UltraScan-SOMO (US-SOMO): An Integrated Hydrodynamic and Small Angle Scattering Data Analysis Software Suite

Alexey Savelyev

Workshop on SAXS and Diffraction Studies,
SLAC National Acceleration Laboratory

28-30 March 2016

SCHOOL OF MEDICINE
UT HEALTH SCIENCE CENTER®
DEPARTMENT OF BIOCHEMISTRY

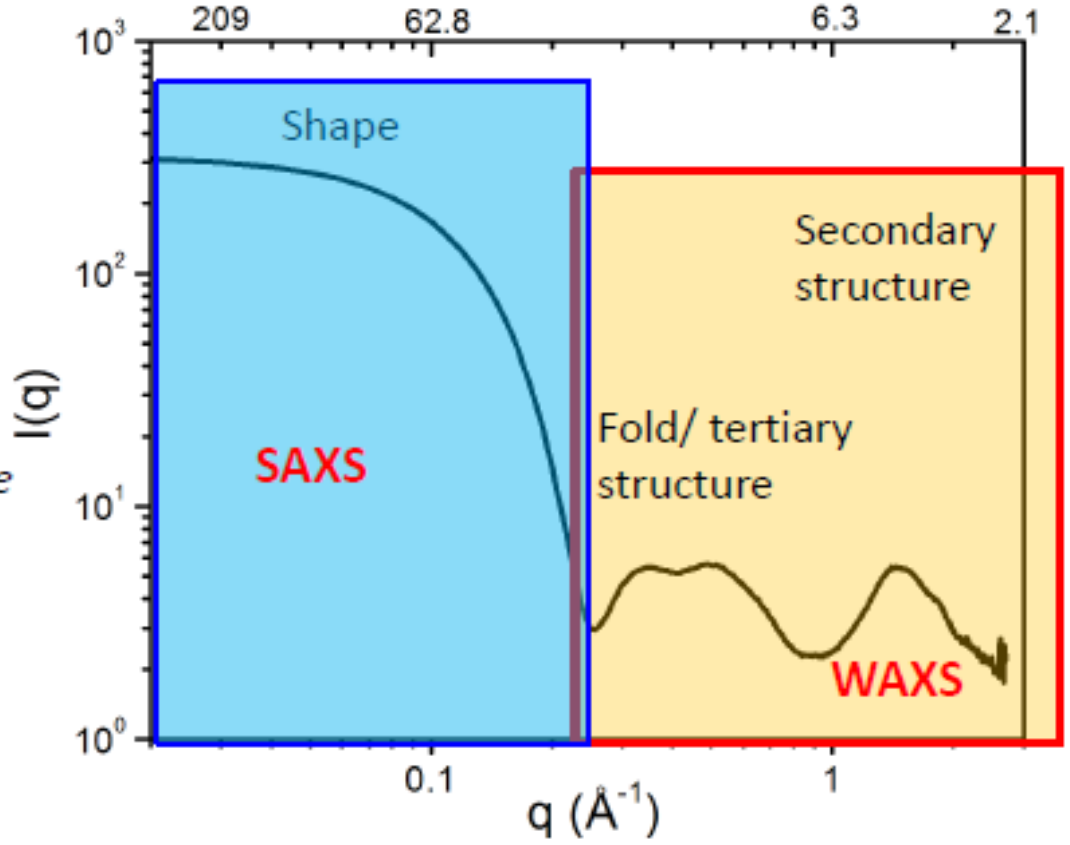
Outline

- SAXS/WAXS data application to validation of computational models of DNA
- Overview of US-SOMO Hydrodynamic and SAS capabilities
- **HPLC-SAXS module analysis tools**
    - Tutorial

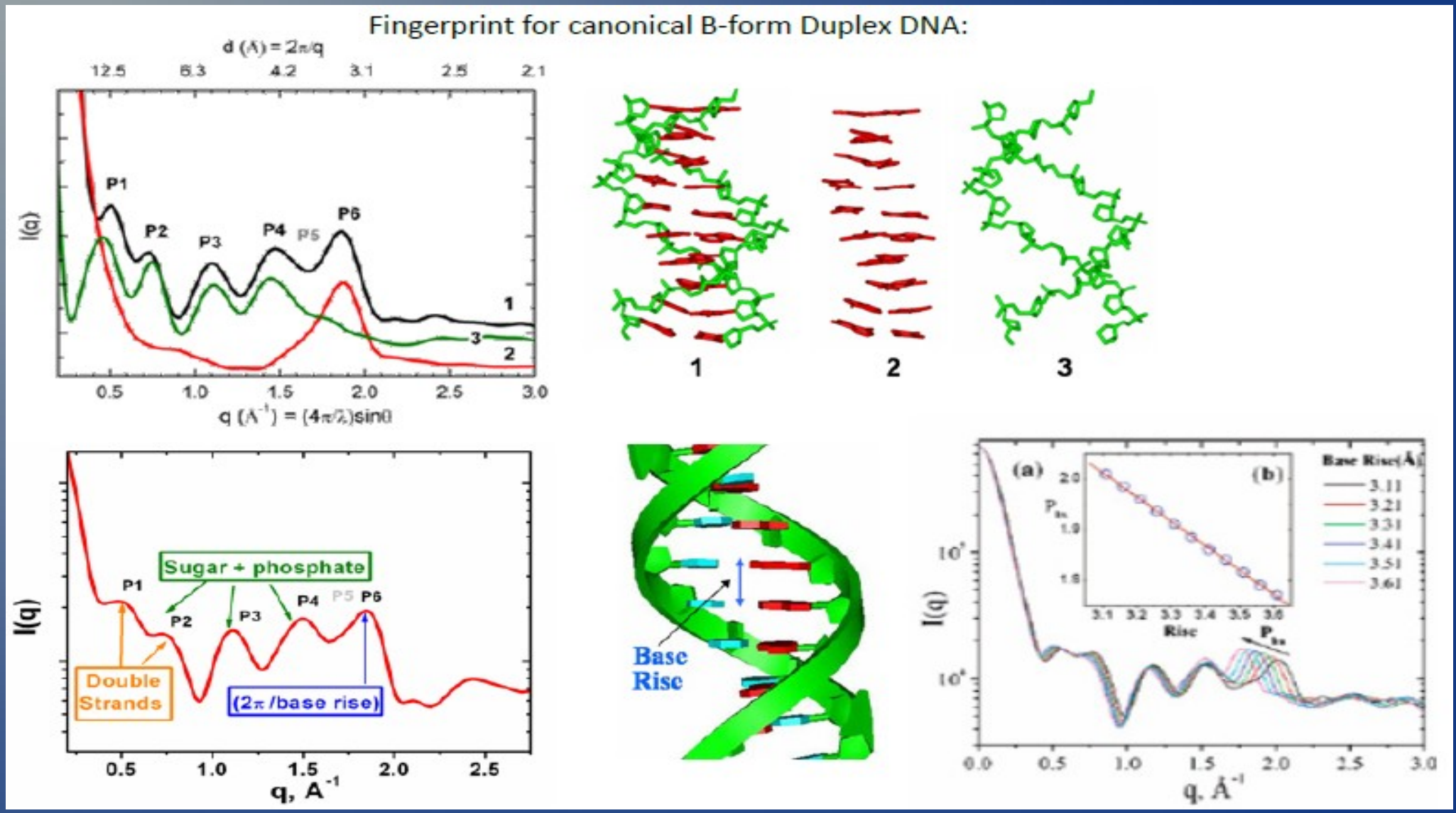
SAXS vs. WAXS Data profile

- SAXS q range: 0 - 0.2~0.4 $Å^{-1}$; WAXS: >0.2 $Å^{-1}$

- Structural Information:
  - SAXS:
    - *Size, shape, MW*
    - *Conformation*
    - *Inter-particle interactions*
    - *Molecular envelope*
  - WAXS:
    - *Fingerprints of internal structure*

