DANSE/SANS Group
University of Tennessee

The Applicability of a P(r) Inversion Method

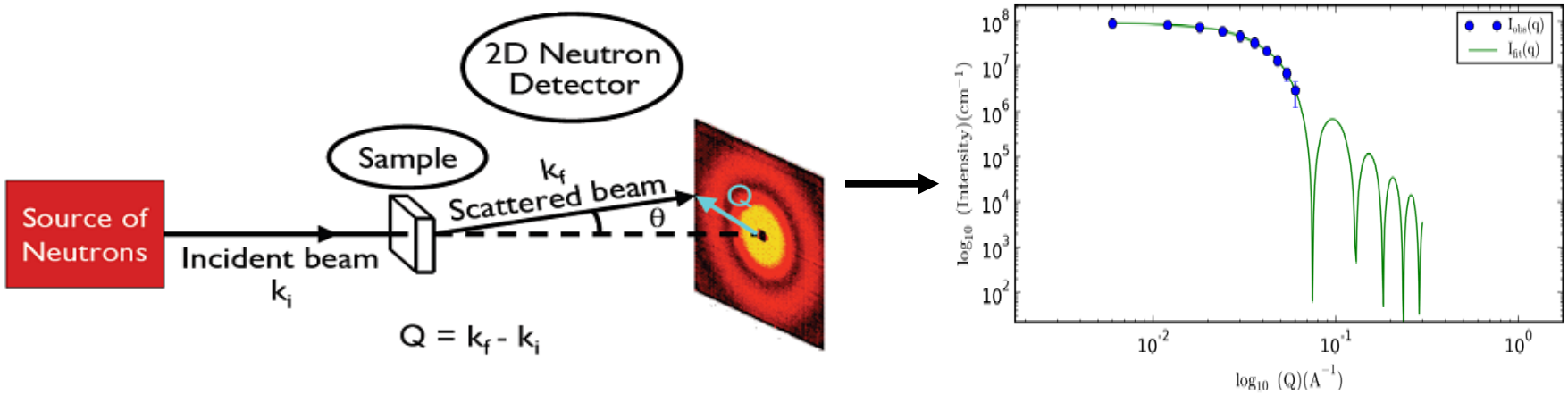
Raiza Cortes Hernandez
University of Puerto Rico, Mayaguez Campus
Advisor: Mathieu Doucet (Paul Butler)
National Institute of Standard and Technology (NIST)
NIST Center for Neutron Research Laboratory (NCNR)

Outline

- Small Angle Neutron Scattering
- The P(r) Technique
- Parameters
- Outputs
- PrView 0.2

Small Angle Neutron Scattering (SANS)

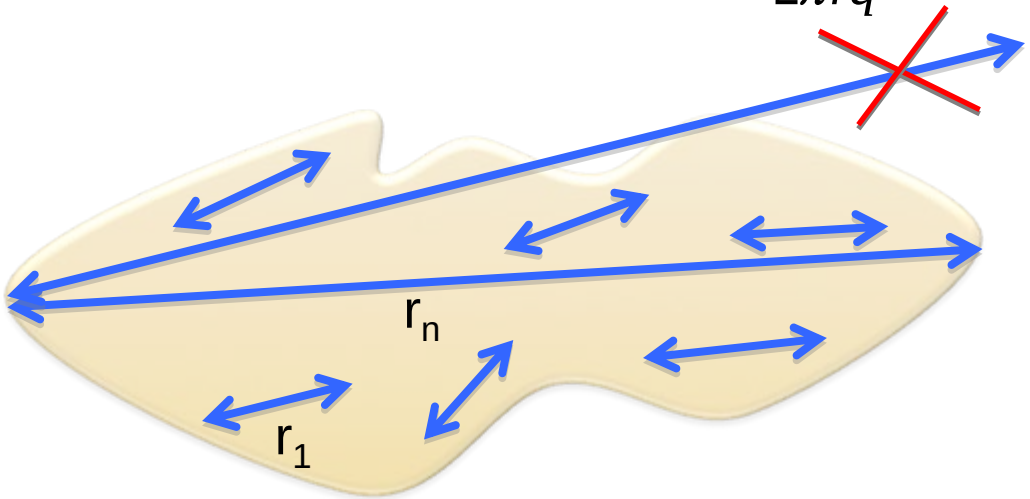
- An incident beam of neutrons is directed onto a sample. Most is transmitted, some is absorbed, and some is scattered.
- A detector is positioned at some distance from the sample, and the scattered intensity is recorded as a function of angle (or Q, the momentum transfer).
- Scattering of neutron measures the structure, shape and size of the sample.

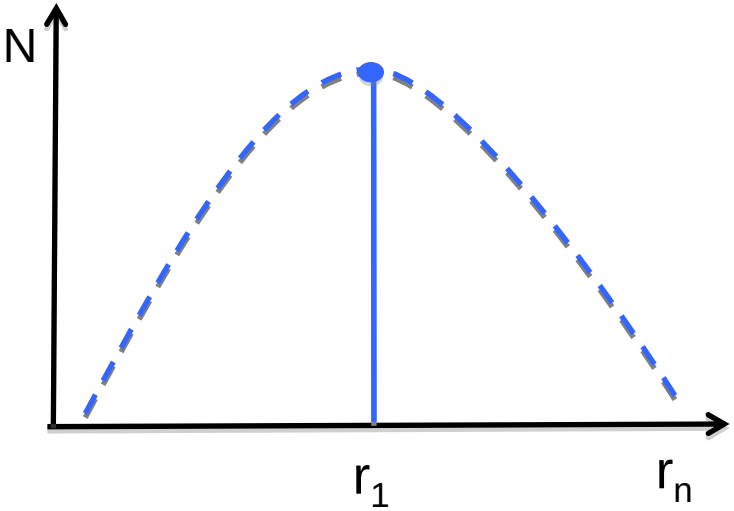


It measures the structure from 10 [Å] to 3000 [Å]

Getting P(r) from I(Q)

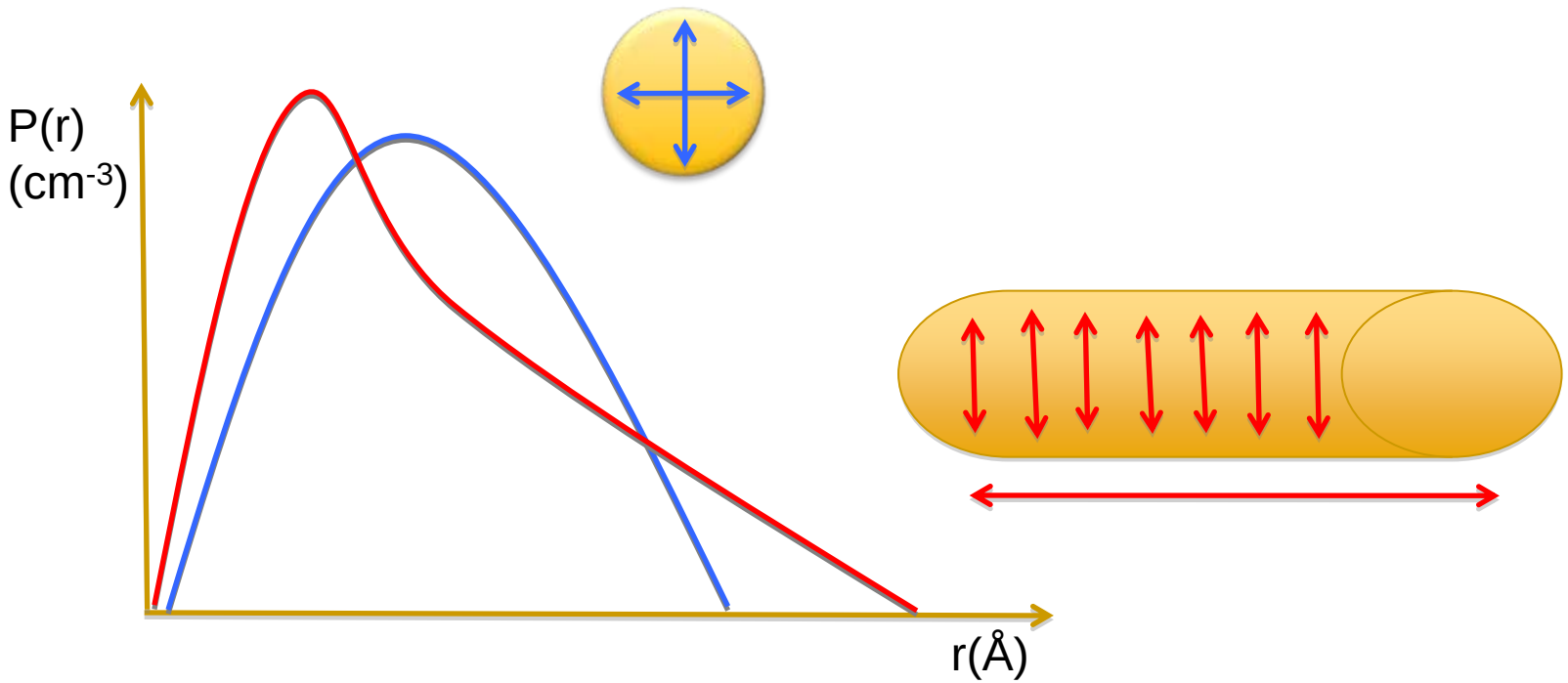
- Simple, well understood shapes can be modeled with analytical expressions to obtain some size parameters.
- For more complex cases, or cases where the experimenter has no idea other techniques must be used,
  - Consider: $$I(q) = \int p(r) \frac{\sin 2\pi rq}{2\pi rq} dr$$





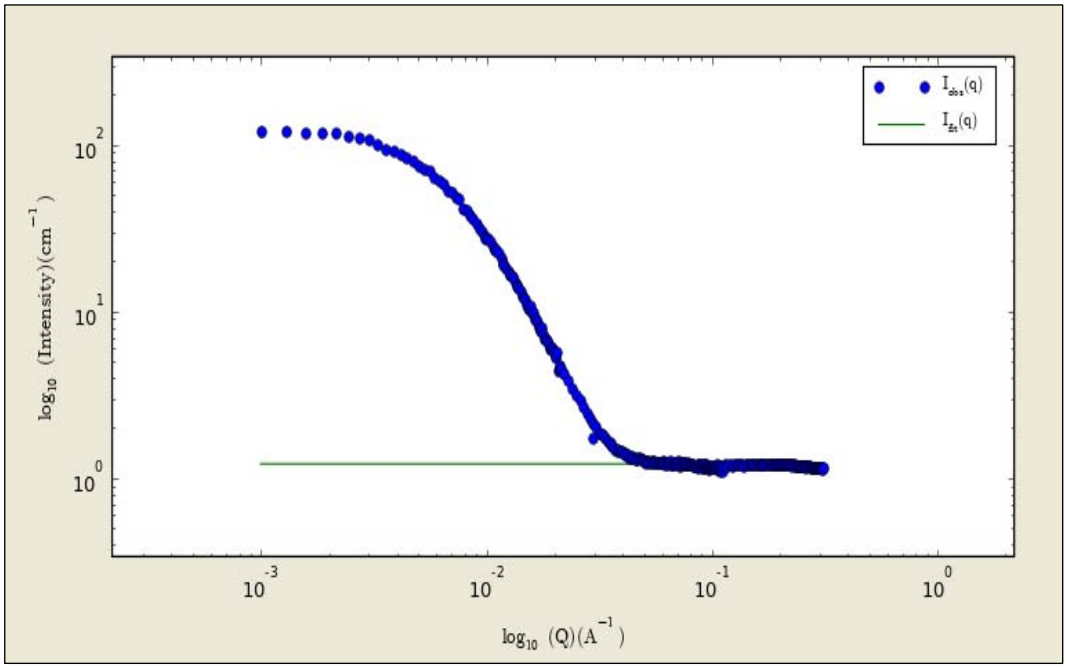
Getting P(r) from I(Q)

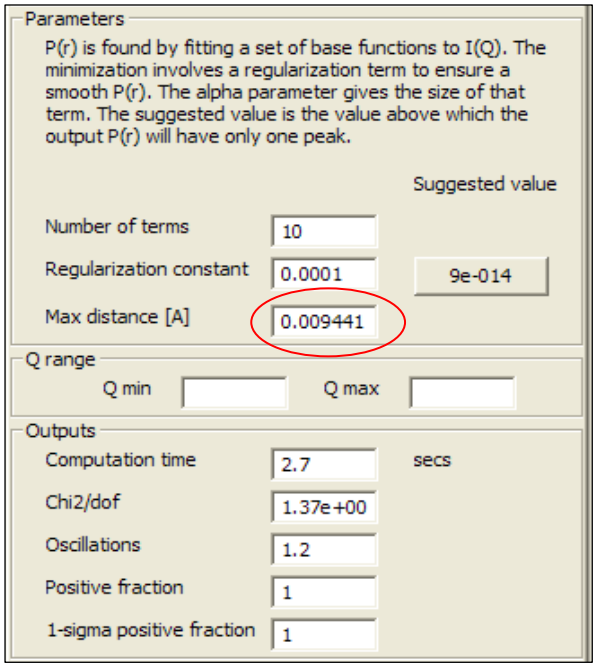
- How it's translated into a shape.



