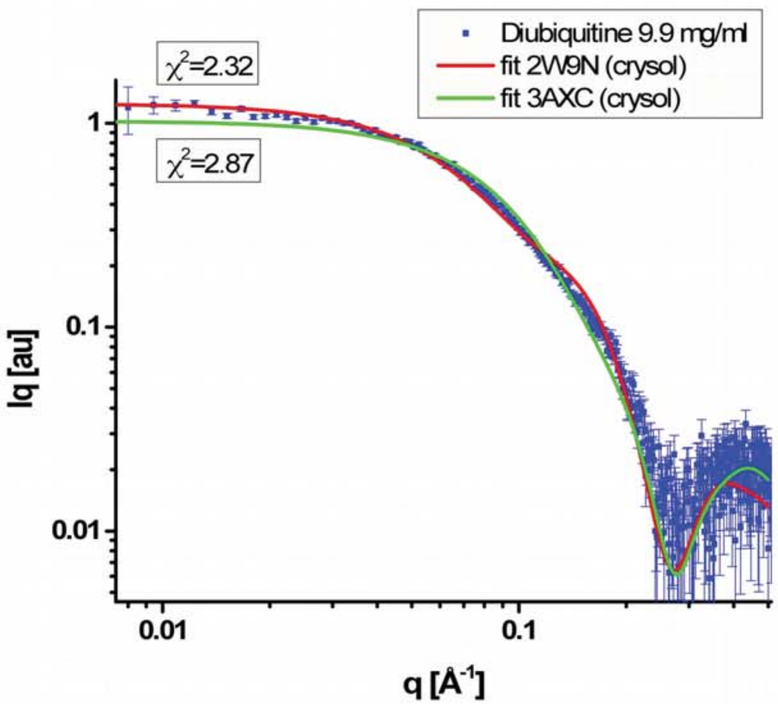
GS-GTP 5 models red
GS-GDP 5 models green

Rigid body modeling

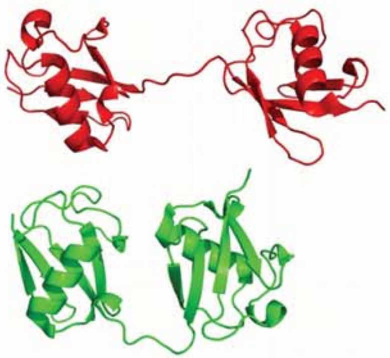
Fit of pieces of structures to SAXS data: eom

```
$ eom
   RANCH  Version 2.0
 Started: ............................................. : Fri Oct 17 16:06:44 2014
 iSeed ................................................ : 1934479814
 Chain type ........................................... : Random
 Sequence file name: .................................. : sequence.seq
 Symmetry: ............................................ : P1
 Symmetry type: ....................................... : aSymmetric
 Number of residues: .................................. : 1142
 Number of domains: ................................... : 2
  ------------------------------------------------------
 Domain number: ....................................... : 1
 Path: ................................................ : diubi1.pdb
 Fixed: ............................................... : 0
 Oligomer: ............................................ : 0
 DNA file name: ....................................... : None
  ------------------------------------------------------
 Domain number: ....................................... : 2
 Path: ................................................ : diubi2.pdb
 Fixed: ............................................... : 0
 Oligomer: ............................................ : 0
 DNA file name: ....................................... : None
 Contact file name: ................................... : None
 Number of structures: ................................ : 10000
 Suffix of the pdb files: ............................. : eom
 Output Folder Path: .................................. : ./
 Number of experimental curves: ....................... : 1
 Experimental data filename (curve 1): ................. : diubi_9p9_subtr.dat
.
```

Sequence file has fasta format (without comments and chain information)