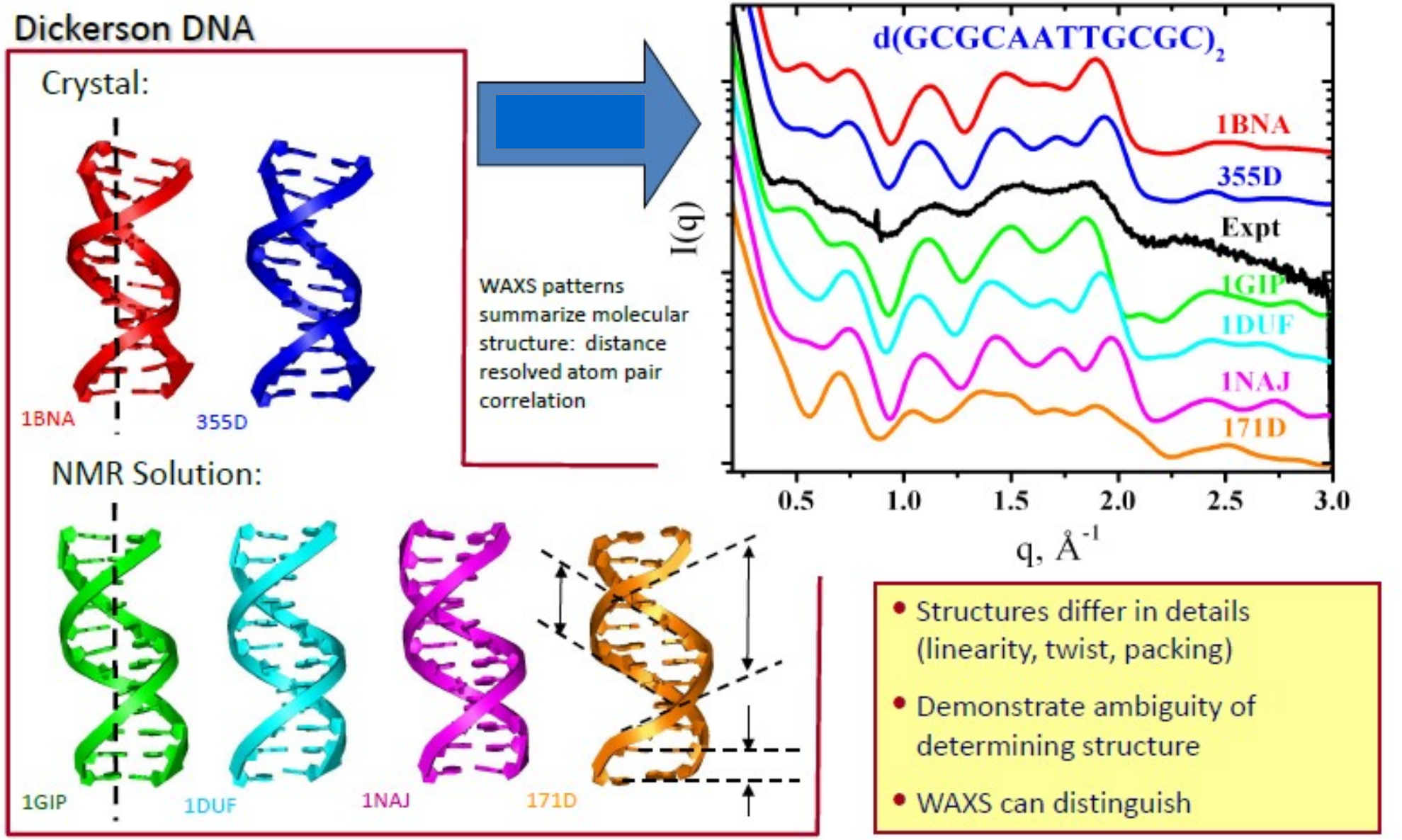


WAXS example 1: fingerprints for DNA conformation



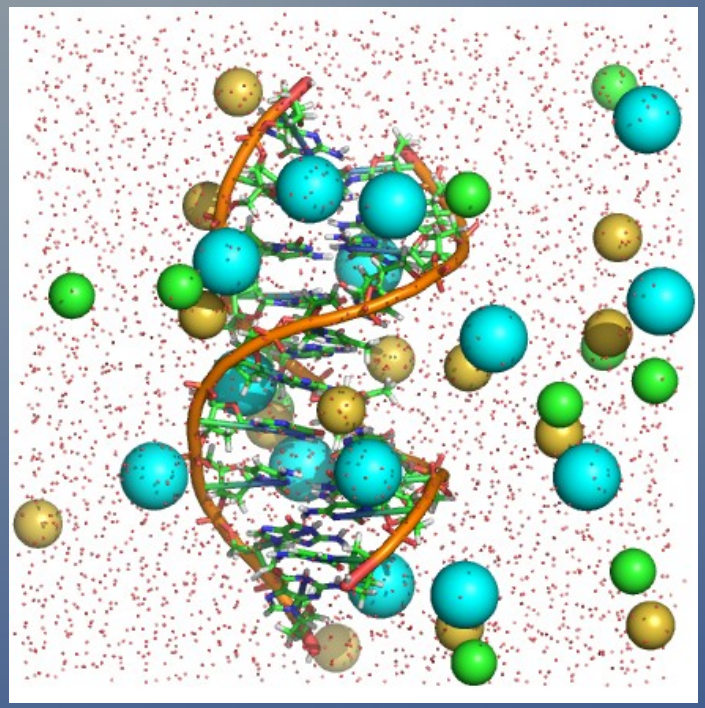
Tiede et al., PNAS 103, 3534, 2006
Tiede et al., JACS 127, 16, 2005

WAXS example 2: differentiate btw NMR and X-ray structures



Zuo, X. & Tiede, D. M. (2005). Resolving conflicting crystallographic and NMR models for solution-state DNA with solution X-ray diffraction. *J Am Chem Soc 127, 16-7.*

SAXS/WAXS example 3: differentiate btw. different all-atom computational DNA models



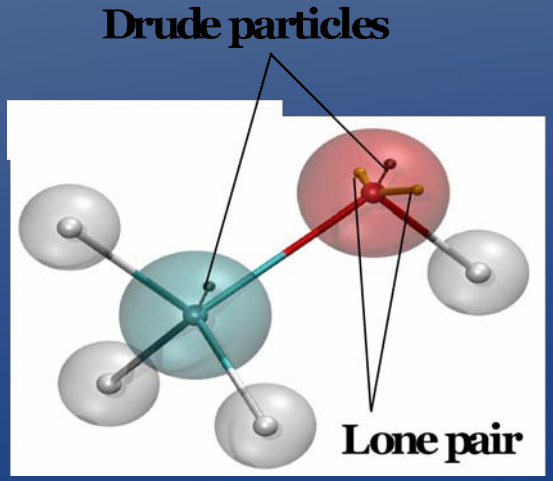
CHARMM C36, Drude and AMBER FFs

- 1DCV (10 b.p.) [B form]
- Ecor1 (12 b.p.) [B form]

AMBER setup:

32-128 CPUs

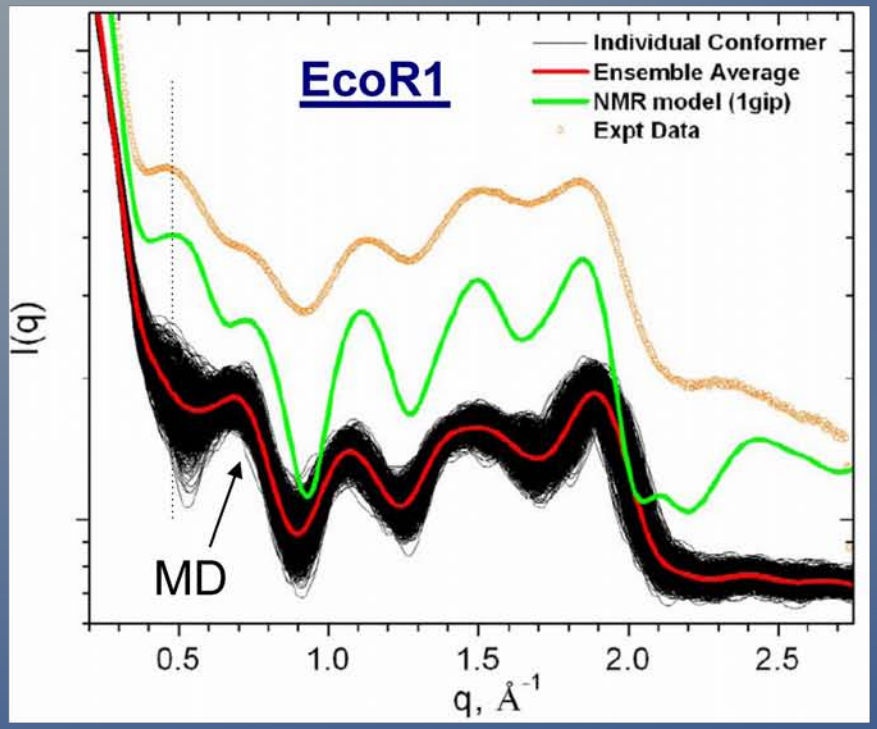
- Parmbsc0 FF for DNA;
- TIP3P water model;
- Cheatham&Joung monovalent ion parameters for Ewald and TIP3P water



New fully polarizable all-atom model for DNA based on Drude oscillator formalism

Savelyev & MacKerell, J Comp Chem, 35, 1219, 2014

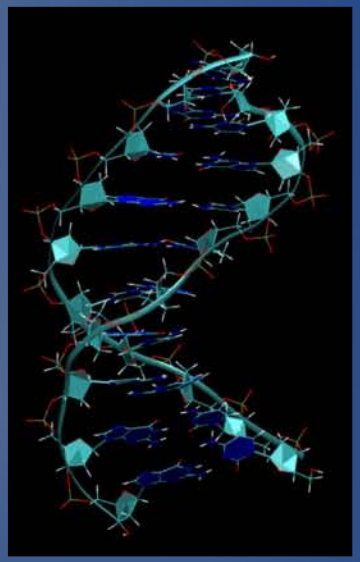
Earlier AMBER MD simulations of EcoR1



D. Tiede et al, PNAS, 103, 3534, 2006

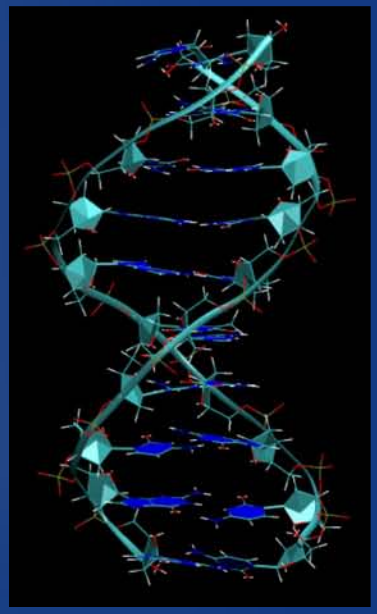
1DCV: AMBER (2014)

?? form



1DCV: DRUDE

B form



THEORY

$$I(\mathbf{q}) = \sum_j^N \sum_k^N A_j A_k e^{i\mathbf{q}\cdot\mathbf{r}_{j,k}}$$

$$I(q) = \langle I(\mathbf{q}) \rangle_\Omega = \sum_j^N \sum_k^N A_j A_k \frac{\sin qr_{j,k}}{qr_{j,k}}$$

$$A_j = f_j(q) - \rho_0 g_j(q)$$

$$g_j(q) = G(q) V_j e^{-q^2 V_j^{2/3}/4\pi}$$

$$G(q) = \frac{V_o}{V_m} e^{-q^2 (V_o^{2/3} - V_m^{2/3})/4\pi}$$

$0 < q < 25\ A^{-1}$

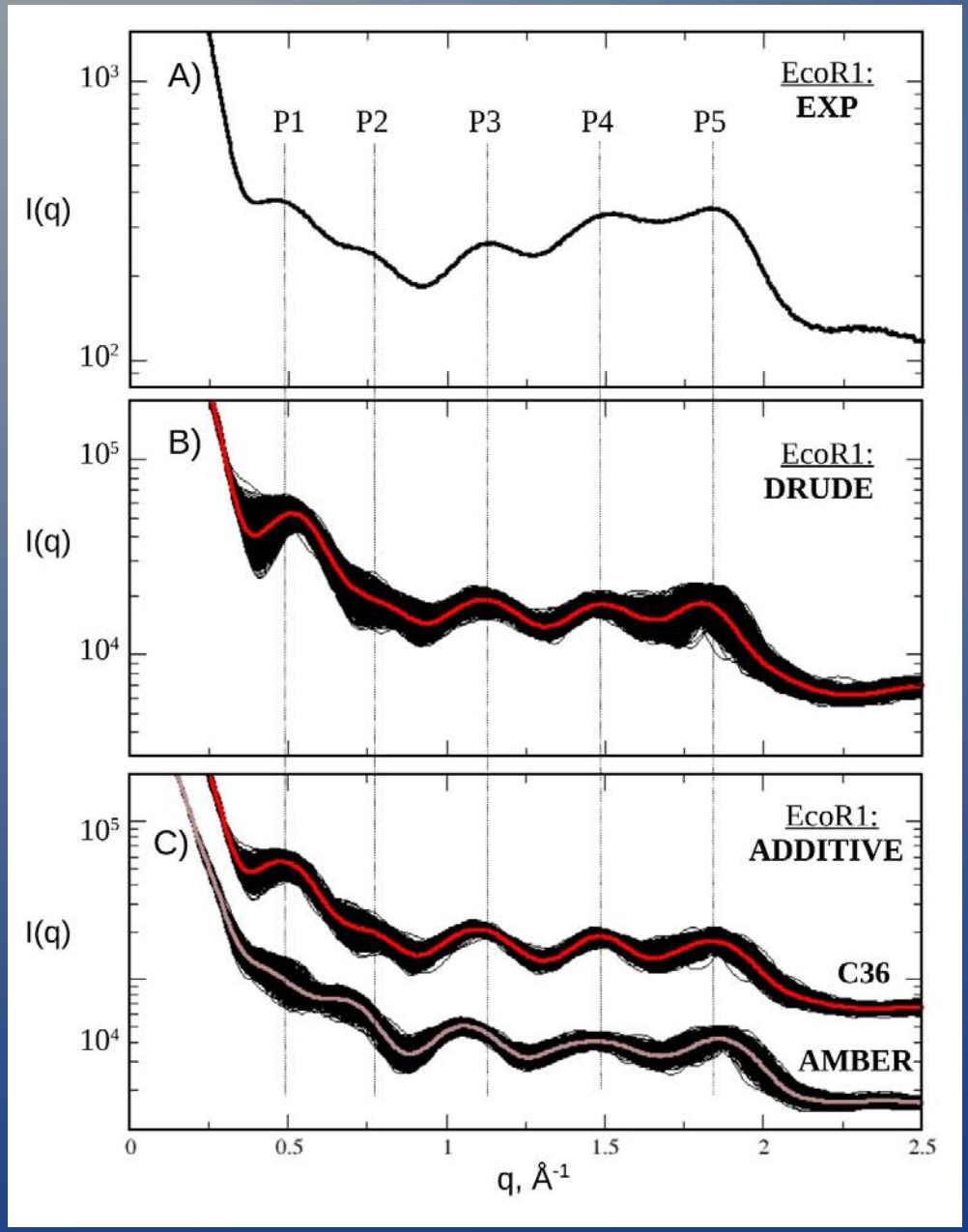
$$f(q) = \sum_{i=1}^4 a_i \exp\left(-b_i \left(\frac{q}{4\pi}\right)^2\right) + c,$$