First Attempt with real data

- The program used to calculate P(r) is PrView 0.1*
- We have an idea of what the shape might be …. but the answer is nonsense…. Despite "playing" for a few days.





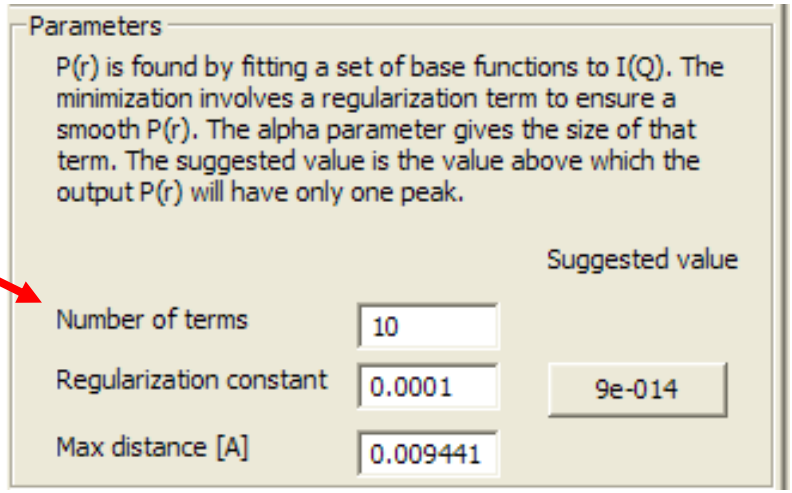
*http://danse.chem.utk.edu/prview.html

Number of Terms Parameter

- It was possible to generate an algorithm to determine how many terms are necessary in the P(r) expansion to obtain a reliable result.

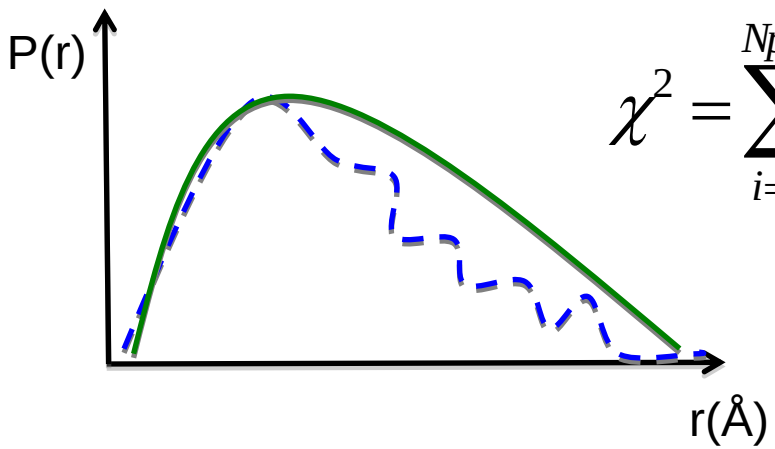
$$p(r) = \sum_{i=1}^{N} c_i \varphi_i(r)$$

Terms needed in the expansion (*Number of terms* parameter in PrView).



Regularization Term Parameter

- We chose to modify a simple algorithm due to P. Moore*.
- A *regularization term* was added to help converge faster to a physically reasonable result.



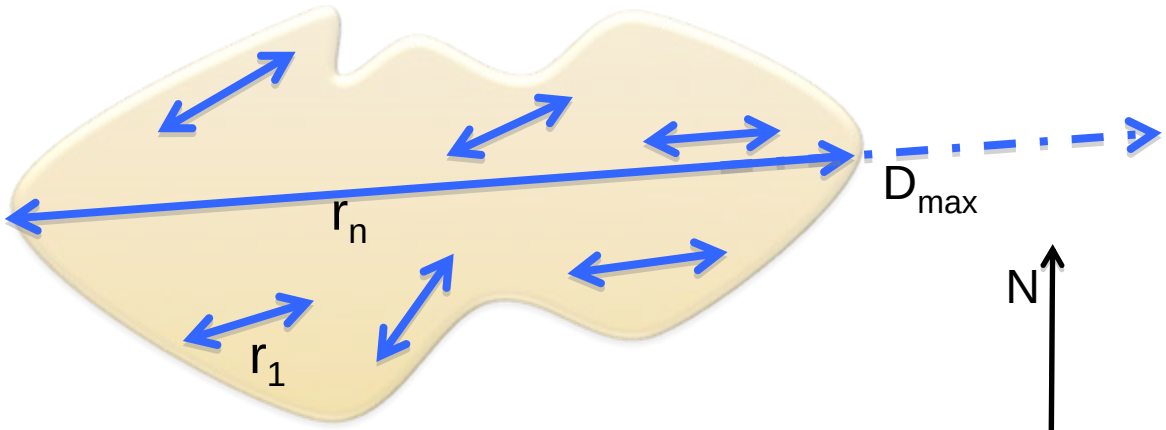
$$\chi^2 = \sum_{i=1}^{Npts} \left(I_i^{obs} - I(q_i) \right)^2 / \sigma_i^2 + \alpha \int \frac{d^2 p(r)}{dr^2} dr$$

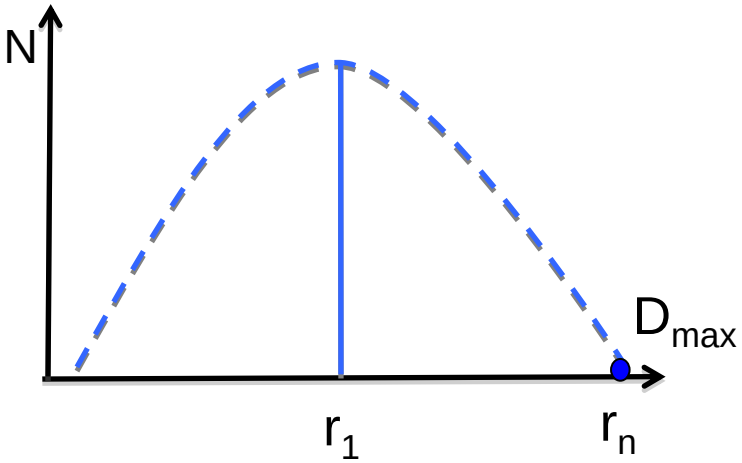
- The minimization is done numerically with a simple linear least squares fit and favors “smooth” results over highly oscillatory ones.

*P. Moore, J. Appl. Cryst. (1980) 13, 168-175.

D_{max} Parameter

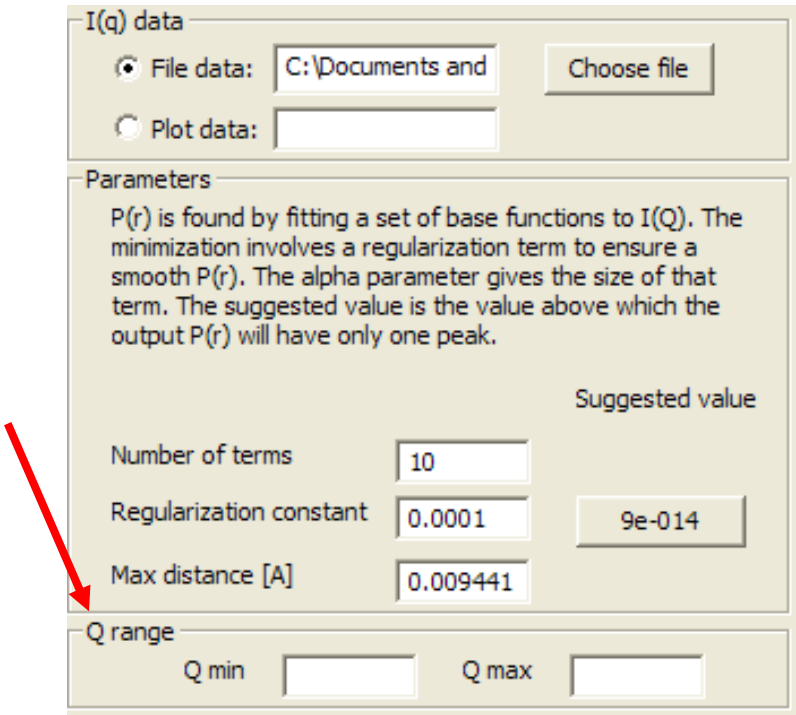
- Maximum Distance Parameter in PrView.
- Is the longest distance that appears inside the shape.

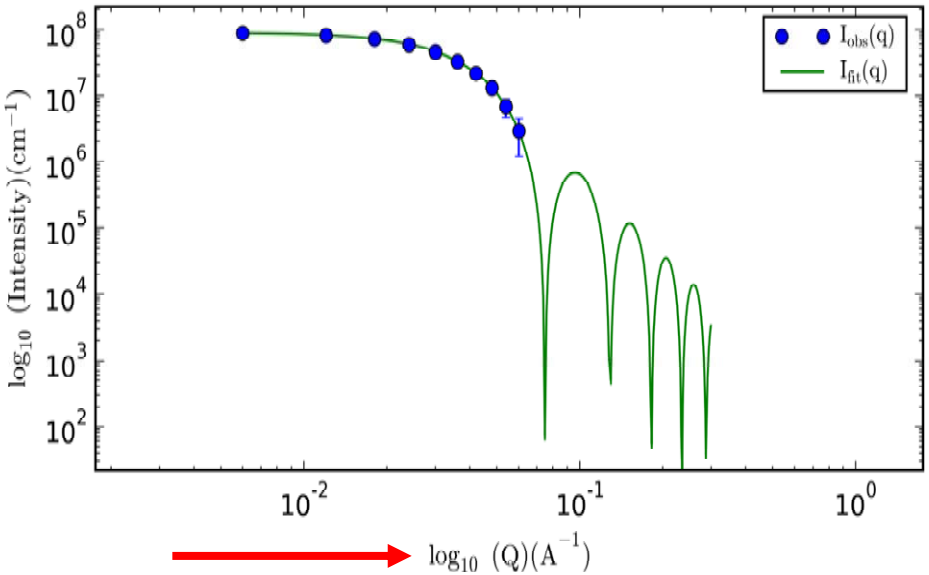




Q Range Choice

- Is the x axis values in the I(Q) graph.