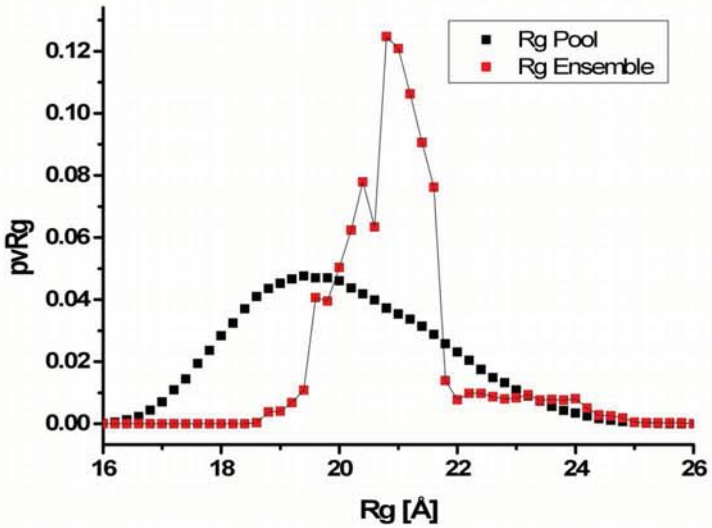


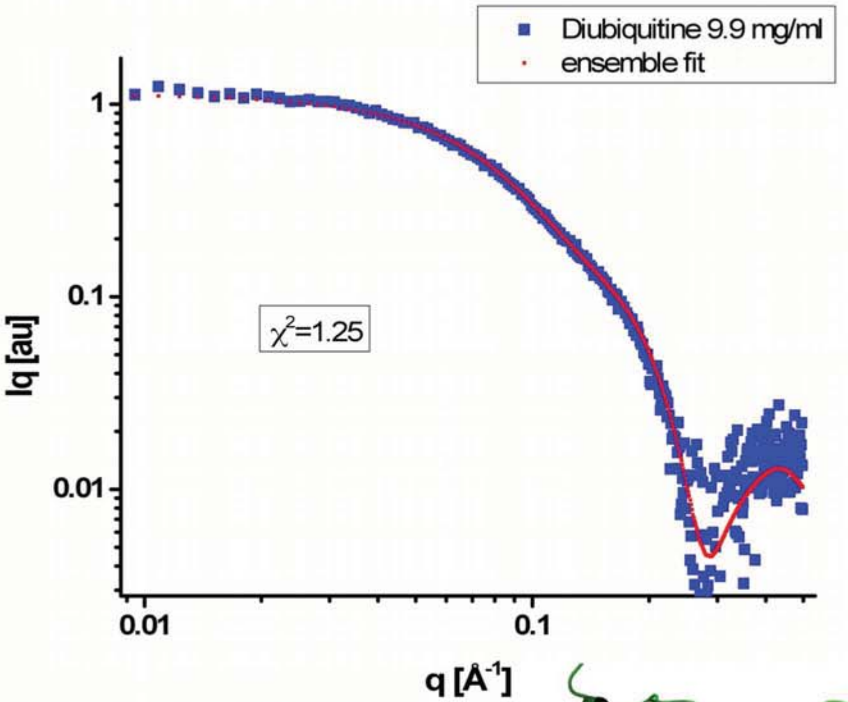
Crystal structures
2W9N (red)
3AXC (green)



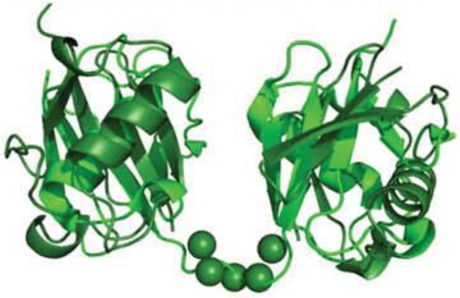
Rigid body modeling

Fit of pieces of structures to SAXS data: eom



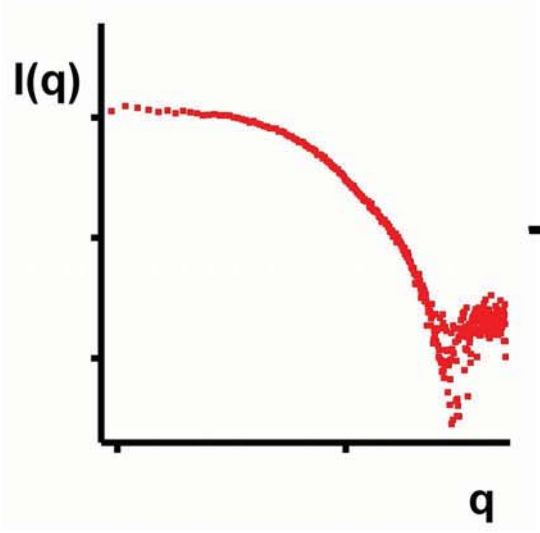


Overlay of Enseble Models and Crystal Structures



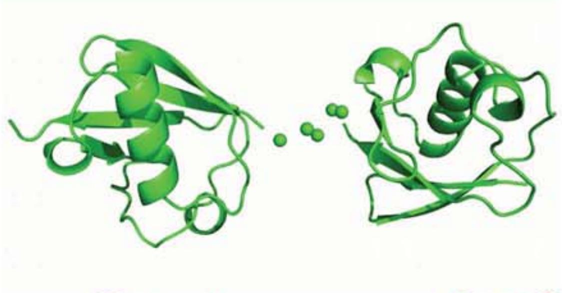
Distance	**15 [Å]**	**9 [Å]**

Rigid body modeling

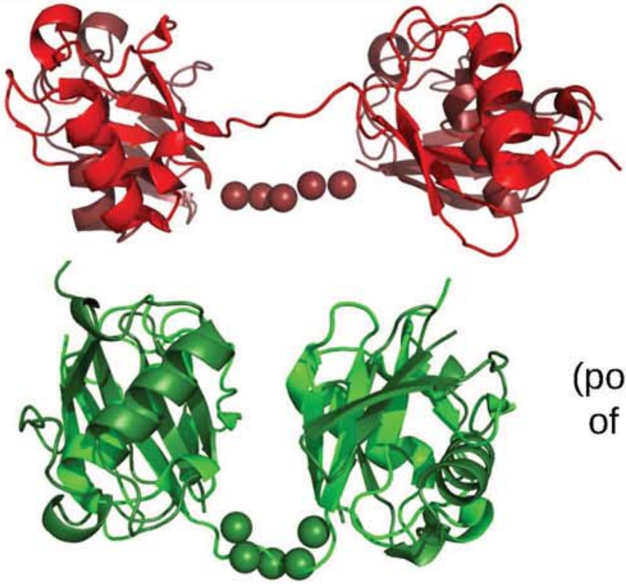


Depending on the initial assumption or question modeling of data with different software leads to different results.

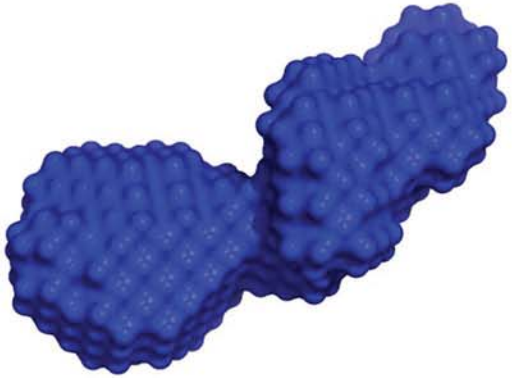
All these results are correct within the limiting factors of the software



Rigid body model
(average structure)



Ensemble
Ratio
38:62
(possible conformations
of a flexible molecule)



Dummy model
(bead model without
any structural
information)

Thanks