(NIST web site)

- Look at the q up to ~2.5 Å^-1
- Process entire MD trajectory (thousands of frames)

Savelyev & MacKerell, J Phys Chem Lett, 6, 212, 2015

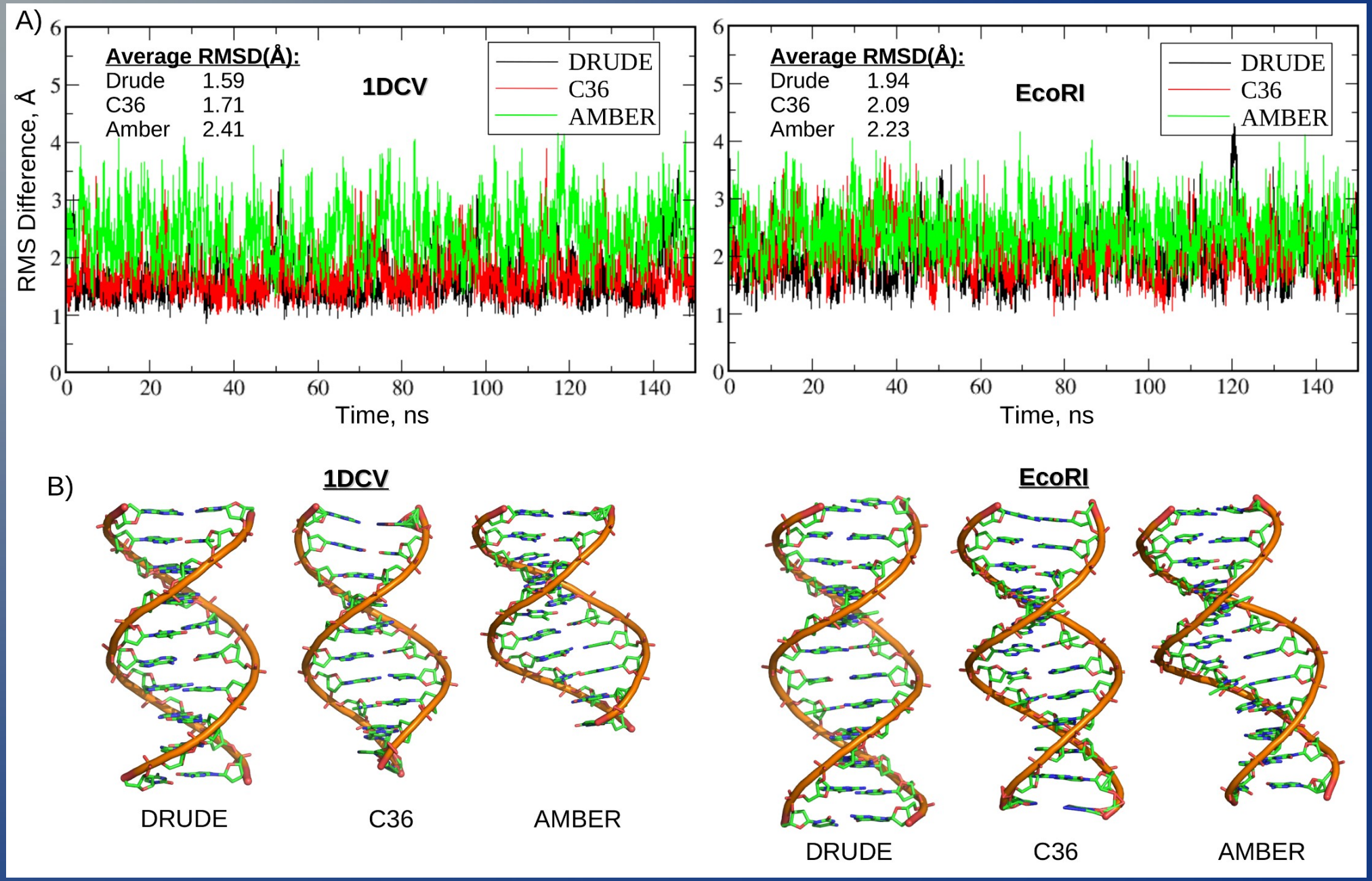
SAXS/WAXS example 3: differentiate btw. different all-atom computational DNA models



Peak	EXP	MD[a]		
		C36	Amber	Drude
EcoR1, 12-bp.				
P1	0.456	0.442	—	0.507
P2	0.750[b]	0.800[b]	0.640[b]	0.785[b]
P3	1.127	1.101	1.055	1.117
P4	1.513	1.478	1.472	1.490
P5	1.834	1.829	1.861	1.803
rmsd [P2 – P5]		0.033	0.084	0.018
rmsd_all		0.030		0.027
1DCV, 10-bp.[c]				
P1	0.510	0.480	—	0.520
P2	0.755[d]	0.800[d]	0.700[d]	0.820[d]
P3	1.180	1.100	1.050	1.110
P4	1.525	1.455	1.435	1.495
P5	1.790	1.825	1.830	1.800
rmsd [P2 – P5]		0.060	0.086	0.050
rmsd_all		0.055		0.045

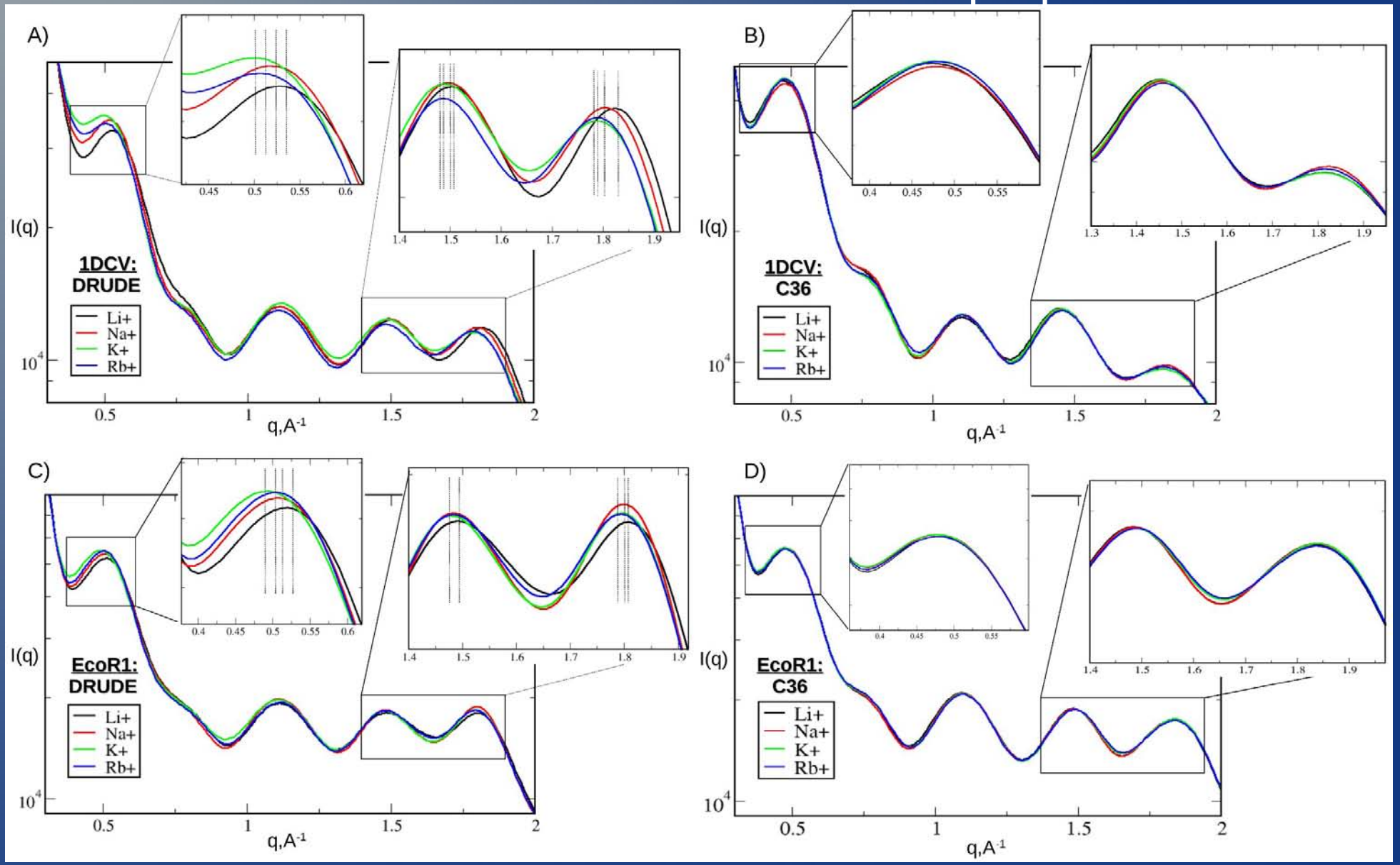
[a] Peak positions were determined from zero crossing points in the first derivative; values are in A^{-1}
[b] Approximate positions of the plato at P2
[c] Only internal 8 base pair were considered to be consistent with the experimentally studied sequence
[d] Approximate positions of the spike at P2

Savelyev & MacKerell, J Comp Chem, 35, 1219, 2014
Savelyev & MacKerell, J Phys Chem B, 118, 6742, 2014
Savelyev & MacKerell, J Phys Chem Lett, 6, 212, 2015