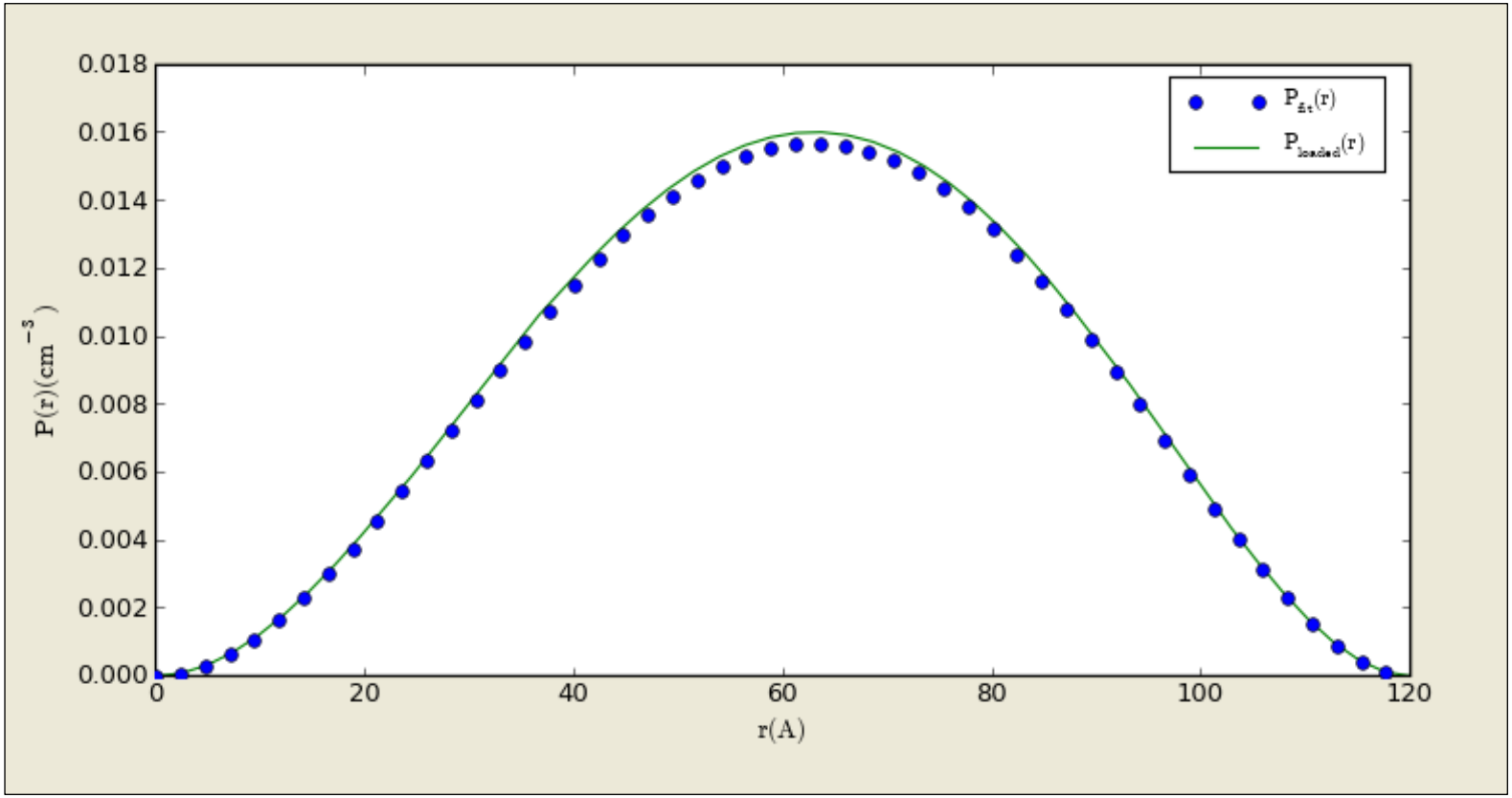




What can we do?

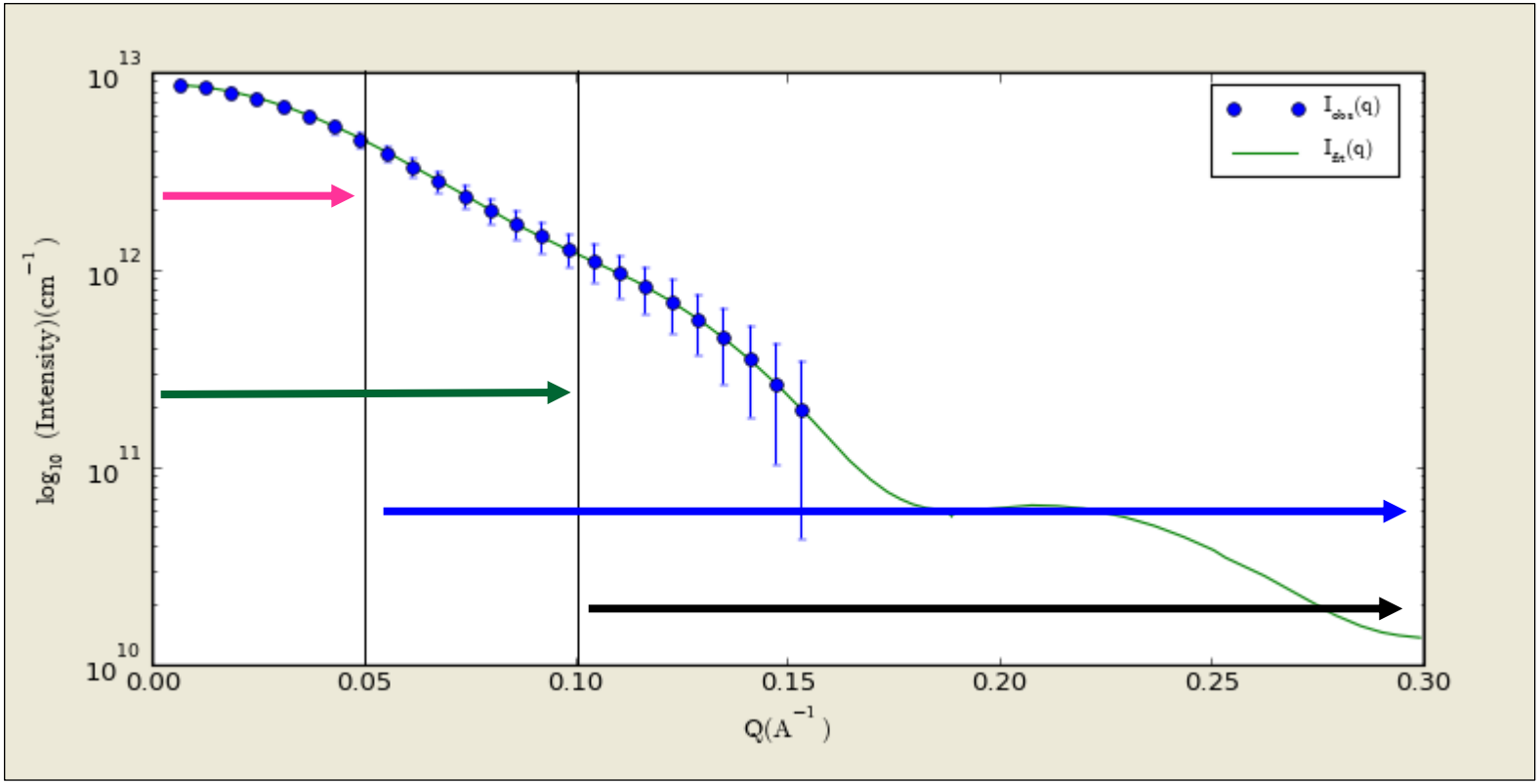
- Start with the simplest things (one step at a time).
- List of geometric shapes used for validation of the technique and our regularization term:
  - Spheres, radius range of $5 \leq R \leq 400$.
  - Cylinders
  - Simulated shapes with a more complicated I(q) distribution like:
    - "Unknown" arrangement of spheres and cylinders.
    - Dumbbells (to thoroughly challenge the regularization term)

Simulated Data for a Sphere



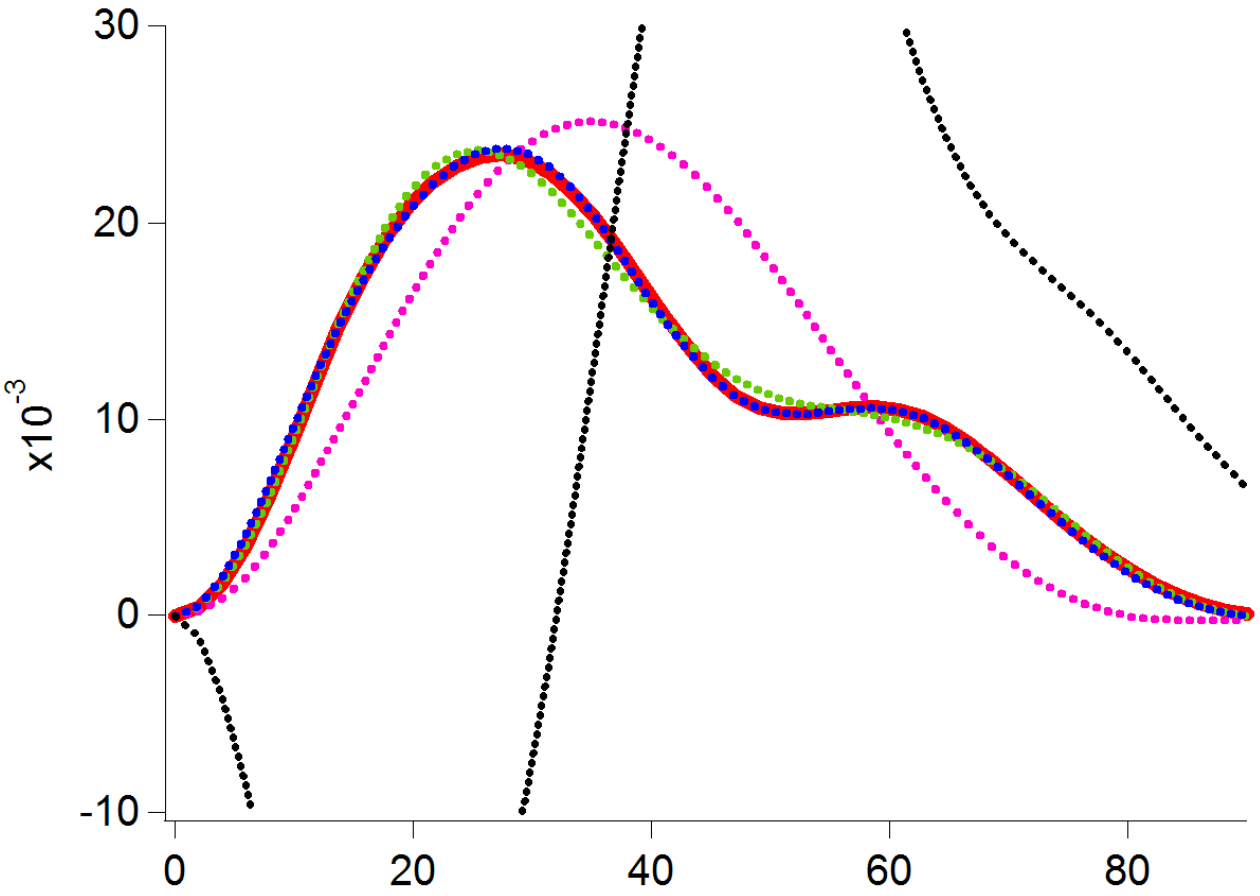
Where Should I Start?

- Explore effect of restricted Q range
- Effect of PrView parameters



Effect of Q range

- A restricted Q range doesn't give reliable results.



Red – Answer

Pink – 0.0061 – 0.05

Green – 0.0061-0.1

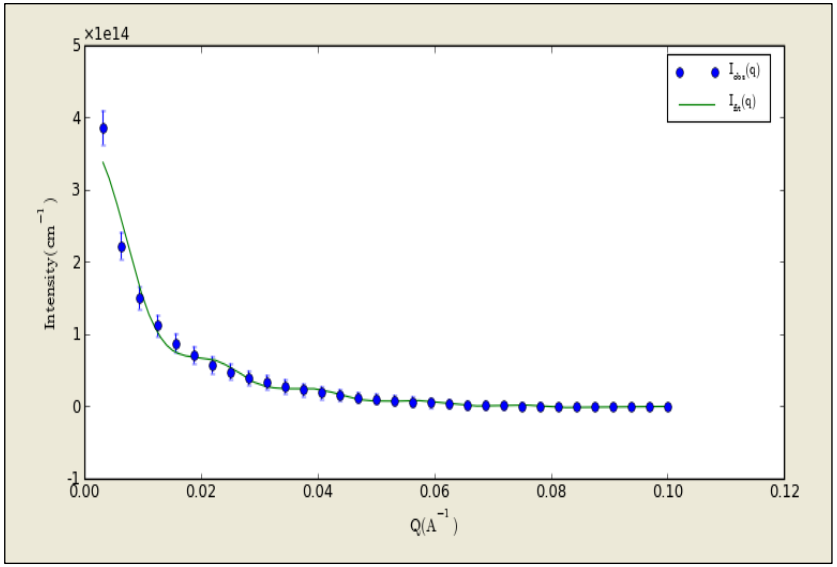
Blue – 0.05-0.3

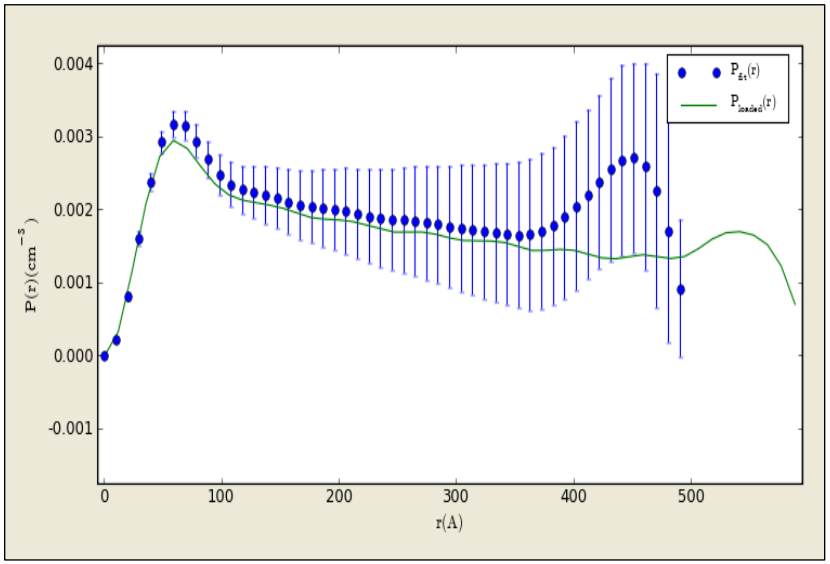
Black – 0.1 – 0.3

Full Q range is 0.0061-0.3

Effect of D_{max} parameter

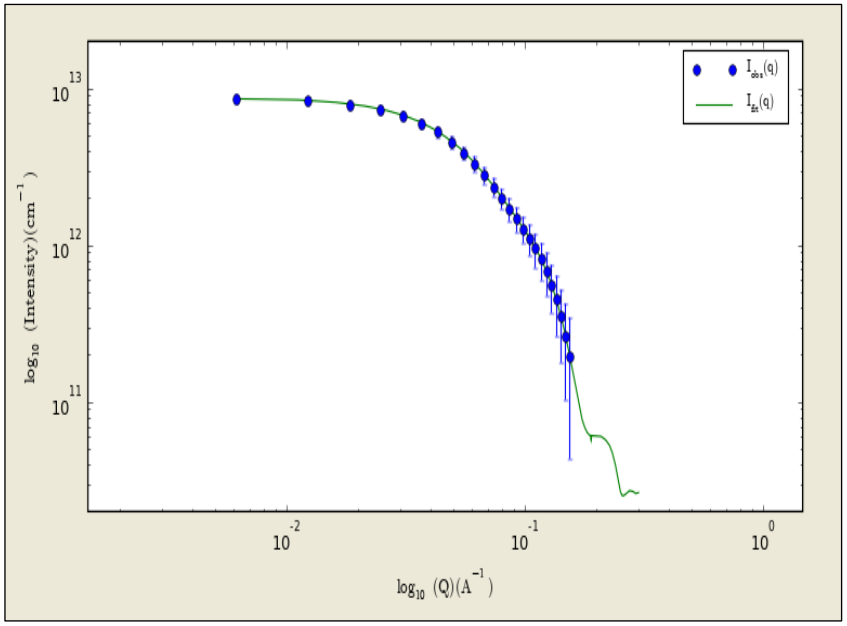
- What happens if the D_{max} is less than the best answer.
  - Cylinder with D_{max} = 500 [Å]
  - Answer D_{max} of 1000 [Å]

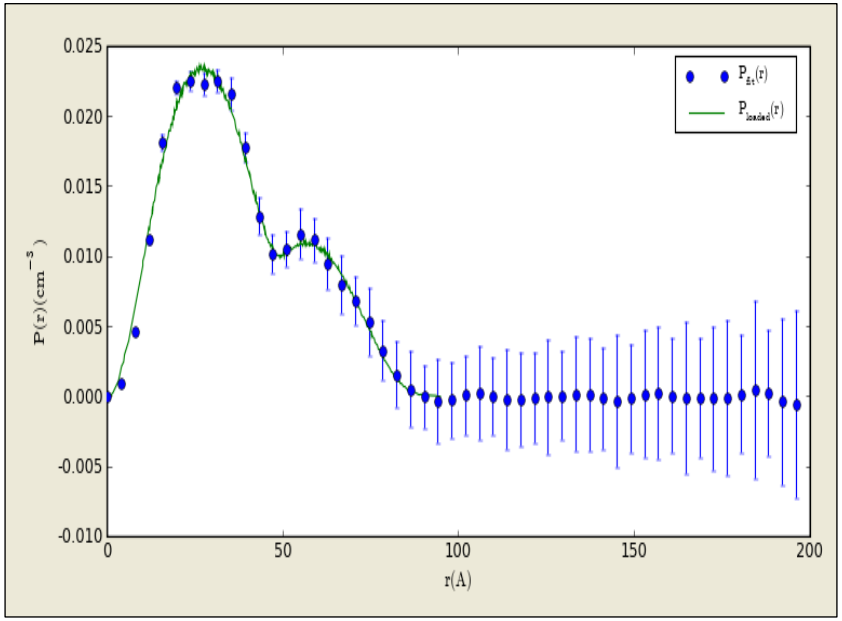




Effect of D_{max} parameter (+ Regularization term)

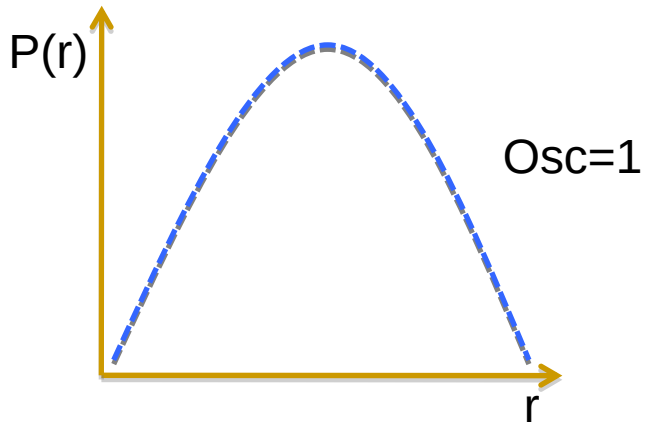
- What happens if D_{max} is greater than the best answer:
  - Dumbbells with D_{max} = 200 [Å]
  - Answer D_{max} = 100 [Å]