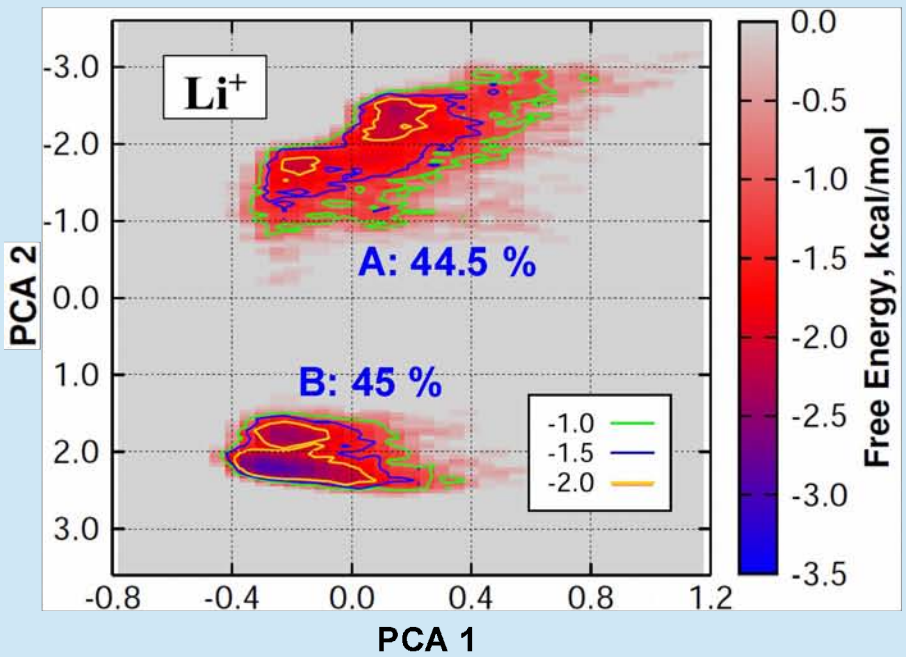
Savelyev & MacKerell, J Phys Chem Lett, 6, 212, 2015

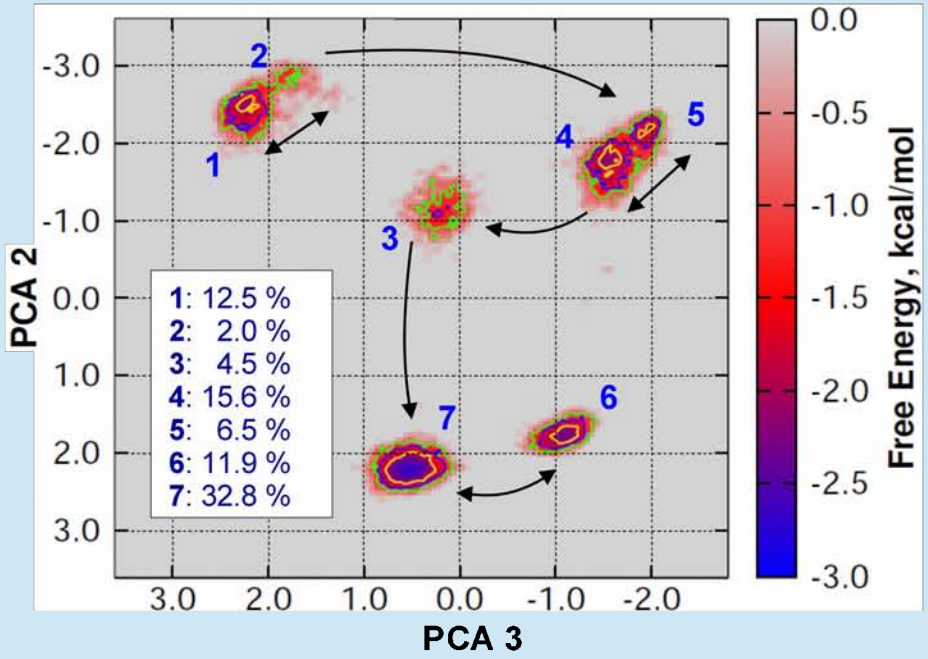
MD simulations predict differential impact of Li+, Na+, K+ and Rb+ ions on DNA conformational properties

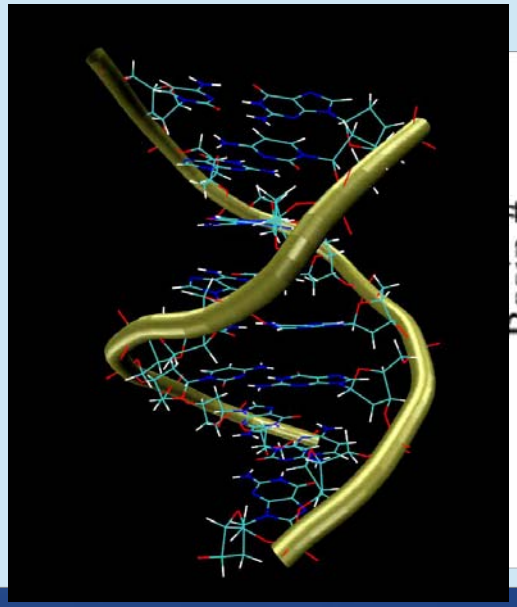


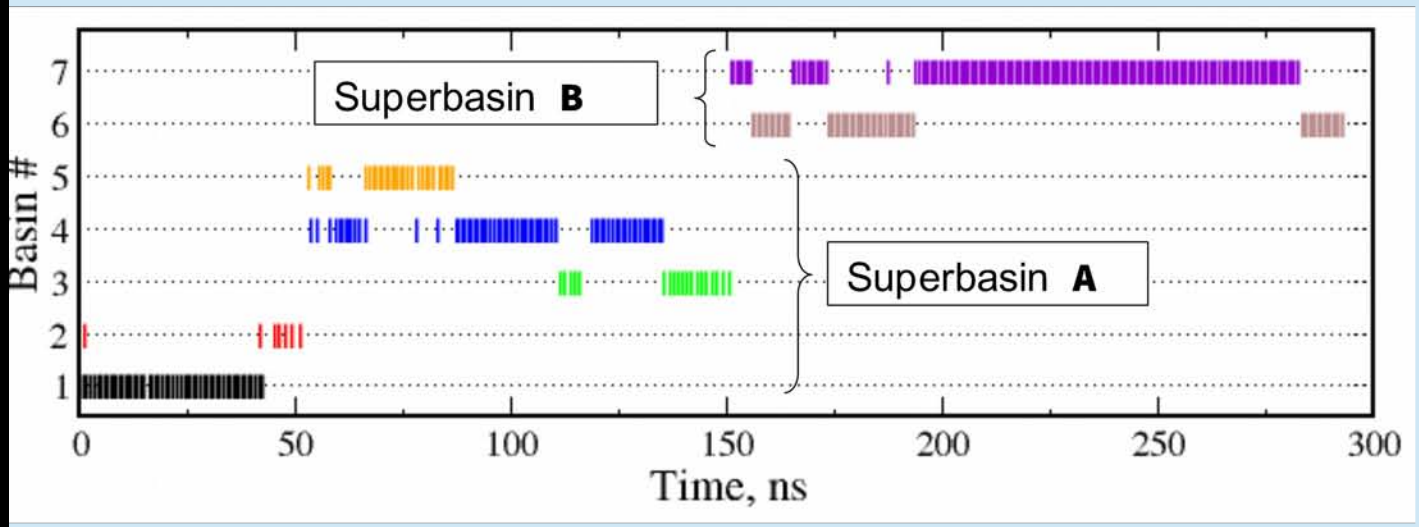
Savelyev & MacKerell, J Phys Chem Lett, 6, 212, 2015

1DCV in LiCl [DRUDE]: Dynamics Inferred from dPCA



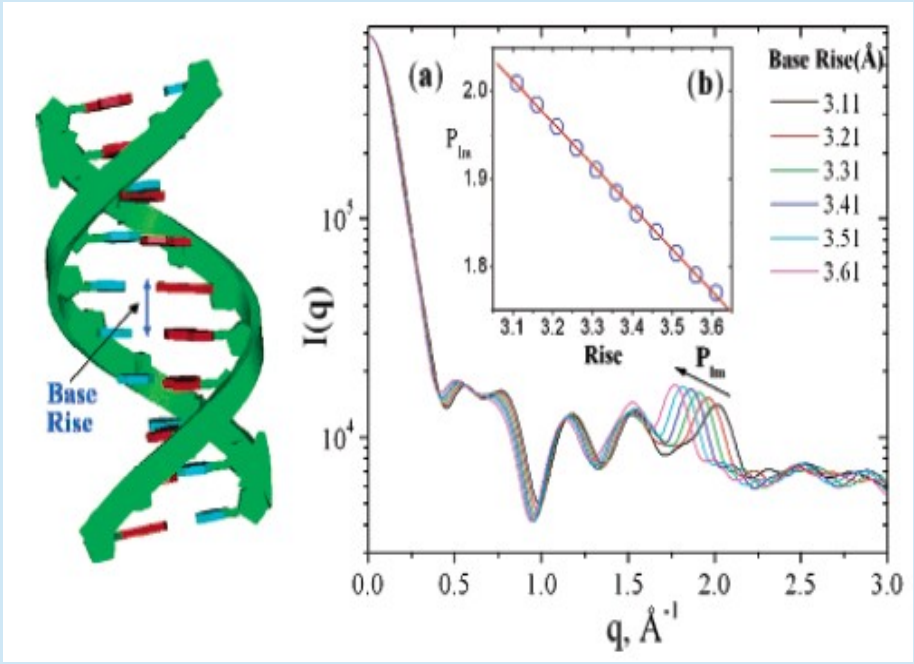






Savelyev, A, *in preparation*

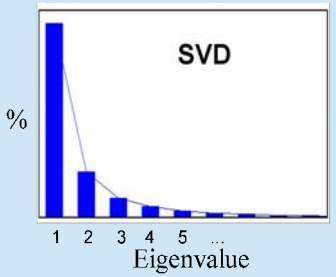
Use of the dPCA to Characterize Solution-State X-ray Spectra



$$C = T\Lambda T^T$$

$$p_i(t) = \mu_i \cdot (x(t) - < x >)$$

$$\Delta G(p_1, p_2) = -k_B T [\ln \rho(p_1, p_2) - \ln \rho_{max}]$$

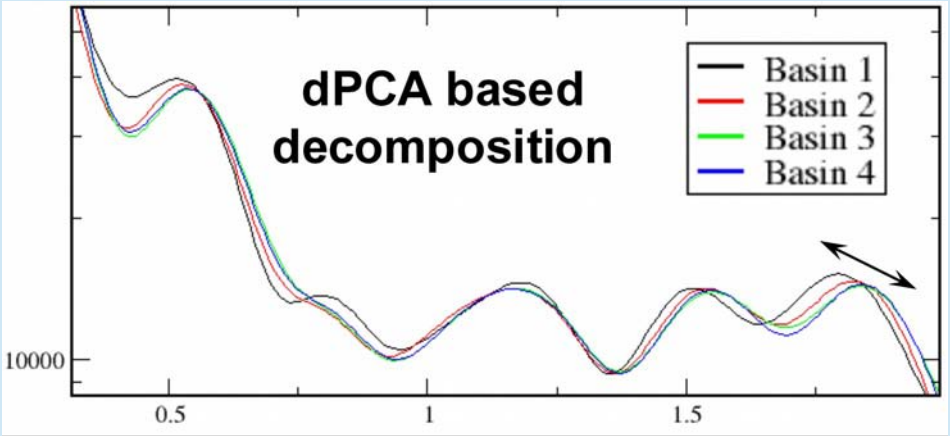


Dihedral Angle PCA (dPCA):

• To eliminate discontinuity problems associated with angular coordinates 0/2π, each angle was instead defined by its SIN and COS components