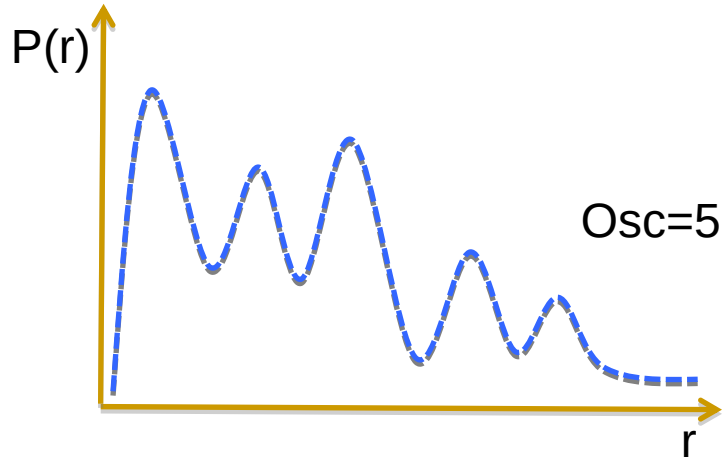




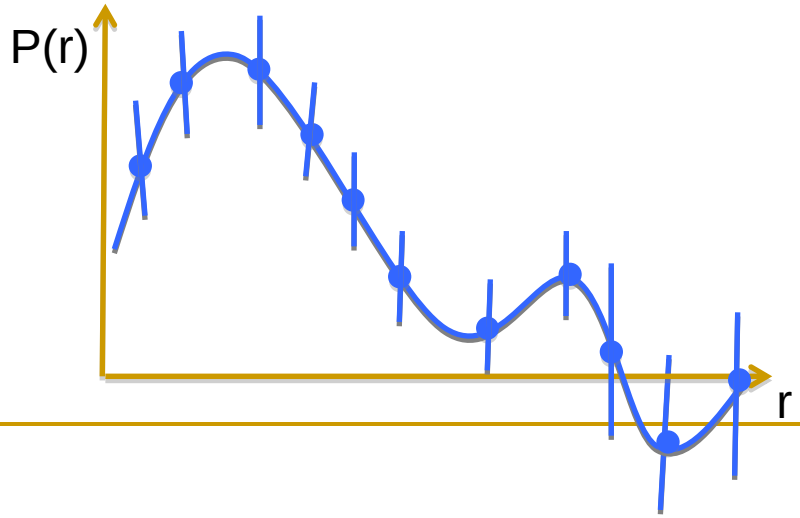
Number of Terms Parameter

- Figures of Merit (Outputs)
- Oscillations



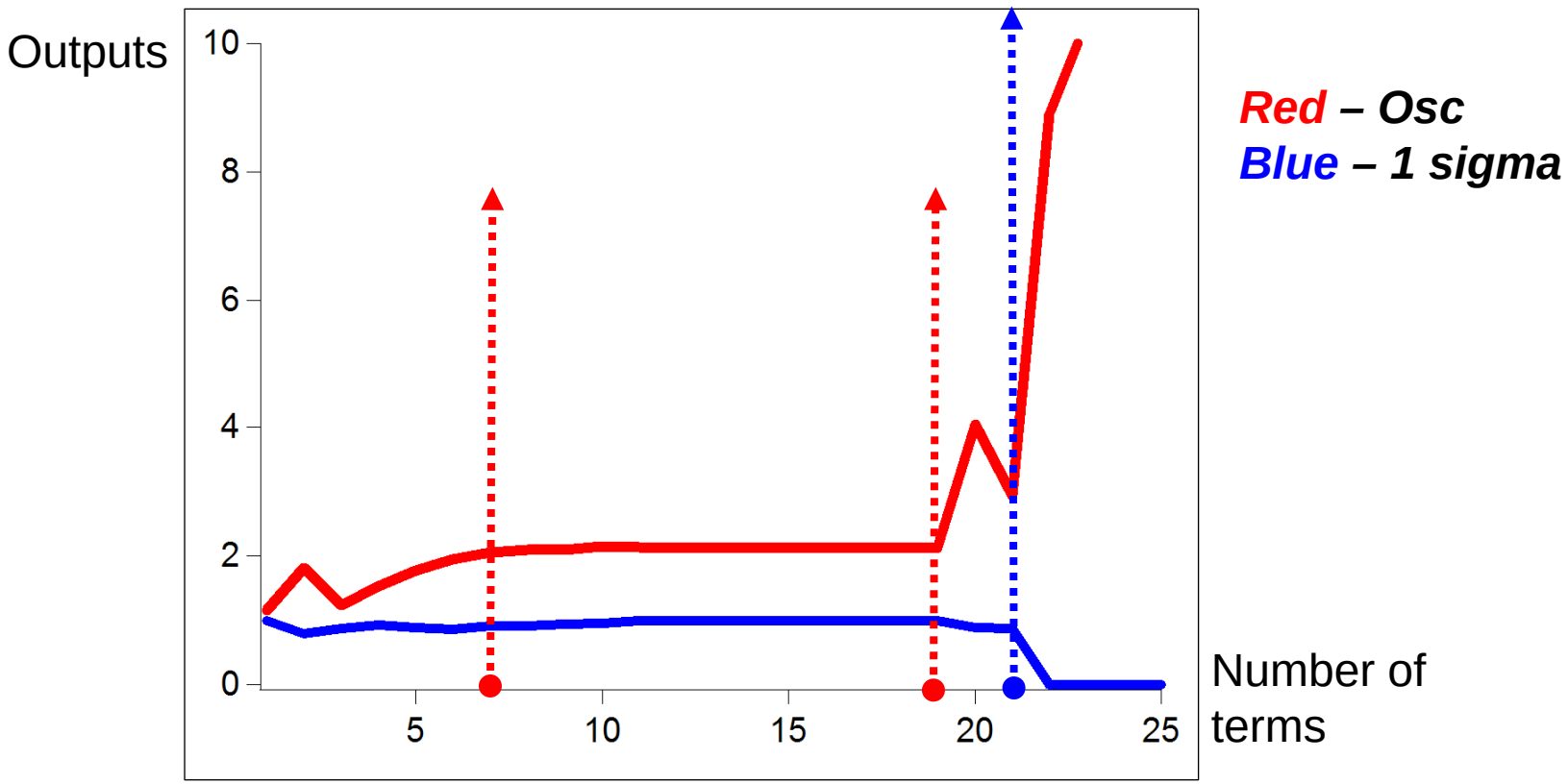


- 1-sigma



Automating the Number of Terms Parameters

- It was found a similarity in each case.



New Results of the First Attempt

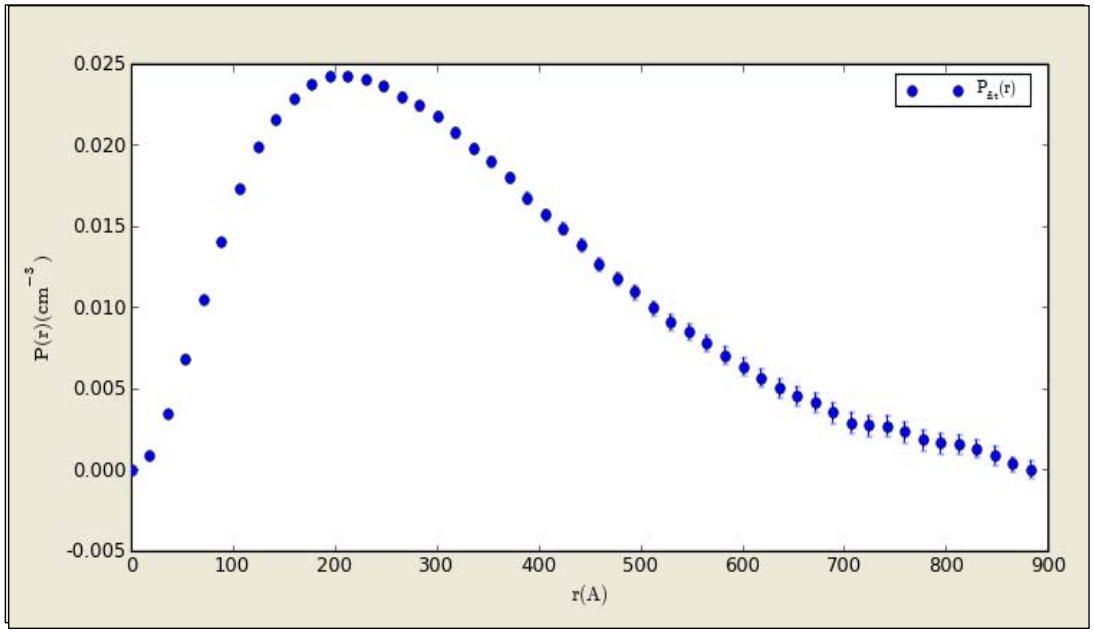
Parameters

P(r) is found by fitting a set of base functions to I(Q). The minimization involves a regularization term to ensure a smooth P(r). The alpha parameter gives the size of that term. The suggested value is the value above which the output P(r) will have only one peak.

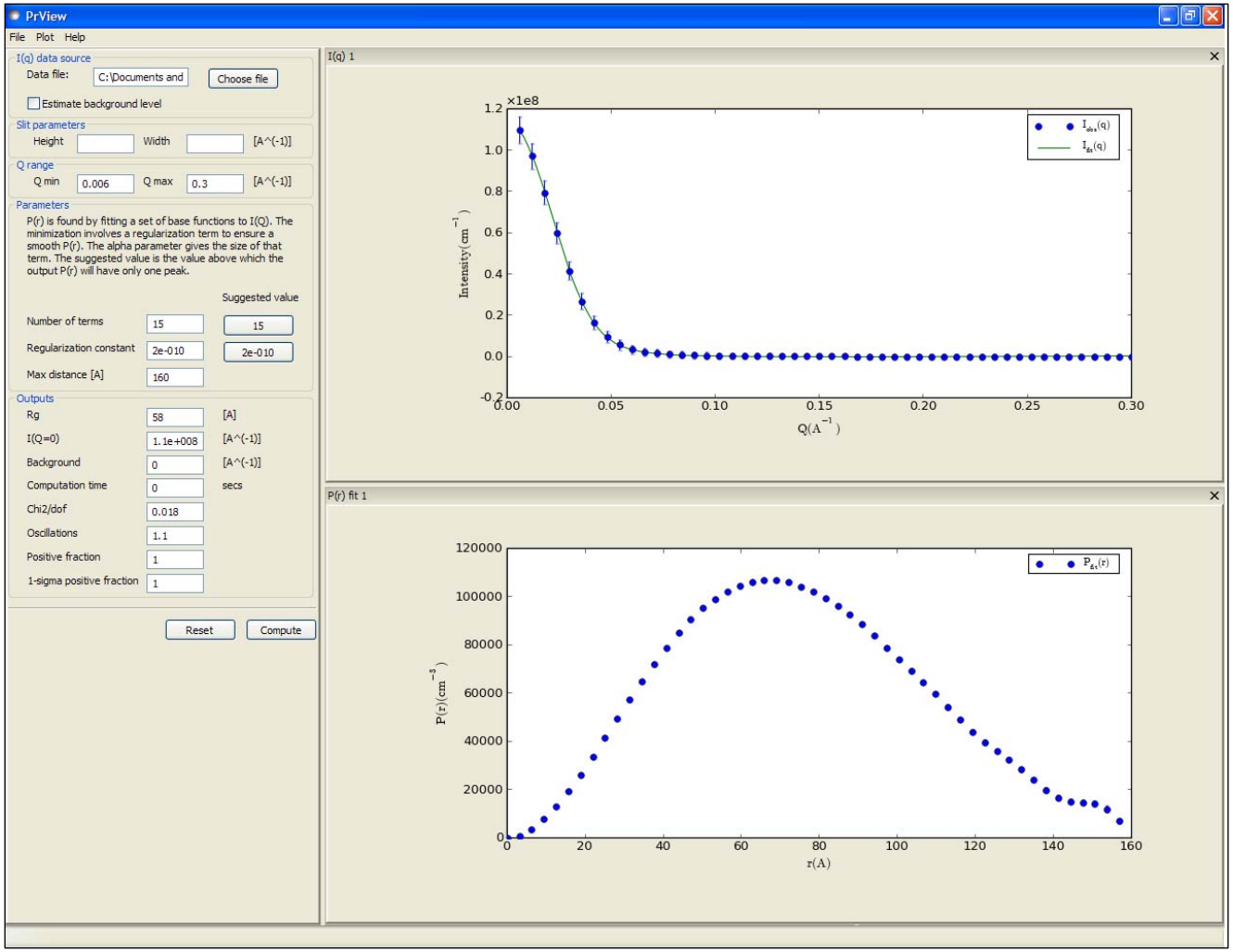
		Suggested value
Number of terms	29	29
Regularization constant	8e+006	8e+006
Max distance [A]	900	

Outputs

Rg	2.6e+002	[A]
I(Q=0)	1.3e+002	[A^(-1)]
Background	1.2	[A^(-1)]
Computation time	0.016	secs
Chi2/dof	1.3e+003	
Oscillations	1.5	
Positive fraction	1	
1-sigma positive fraction	1	



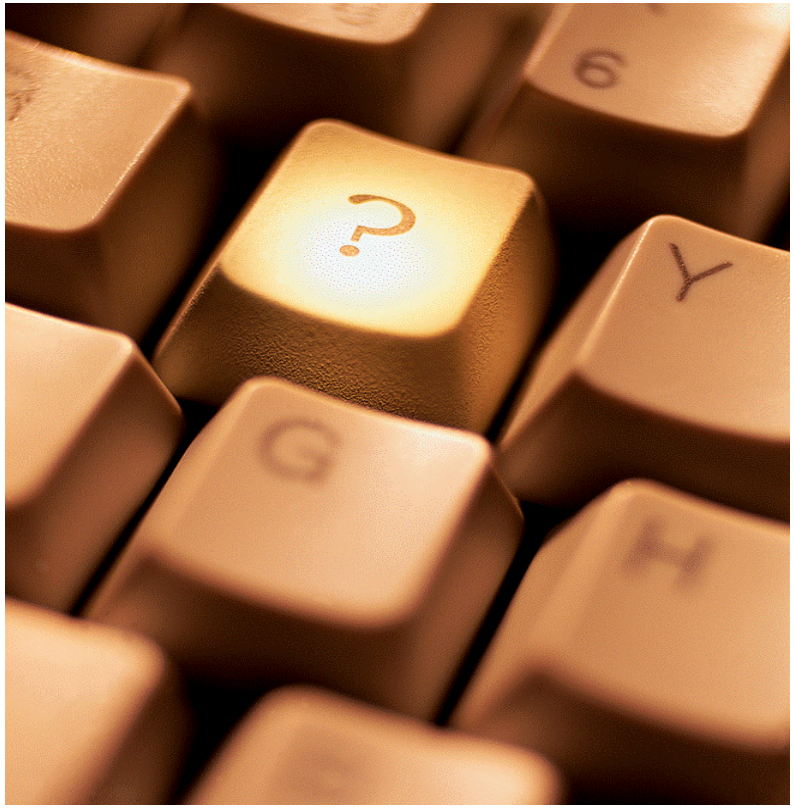
PrView 0.2


PrView
File Plot Help
I(q) data source
Data file: C:\Documents and
Choose file
Estimate background level
Slit parameters
Height
Width
[A^(-1)]
Q range
Q min 0.006
Q max 0.3
[A^(-1)]
Parameters
P(r) is found by fitting a set of base functions to I(Q). The minimization involves a regularization term to ensure a smooth P(r). The alpha parameter gives the size of that term. The suggested value is the value above which the output P(r) will have only one peak.
Suggested value
Number of terms 15 15
Regularization constant 2e-010 2e-010
Max distance [A] 160
Outputs
Rg 58 [A]
I(Q=0) 1.1e+008 [A^(-1)]
Background 0 [A^(-1)]
Computation time 0 secs
Chi2/dof 0.018
Oscillations 1.1
Positive fraction 1
1-sigma positive fraction 1
Reset
Compute
I(q) 1
×1e8
Intensity(cm^-1)
Q(A^-1)
P(r) fit 1
P(r)(cm^-3)
r(A)

Conclusion

- New Version of PrView (0.2) with an automated number of terms parameter.
- The users know that a restricted Q range doesn't give reliable result and a way of getting an approximation of the D_{max} input parameter.
- Easier for the users to get an answer or an approximation of the answer.

Questions?



- Deleted slides

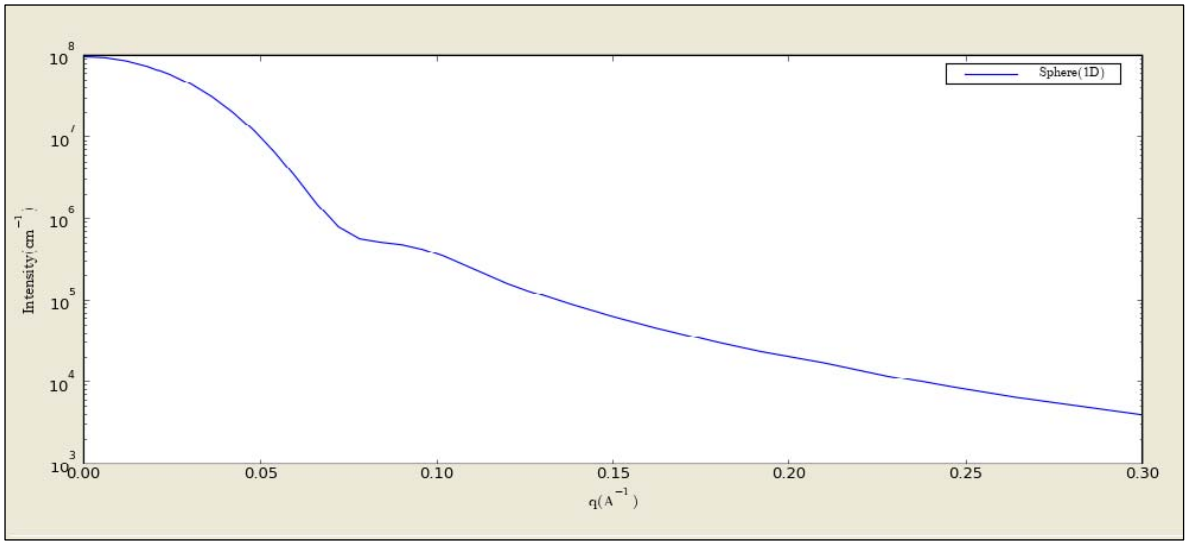
Technique for getting P(r)

- The coefficient of each base function is found by minimizing the following:

$$\chi^2 = \sum_{i=1}^{Npts} \left(I_i^{obs} - I(q_i) \right)^2 / \sigma_i^2 + \alpha \int \frac{d^2 p(r)}{dr^2} dr$$

Evaluating P(r) for simulated I(q) Data

- Simulated data for a Sphere was generated with different radius, polydispersity and Q range. To determine the types of system that the technique is good for, how sensitive it is to the length and to predict the size of the regularization term.



<u>Example:</u>

Radius 60

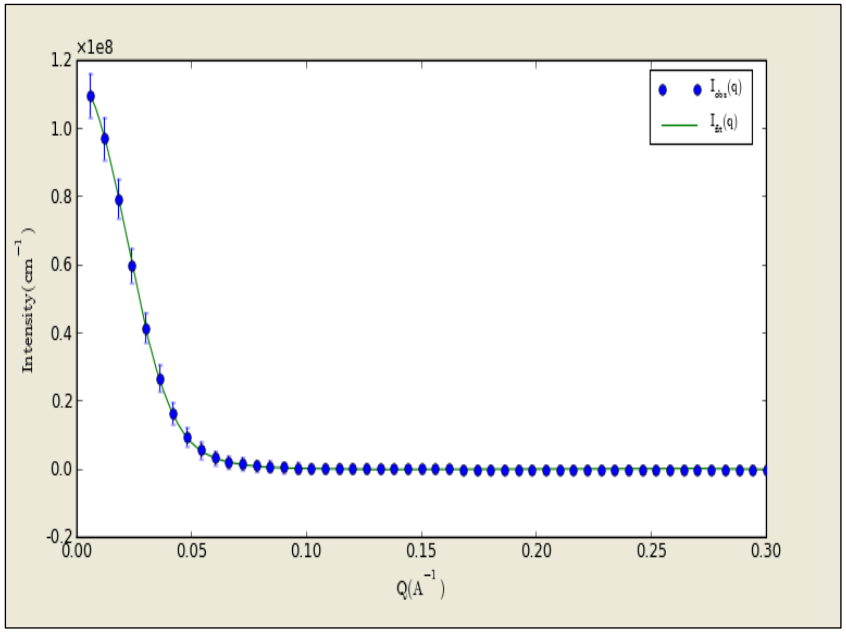
Disp 10

Evaluating P(r) for simulated I(q) Data

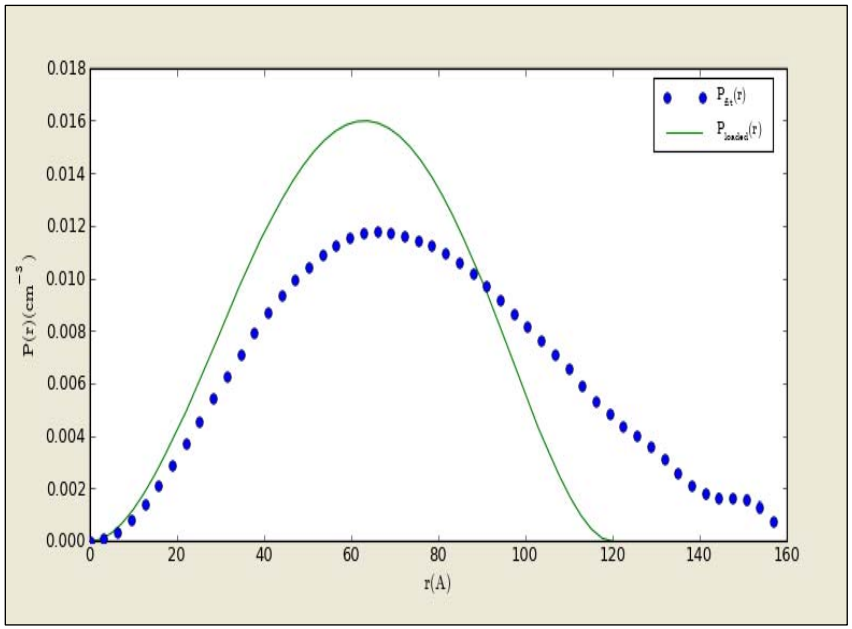
- Also the answer for the simulated data was generated, which was used for comparing.
- To know the types of system that the technique works for, the simulated data was loaded into PrView, we observe how the system change with different distances and Q ranges and compare it with the answer.
- Every case was divided in different Q ranges:
  - 0.001-0.3 (A^{-1}) (full range for Sphere)
  - 0.001-0.02 (A^{-1})
  - 0.001-0.05 (A^{-1})
  - 0.001-0.08 (A^{-1})
  - 0.02-0.08 (A^{-1})
  - 0.25-0.08 (A^{-1})
  - 0.05-0.3 (A^{-1})
- It was also divided in different distances, this depending upon the radius of the sphere.

Results

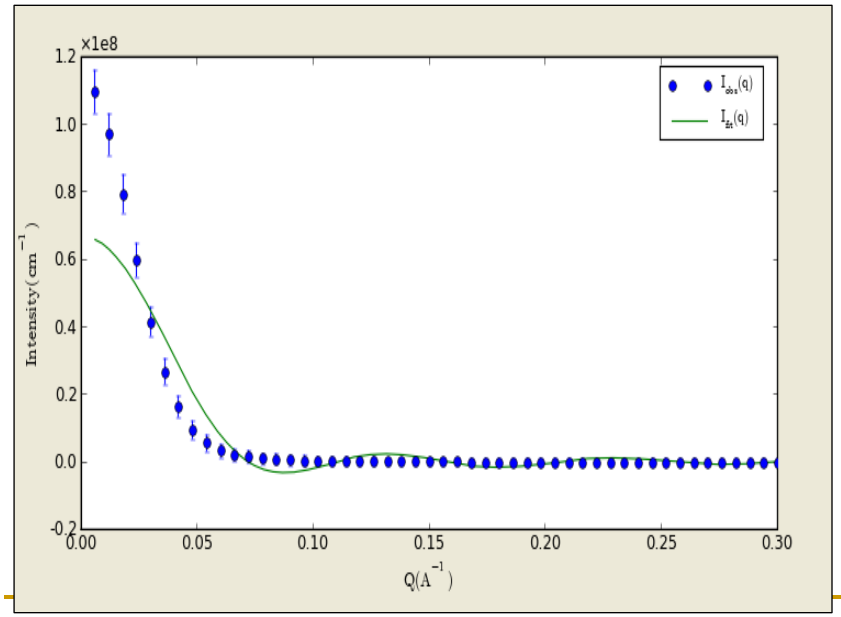
- Example:
  - Sphere with Radius 60 and polydispersity 20.
  - Best Answer: Max Distance 160 [A].