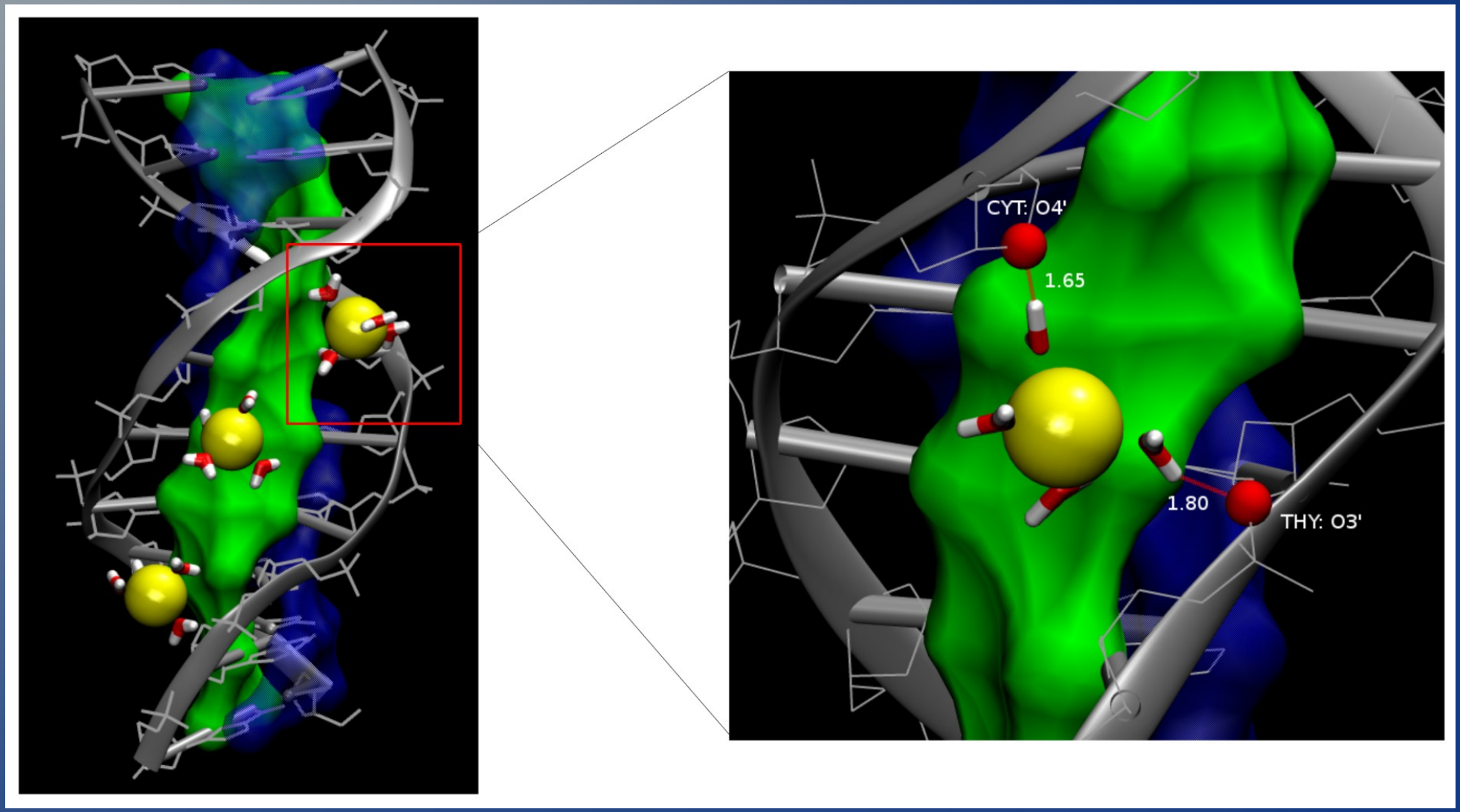
$$\varphi \mapsto \begin{cases} x = \cos\varphi \\ y = \sin\varphi \end{cases}$$

Tiede et al, JACS, 127, 16, 2005

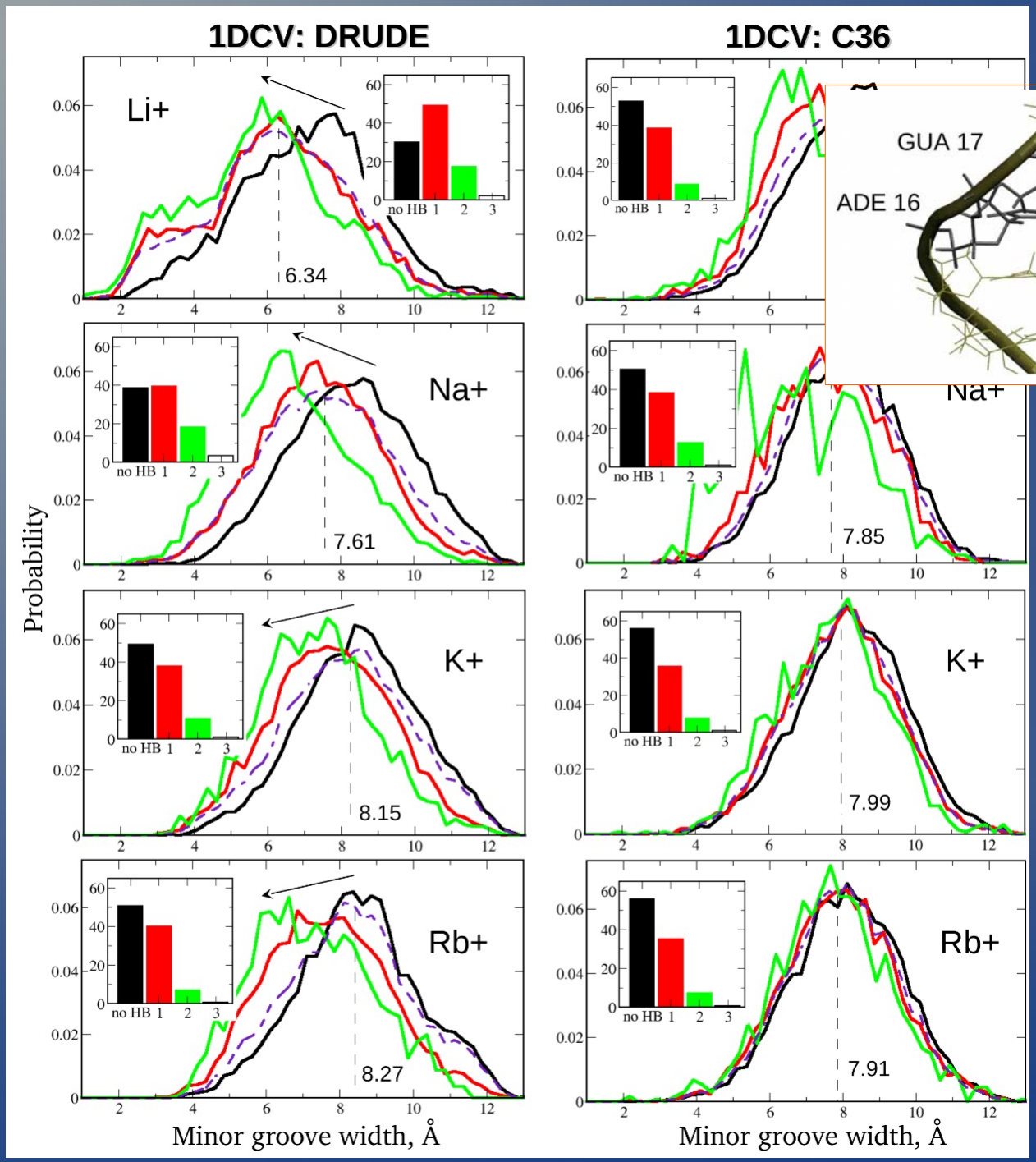


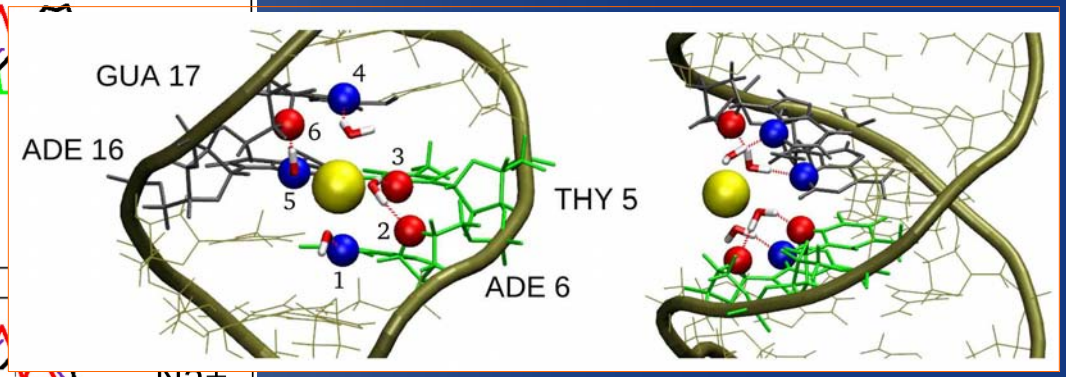
Savelyev, A, *in preparation*

Hydrated cations modulate minor groove via water-mediated hydrogen bonds with DNA strands



Savelyev & MacKerell, J Chem Theory Comput, 11, 4473, 2015





SAXS experiments on DNA in various ionic buffers are planned at NIST

Savelyev & MacKerell,
J Chem Theory Comput, 11, 4473, 2015

New developments in the UltraScan SOlution MOdeler (US-SOMO) Software Suite

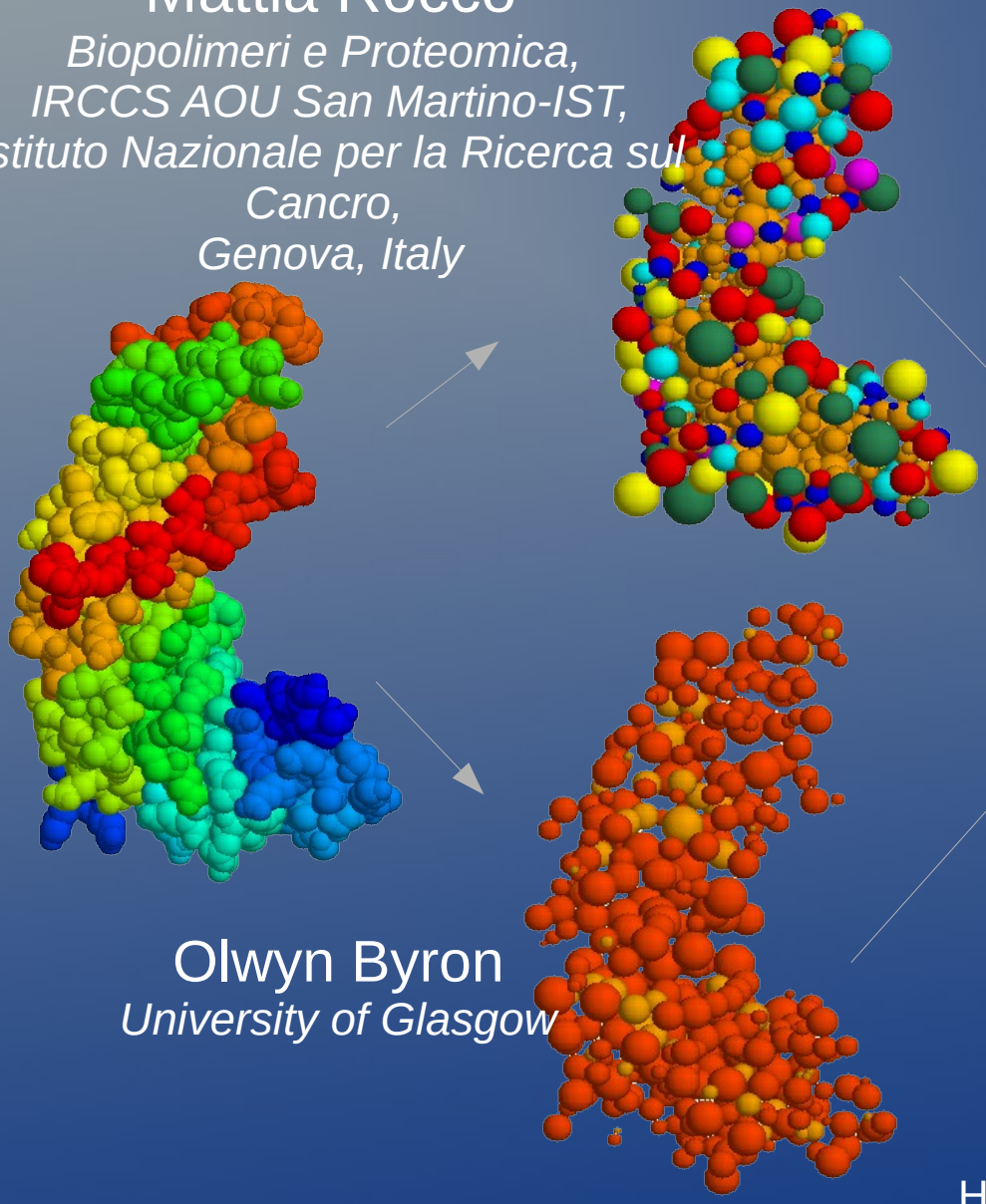
E. H. Brookes[a]
J. Pérez[b]
P. Vachette[c]
A. Savelyev[a]
M. Rocco[d]