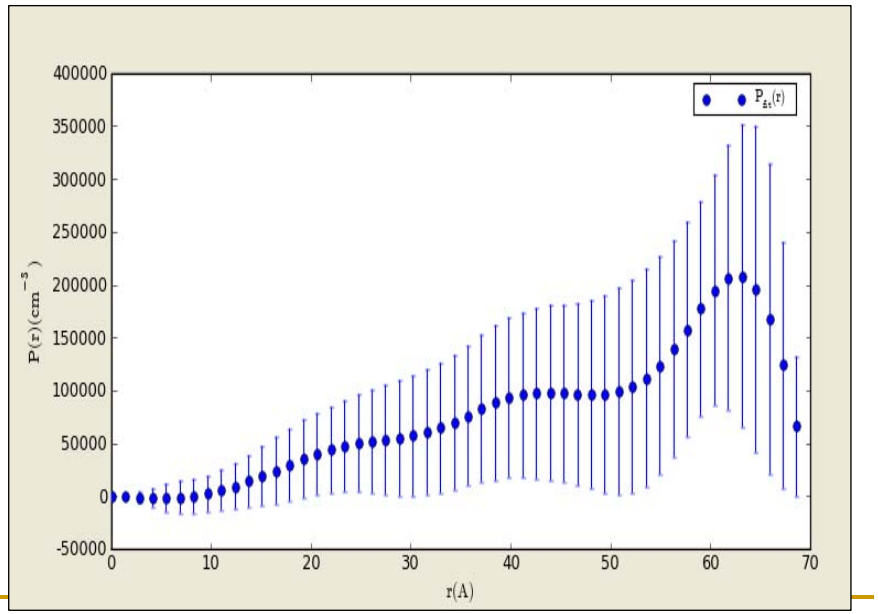


Intensity Graph



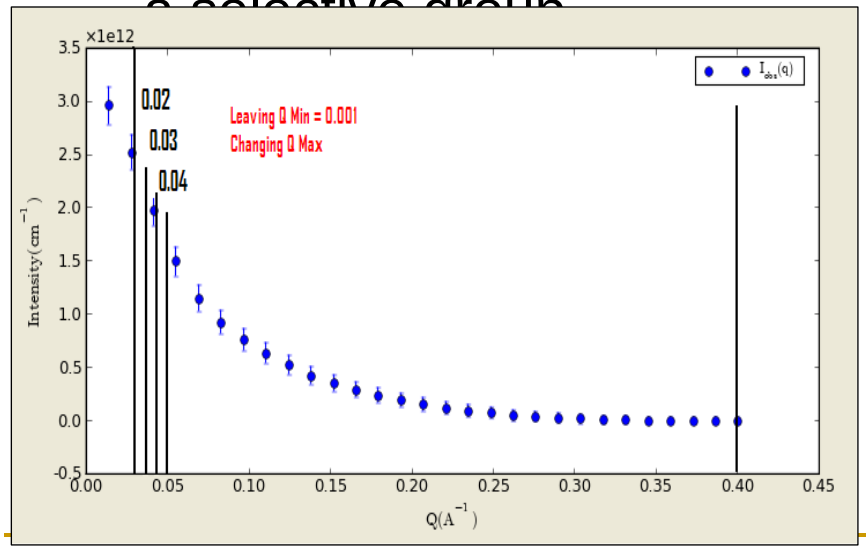
P(r) Graph

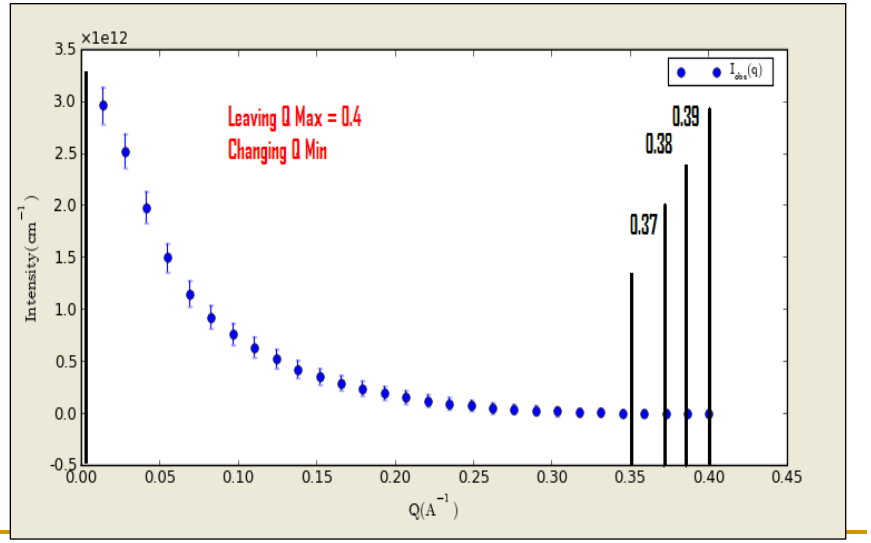

x1e8
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Intensity(cm^{-1})
Q(A^{-1})
$I_{obs}(q)$
$I_{fit}(q)$


400000
350000
300000
250000
200000
150000
100000
50000
0
-50000
0
10
20
30
40
50
60
70
P(r)(cm^{-3})
r(A)
$P_{fit}(r)$

Another Sample Geometry

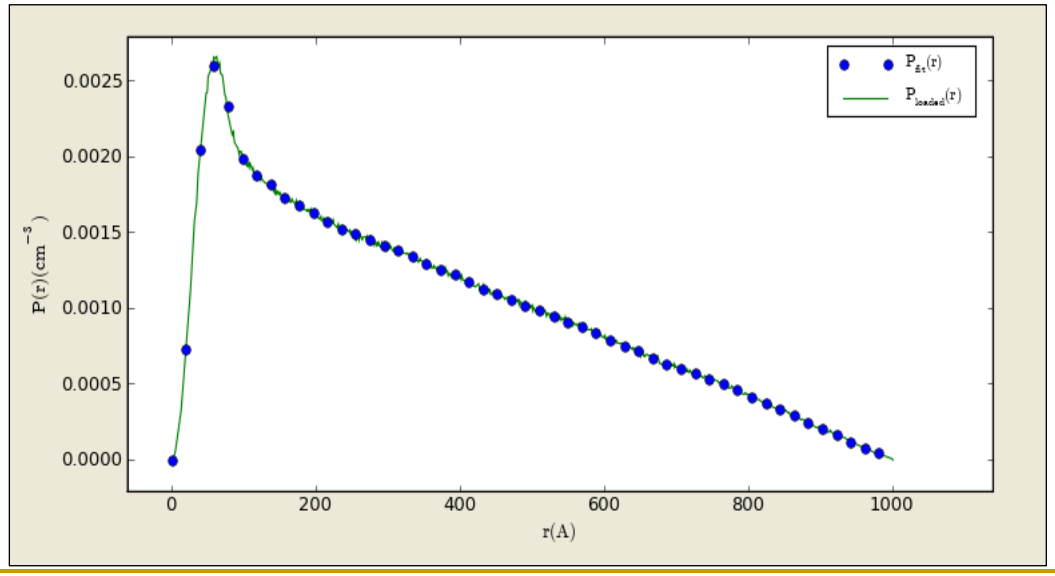
- The same test was performed in three files with cylindrical geometrical shape to see if this rules worked with other shapes.
- Also, the test for Q ranges perform for this shape was more specific. This means that all ranges were checked instead of a selective group.





Results

- Same result as in the cases shown before.
- Different shapes of cylinders, ones that need a bigger distance to work.
- Start changing the number of terms.

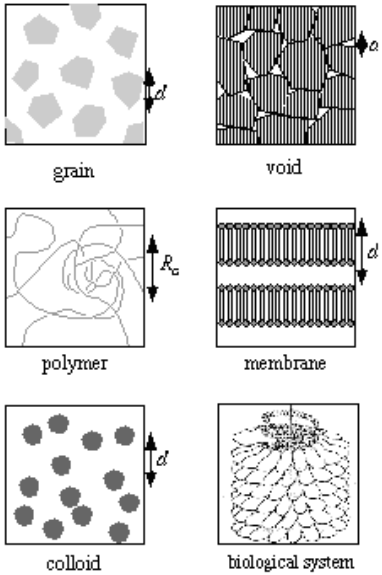


Example:

Cylinder with distance 1000 [A] that needed a number of terms = 29.

P(r) Inversion Method (TO END)

- I(Q) contains the information of interest: the shape, size ,and orientation of the structures in the sample.
- I(Q) as a function Q can be written in terms of a pair correlation function P(r), which gives the probability distribution of distances between any two point in the system.

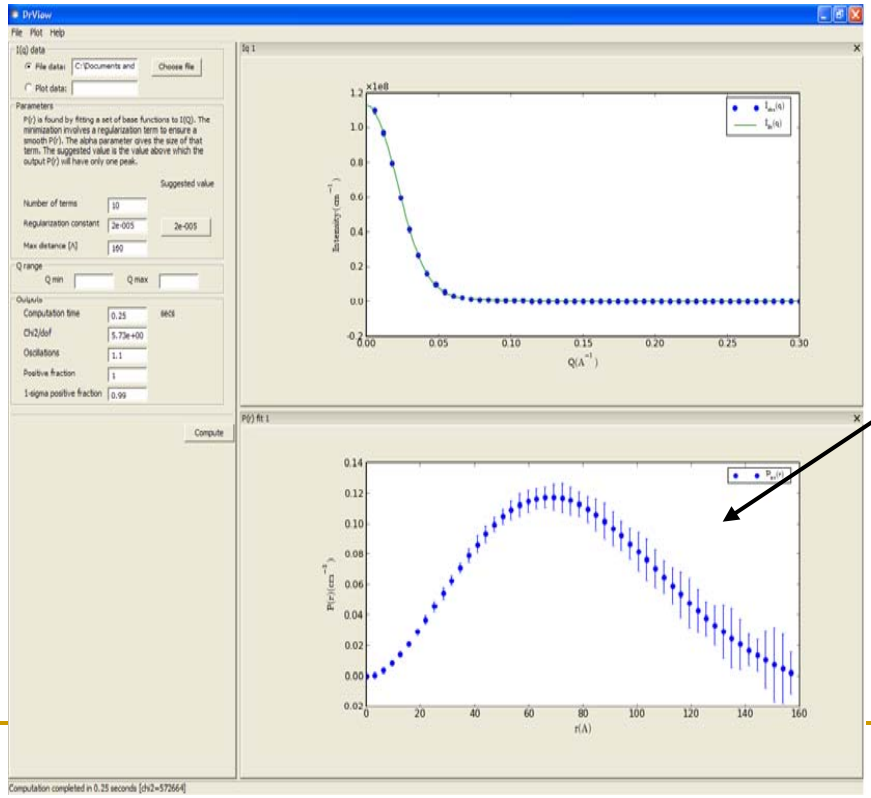


Technique for getting P(r) (TO END)

- The technique used for getting P(r) is a modified version of the technique described in P.Moore, J. Appl. Cryst. (1980) 13, 168-175.
- To evaluate P(r) we write it as an expansion of n terms of base functions:
- We can then re-write I(Q) as:
- A *regularization term* was added to ensure that the output is smooth. It is estimated numerically, and the minimization is done with a simple linear square fit.

Software Package (TO END)

- The program that allows us to generate P(r) from a given I(Q), is called ***PrView***.



of

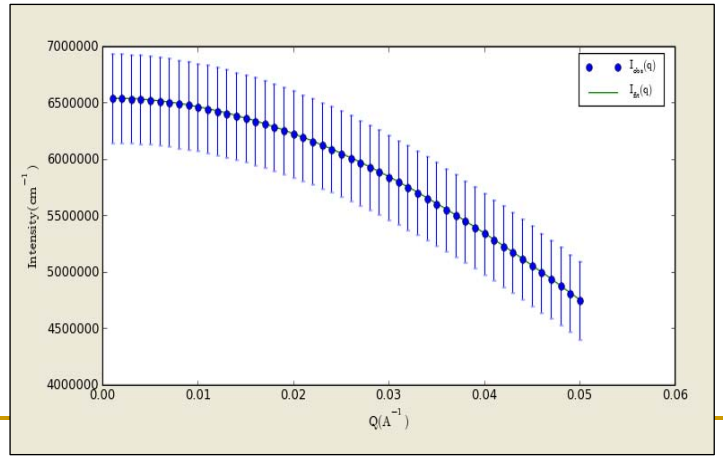
Example of a Sphere with Radius 60 [Å]

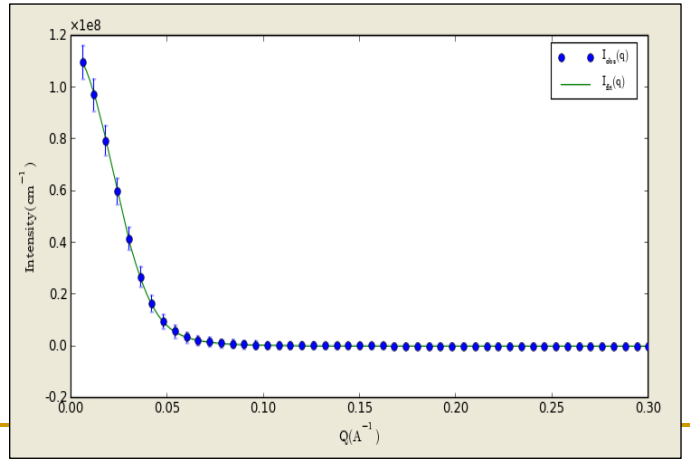
Evaluating P(r) for simulated I(q) Data (TO END)

- One example of a generated data case for P(r) is:
  - Sphere Radius 60
    - Dispersion used was 5, 10 and 20.
    - Distances used were 50, 100, 120, 140 and 160.
    - Q ranges mention before.
- All these cases were plot on Igor Pro to see the Q ranges that help the user get an approximation of the real answer.
- The estimated value for the regularization constant was also tested to see if it really helps the user get an accurate answer of P(r). This was done, generating P(r) in PrView for the different cases; the suggested value for the regularization constant was used.