[a] *Dept. of Biochemistry, U. of Texas Health Science Center at San Antonio, San Antonio, TX*
[b] *Beamline SWING, Synchrotron SOLEIL, Gif-sur-Yvette, France*
[c] *2BC, UMR 9198, Orsay, France*
[d] *Biopolimeri e Proteomica, IRCCS AOU San Martino-IST, Genova, Italy*

US-SOMO: Hydrodynamic calculations

Mattia Rocco

Biopolimeri e Proteomica,
IRCCS AOU San Martino-IST,
Istituto Nazionale per la Ricerca sul Cancro,
Genova, Italy



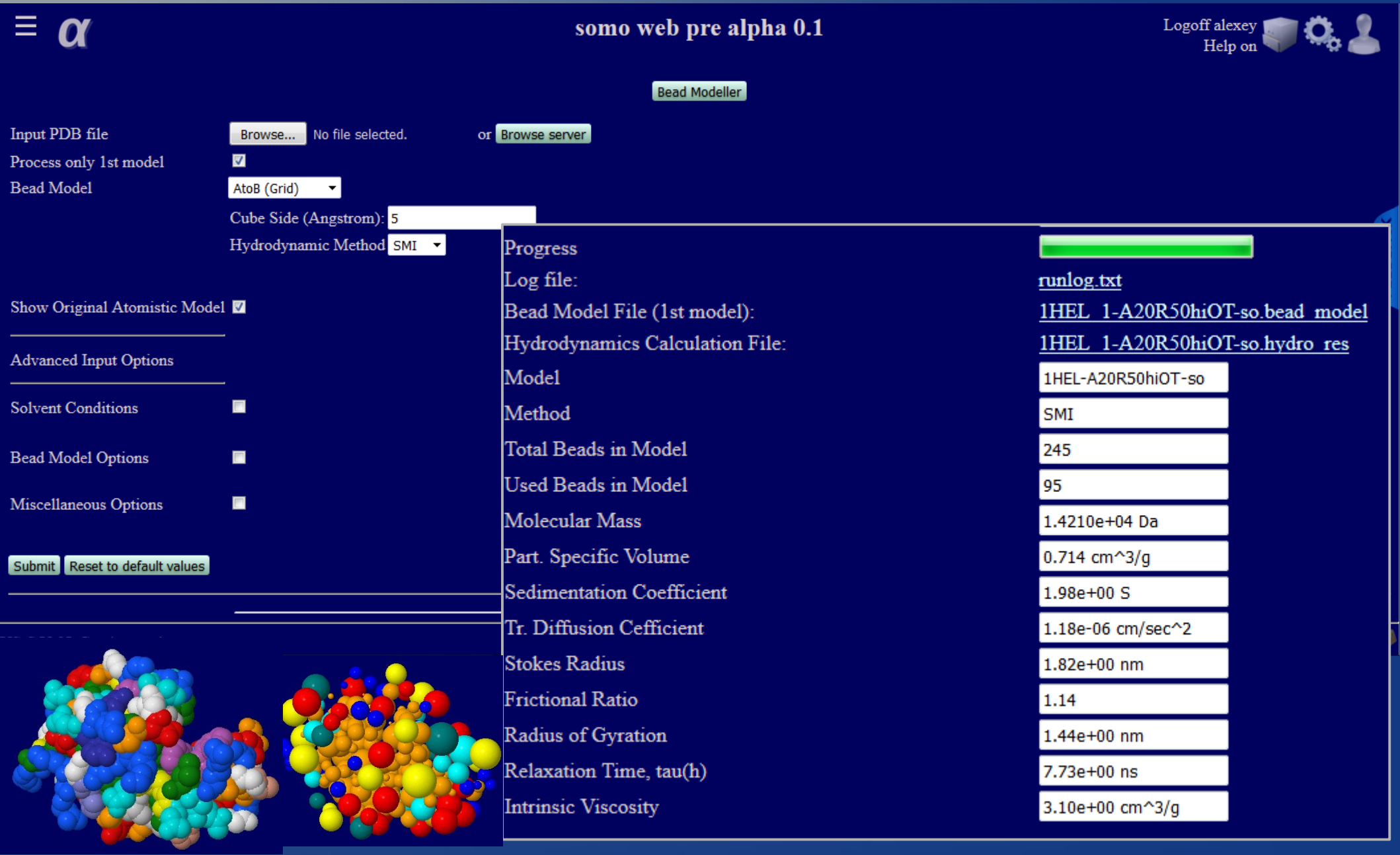
Olwyn Byron

University of Glasgow

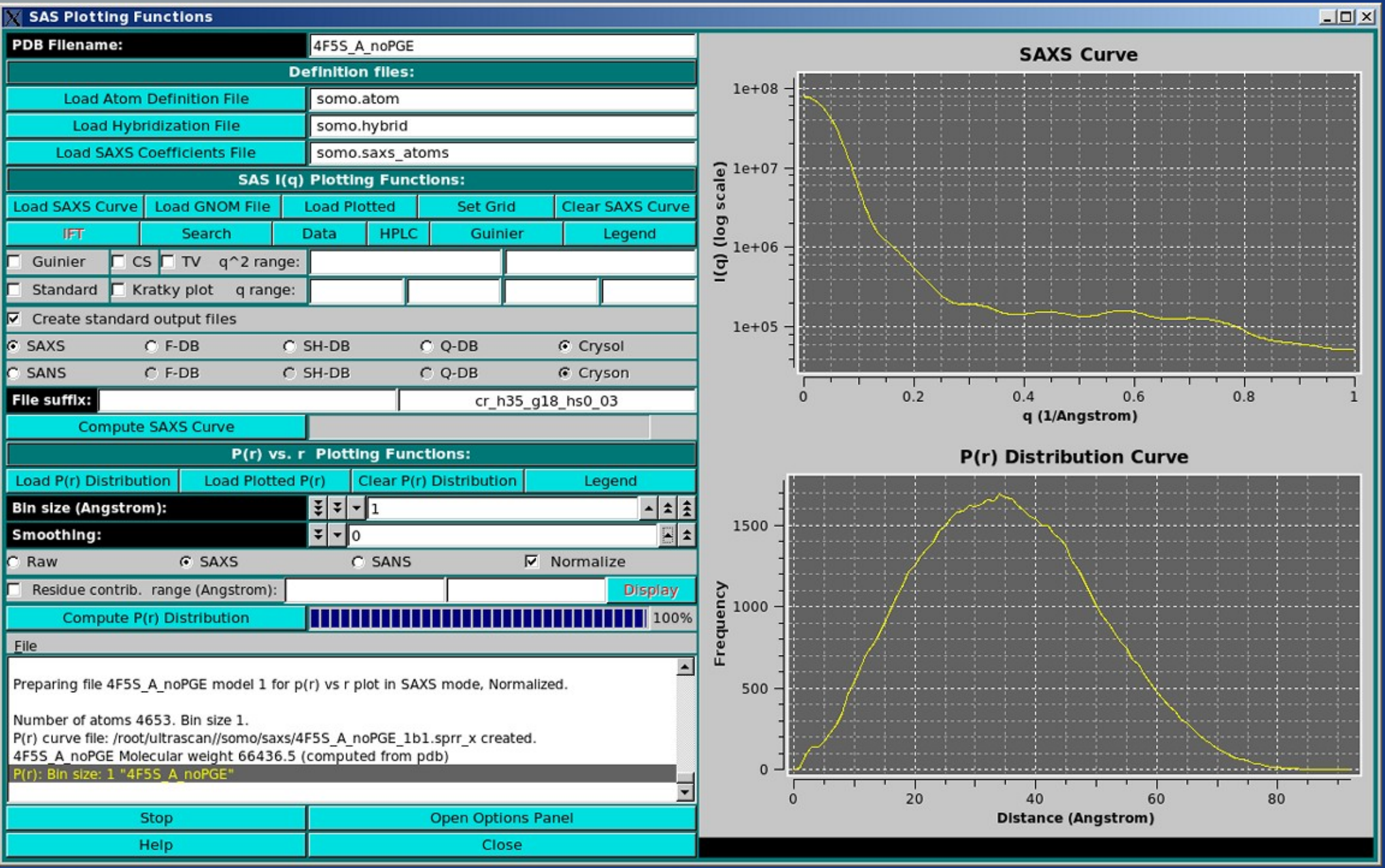
SOMO Hydrodynamic Results (Water at 20&C) (Density 1.00194 cP, Viscosity 0.998234 g/ml)	
Model:	3GKO_no_hetatm-a2b
Total Beads in Model:	439
Used Beads in Model:	369
Molecular Mass:	3.3437e+04 Da
Part. Specif. Volume:	0.733 cm^3/g
Sedimentation Coefficient s:	2.72e+00 S
Tr. Diffusion Coefficient D:	7.40e-07 cm/sec^2
Stokes Radius:	2.90e+00 nm
Frictional Ratio:	1.36
Radius of Gyration:	2.62e+00 nm
Relaxation Time, tau(h):	3.21e+01 ns
Intrinsic Viscosity:	5.06e+00 cm^3/g
View ASA Results File	View Bead Model File
View Full Hydrodynamics Results File	
Help	Close

Hydrodynamic tensor equation by matrix inversion or boundary element method or Zeno method (non rotational)