Results (TO END)

- After watching carefully all the graph and comparing all the results with the answer we got to the conclusion that the Q ranges that work were:
  - 0.001 - 0.3, this is the full range
  - 0.001 - 0.02
  - 0.001 - 0.8
  - 0.02 - 0.8 (tested in some cases)





Evaluating P(r) for simulated I(q) Data (TO END)

Example of how the Intensity vs Q graph was divided.

Results

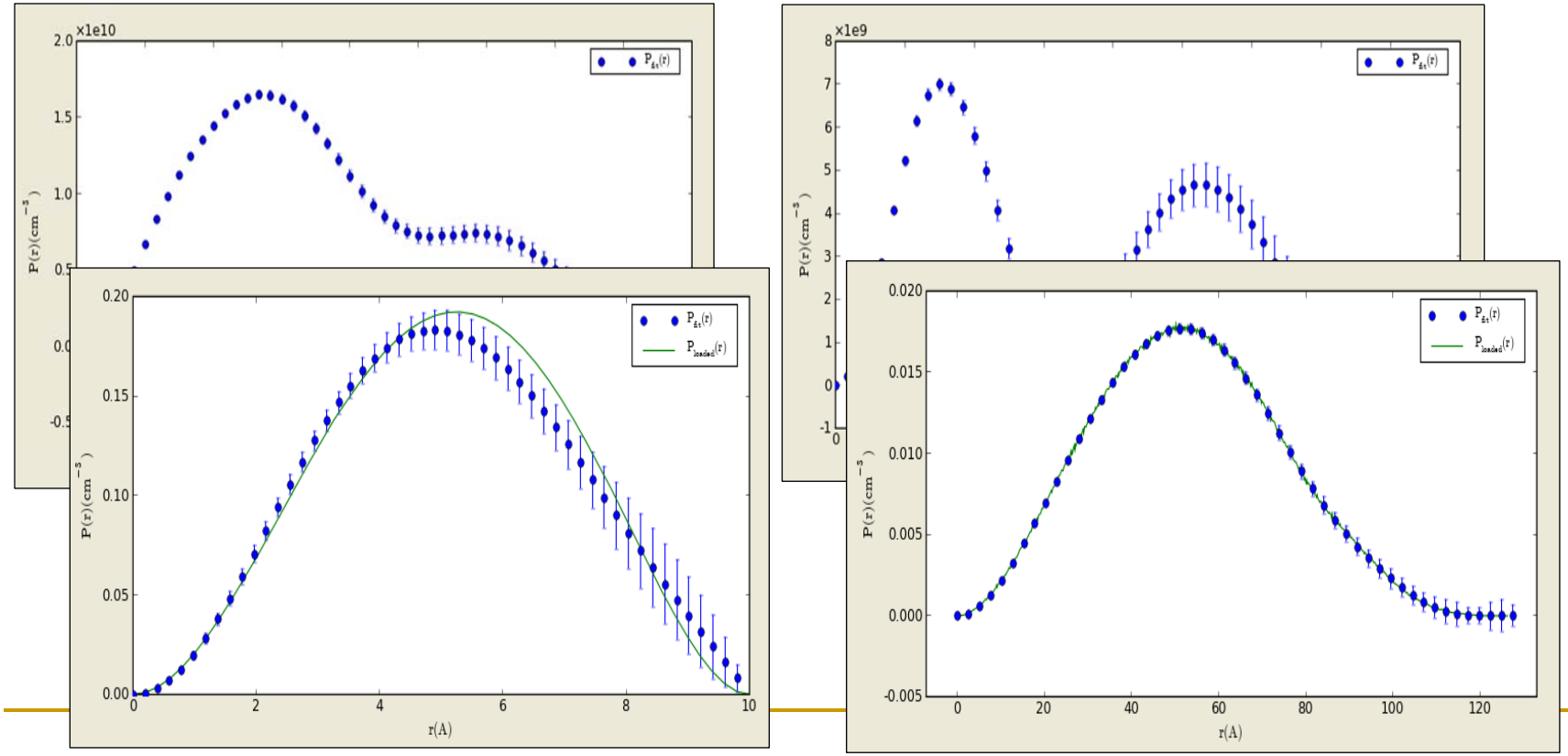
- The polydispersity (disp) changes the distance; the bigger the polydispersity, the bigger the distance it needs.
- If the distance is less than the best answer:
  - The user is going to see more oscillations with error bars in the P(r) curve, or if the guess is to small it can appear a straight line instead of a curve, meaning the answer is very inaccurate.
  - The output values are going to be large numbers, and they are supposed to be near 1.(only in sphere)???
  - The fit in the intensity graph also helps the user get a good approximation.
  - This rules apply also to the Q ranges mentioned before, that were already known to work.

Results

- If the distance is greater than the best answer:
  - The P(r) curve will look correct. The only way of knowing that the curve is incorrect, is the fact that the curve will be longer. Also, if the answer is very inaccurate, P(r) would be a line.
  - The intensity graph will also give a good fit, there is no way of knowing is wrong by just looking at this graph.
  - Again, this rules applies to the Q ranges mention before, known already by the user to work.

Results

- Another result with other geometrical shapes.



Main Goal

- Positive fraction, is the fraction of the integral of the absolute value of P(r) that is positive.
- 1-sigma positive fraction, is the fraction of the integral of the absolute value of P(r) that is at least one standard deviation above zero.

- How to get a good approximation for the Number of Terms (NT)?
- A python script was created to generate the data from different cases, this data was the output parameters.
- This data was plotted on IgorPro.

Algorithm

- A python script was created to get that constant range.
- Steps:
  - 1st: Generate data for NT (max=50), Oscillation, and 1-sigma.
  - 2nd: Get a list of 1-sigma with values 0.1 and 0.9, if those values don't exist it get 0.8 or 0.7.
  - 3rd: Get the values of oscillation in that same range.
  - 4th: Finds the median of the oscillation list.
  - 5th: Compare the median with the oscillation list.
  - 6th: Get the values for number of terms in that range.
  - 7th: Find the median of the number of terms list.

Final Goal: automate the process (dummy proof)

- It was possible to generate an algorithm to determine how many terms are necessary in the P(r) expansion to obtain a reliable result.

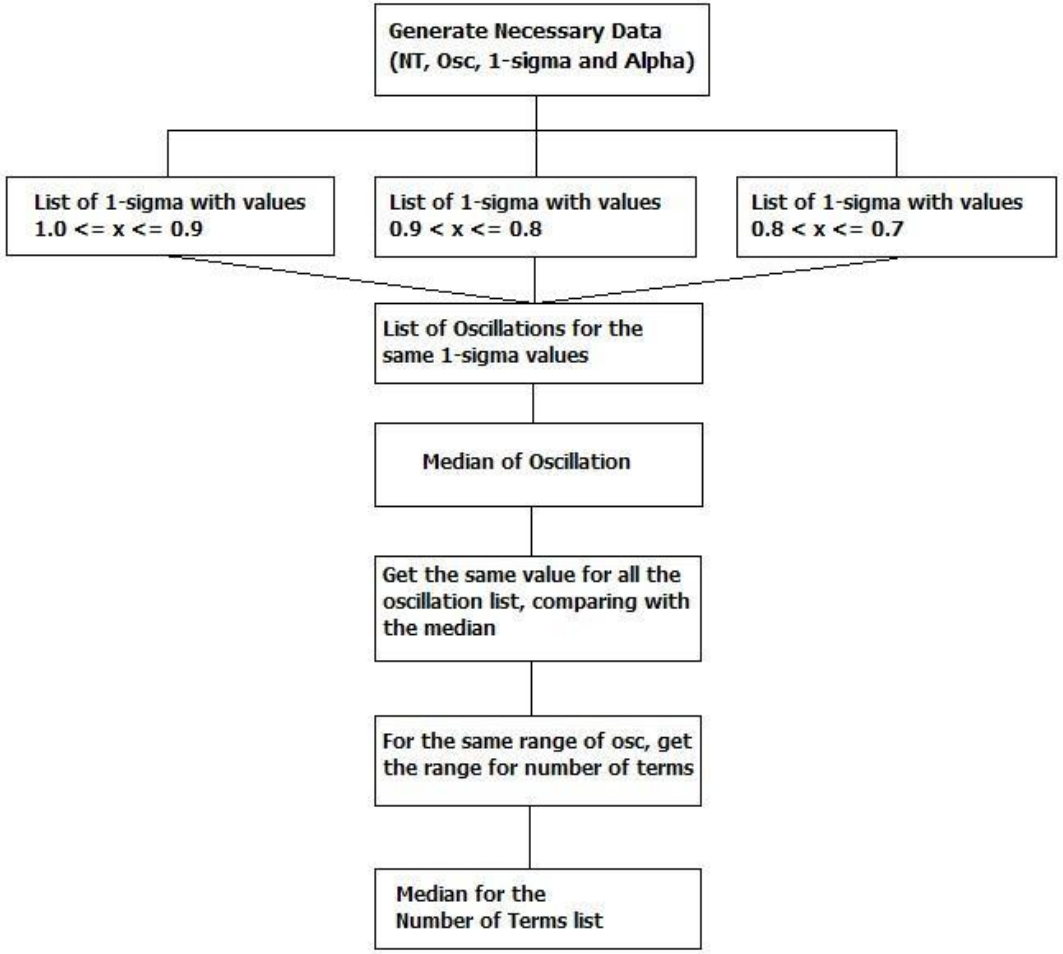
Terms needed in the expansion (*Number of terms* parameter in PrView).

- Output Parameters:
  - <u>Oscillations</u>
  - <u>Regularization Constant</u> (Alpha)

Effect of Dmax Parameter

- For the user to know how to get the length, or Max Distance (Dmax) of the system there was another test performed.
- We want to known what happens if Dmax is less or greater than the answer.
- This way the user will get a way of knowing if Dmax is inaccurate.

Algorithm


Generate Necessary Data (NT, Osc, 1-sigma and Alpha)
List of 1-sigma with values 1.0 <= x <= 0.9
List of 1-sigma with values 0.9 < x <= 0.8
List of 1-sigma with values 0.8 < x <= 0.7
List of Oscillations for the same 1-sigma values
Median of Oscillation
Get the same value for all the oscillation list, comparing with the median
For the same range of osc, get the range for number of terms
Median for the Number of Terms list

Algorithm

- Example:

Number of terms	Oscillation	Error
10	1.55110147206	0.99049328983
11	1.56463143858	0.990968545217
12	1.57528834225	0.991798334557
13	1.58334383972	0.991985481742
14	1.58968072793	0.992304376172
15	1.59438917498	0.99249516835
16	1.59801511828	0.99264643553
17	1.60063322776	0.992839545502
18	1.60260847523	0.992983638983
19	1.60393081587	0.993159578959
20	2.328674351	0.971644858232
21	1.97654711075	0.982072994389
22	19.2428263202	0.0
23	19.7812274082	0.061676875771
24	19.7147099715	0.0
25	19.7027526692	0.498931755731
26	19.7028503269	0.478278426382
27	19.6400267948	0.499941995901
28	19.6349418503	0.499930677587
29	19.6253257512	0.499999700639
30	19.6188429878	0.49999879444