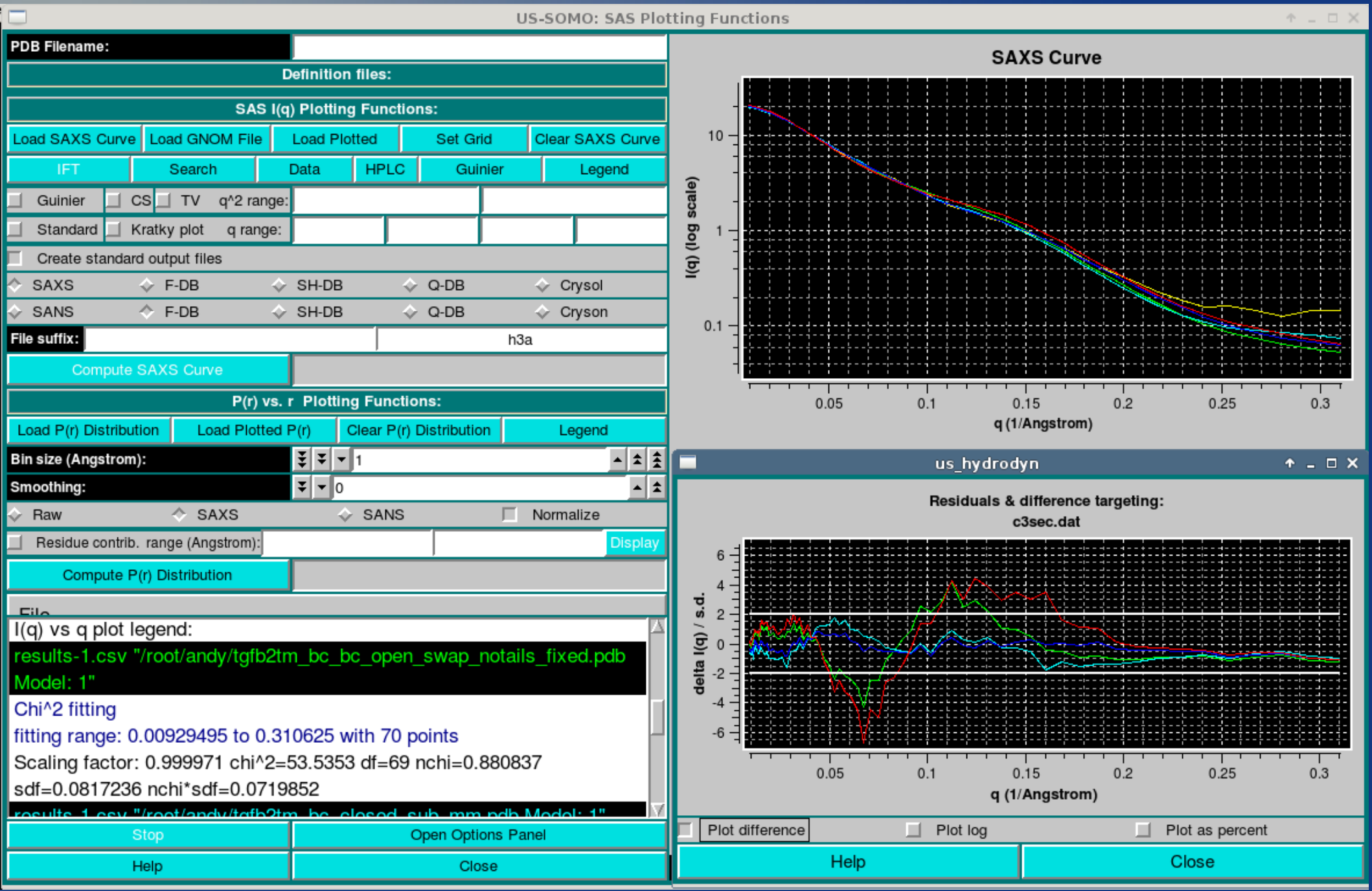
US-SOMO: GenApp Web-portal



US-SOMO SAS

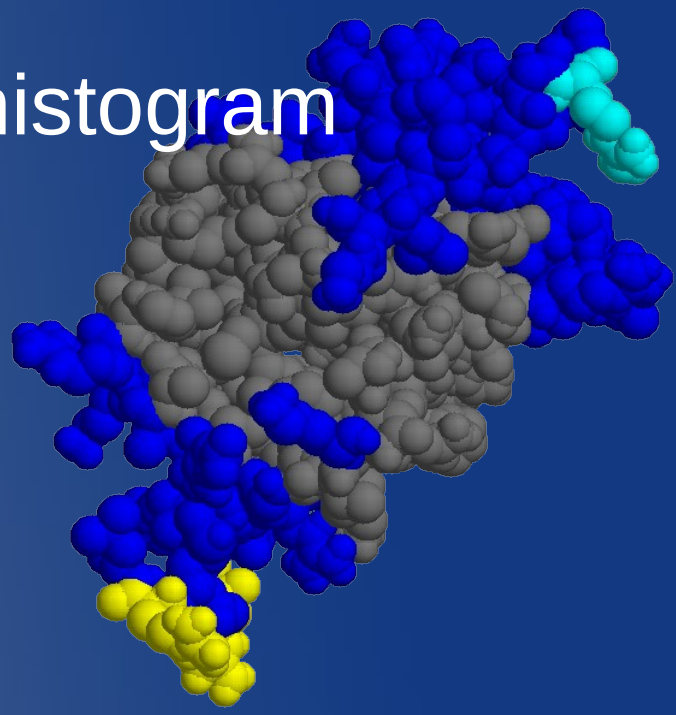

SAS Plotting Functions
PDB Filename:
4F5S_A_noPGE
Definition files:
Load Atom Definition File
somo.atom
Load Hybridization File
somo.hybrid
Load SAXS Coefficients File
somo.saxs_atoms
SAS I(q) Plotting Functions:
Load SAXS Curve
Load GNOM File
Load Plotted
Set Grid
Clear SAXS Curve
IFT
Search
Data
HPLC
Guinier
Legend
Guinier
CS
TV
q^2 range:
Standard
Kratky plot
q range:
Create standard output files
SAXS
F-DB
SH-DB
Q-DB
Crysol
SANS
F-DB
SH-DB
Q-DB
Cryson
File suffix:
cr_h35_g18_hs0_03
Compute SAXS Curve
P(r) vs. r Plotting Functions:
Load P(r) Distribution
Load Plotted P(r)
Clear P(r) Distribution
Legend
Bin size (Angstrom):
1
Smoothing:
0
Raw
SAXS
SANS
Normalize
Residue contrib. range (Angstrom):
Display
Compute P(r) Distribution
100%
File
Preparing file 4F5S_A_noPGE model 1 for p(r) vs r plot in SAXS mode, Normalized.
Number of atoms 4653. Bin size 1.
P(r) curve file: /root/ultrascan//somo/saxs/4F5S_A_noPGE_1b1.sprr_x created.
4F5S_A_noPGE Molecular weight 66436.5 (computed from pdb)
P(r): Bin size: 1 "4F5S_A_noPGE"
Stop
Open Options Panel
Help
Close
SAXS Curve
I(q) (log scale)
1e+08
1e+07
1e+06
1e+05
0
0.2
0.4
0.6
0.8
1
q (1/Angstrom)
P(r) Distribution Curve
Frequency
1500
1000
500
0
0
20
40
60
80
Distance (Angstrom)

US-SOMO SAS


US-SOMO: SAS Plotting Functions
PDB Filename:
Definition files:
SAS I(q) Plotting Functions:
Load SAXS Curve
Load GNOM File
Load Plotted
Set Grid
Clear SAXS Curve
IFT
Search
Data
HPLC
Guinier
Legend
Guinier
CS
TV
q^2 range:
Standard
Kratky plot
q range:
Create standard output files
SAXS
F-DB
SH-DB
Q-DB
Crysol
SANS
F-DB
SH-DB
Q-DB
Cryson
File suffix:
h3a
Compute SAXS Curve
P(r) vs. r Plotting Functions:
Load P(r) Distribution
Load Plotted P(r)
Clear P(r) Distribution
Legend
Bin size (Angstrom):
1
Smoothing:
0
Raw
SAXS
SANS
Normalize
Residue contrib. range (Angstrom):
Display
Compute P(r) Distribution
I(q) vs q plot legend:
results-1.csv "/root/andy/tgfb2tm_bc_bc_open_swap_notails_fixed.pdb Model: 1"
Chi^2 fitting
fitting range: 0.00929495 to 0.310625 with 70 points
Scaling factor: 0.999971 chi^2=53.5353 df=69 nchi=0.880837
sdf=0.0817236 nchi*sdf=0.0719852
Stop
Open Options Panel
Help
Close
SAXS Curve
I(q) (log scale)
10
1
0.1
0.05
0.1
0.15
0.2
0.25
0.3
q (1/Angstrom)
us_hydrodyn
Residuals & difference targeting:
c3sec.dat
delta I(q) / s.d.
6
4
2
0
-2
-4
-6
0.05
0.1
0.15
0.2
0.25
0.3
q (1/Angstrom)
Plot difference
Plot log
Plot as percent
Help
Close

US-SOMO: P(r) structural histogram

- Compute P(r)
- Select distance range
- Display residues colored by contribution



P(r) vs. r Plotting Functions:

Bin size (Angstrom): 2

Raw SAXS SANS Normalize

Residue contrib. range (Angstrom): 40 50 Display

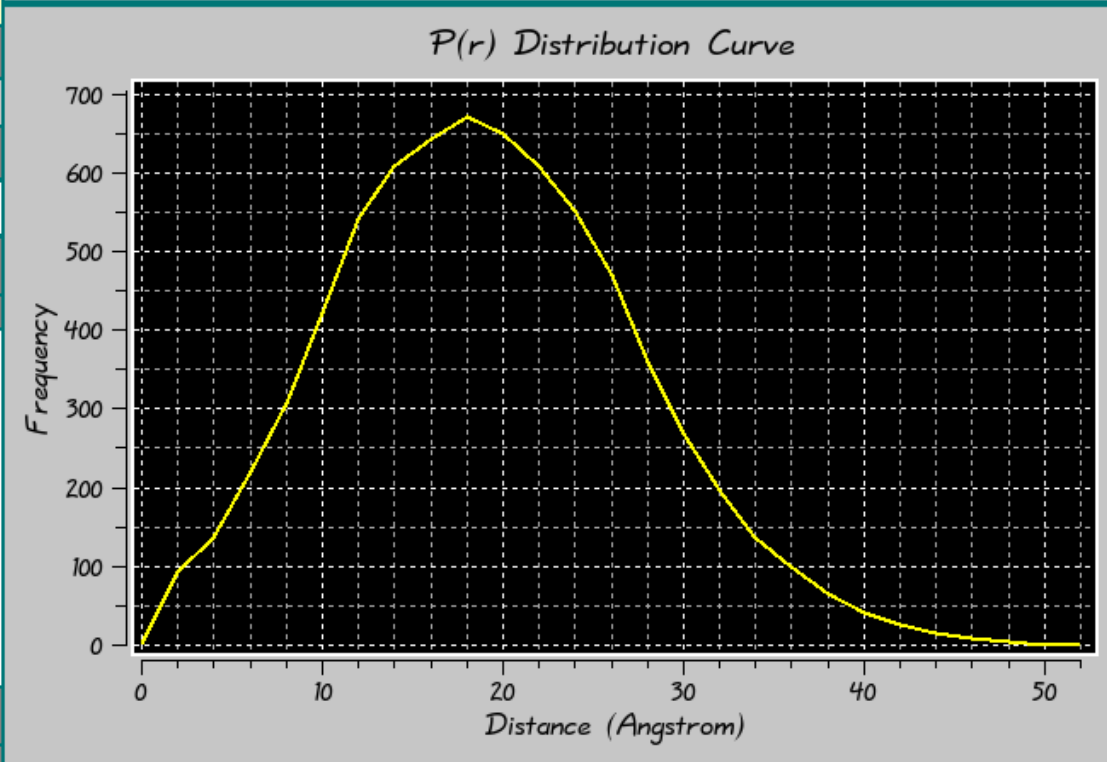
Compute P(r) Distribution 100%

Load P(r) Distribution | Load Plotted P(r) | Clear P(r) Distribution | Legend

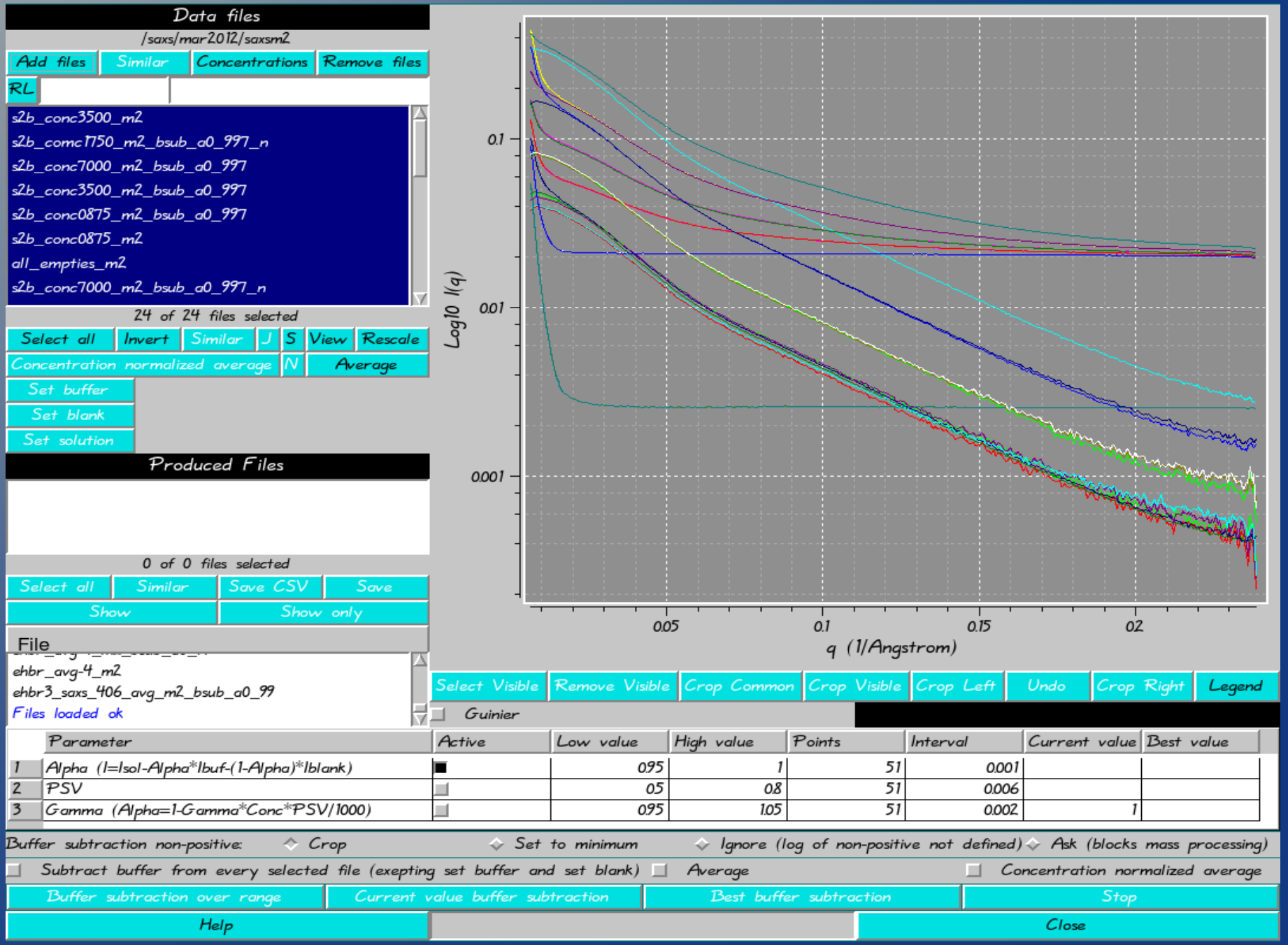
File

Number of atoms 1000. Bin size 2.
P(r) curve
file: /root/ultrascan//somo/saxs/1HEL_s_equi_tp_6_tm10000_m-9_1b2.sprr_
created.
1HEL_s_equi_tp_6_tm10000_m-9 Molecular weight 14289.7 (computed from pdb)
P(r): Bin size: 2 "1HEL_s_equi_tp_6_tm10000_m-9"

Stop | Open Options Panel

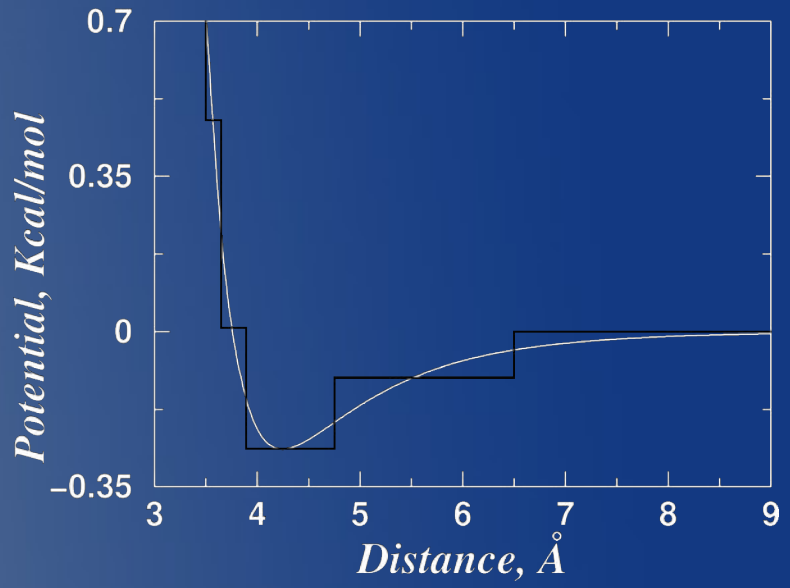


US-SOMO: SAS data treatment



Discrete MD

- Explores conformational space
- Faster than standard MD
- Discretizes the potential function
- 20,000 * 50 fs = 1 ns
- Implicit solvent
- Anderson thermostat



Ding F, Dokholyan NV. *Emergence of protein fold families through rational design.* *Public Library of Science,* Comput Biol. (2006) 2(7):e85

US-SOMO: Cluster DMD Setup

Cluster DMD setup

	PDB file	Active	Relax temp kcal/mol/kB	Relax time * 50fs	Relax PDB output timestep	Relax PDB output count	Run temp kcal/mol/kI	Run time * 50fs	Run PDB output timestep	Run PDB output count	Static range
1	1HEL_s_equi_tp_6_tm10000_m-1.pdb	■	.7	100	50	2	.5	10000	200	50	
2	1HEL_s_equi_tp_6_tm10000_m-10.pdb	■	.7	100	50	2	.5	10000	200	50	
3	1HEL_s_equi_tp_6_tm10000_m-11.pdb	■	.7	100	50	2	.5	10000	200	50	
4	1HEL_s_equi_tp_6_tm10000_m-12.pdb	■	.7	100	50	2	.5	10000	200	50	
5	1HEL_s_equi_tp_6_tm10000_m-13.pdb	■	.7	100	50	2	.5	10000	200	50	
6	1HEL_s_equi_tp_6_tm10000_m-14.pdb	■	.7	100	50	2	.5	10000	200	50	
7	1HEL_s_equi_tp_6_tm10000_m-15.pdb	■	.7	100	50	2	.5	10000	200	50	
8	1HEL_s_equi_tp_6_tm10000_m-16.pdb	■	.7	100	50	2	.5	10000	200	50	
9	1HEL_s_equi_tp_6_tm10000_m-17.pdb	■	.7	100	50	2	.5	10000	200	50	
10	1HEL_s_equi_tp_6_tm10000_m-18.pdb	■	.7	100	50	2	.5	10000	200	50	
11	1HEL_s_equi_tp_6_tm10000_m-19.pdb	■	.7	100	50	2	.5	10000	200	50	
12	1HEL_s_equi_tp_6_tm10000_m-2.pdb	■	.7	100	50	2	.5	10000	200	50	
13	1HEL_s_equi_tp_6_tm10000_m-20.pdb	■	.7	100	50	2	.5	10000	200	50	

US-SOMO

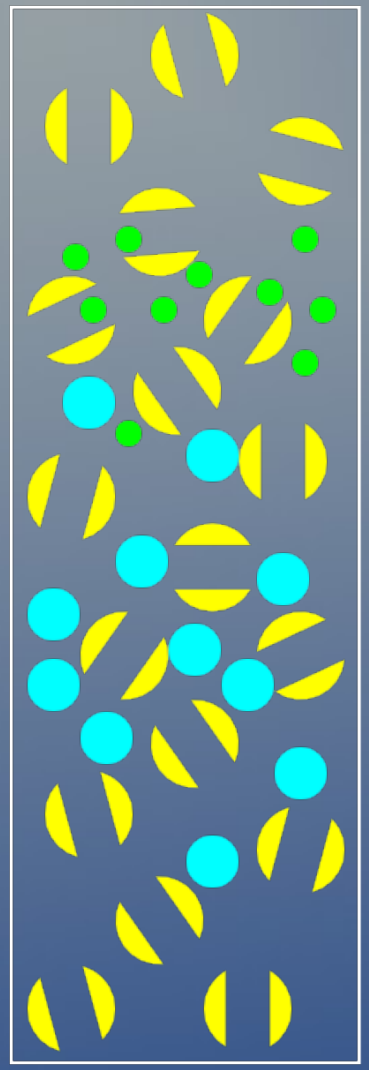
- Bead model generation from atomic structures
- Hydrodynamic parameter calculation
- SAS
  - Multiple Debye for I(q) from (explicitly hydrated) atomic structures or bead models
    - Full, Quick (FoXS), S.H., and external interface to CRYSOL/N, FoXS
    - Customizable excluded volumes, scattering factors
    - Supports common 4 term Gaussian scattering factors and 5 term
      - *Waasmaier,. D. and Kirfel, A, New analytical scattering-factor functions for free atoms and ions, Acta Cryst. (1995). A51, 416-431*
    - CUDA Full Debye
    - S.H. Intel MIC Debye (Initial testing 10k Scatters < 500ms)
  - Rg, Rc, Rt
  - P(r)
  - Best fit, NNLS fitting
  - Buffer subtraction tool
  - Batch & cluster computations
  - DMD on cluster for expanding conformation space
  - HPLC-SAXS tools

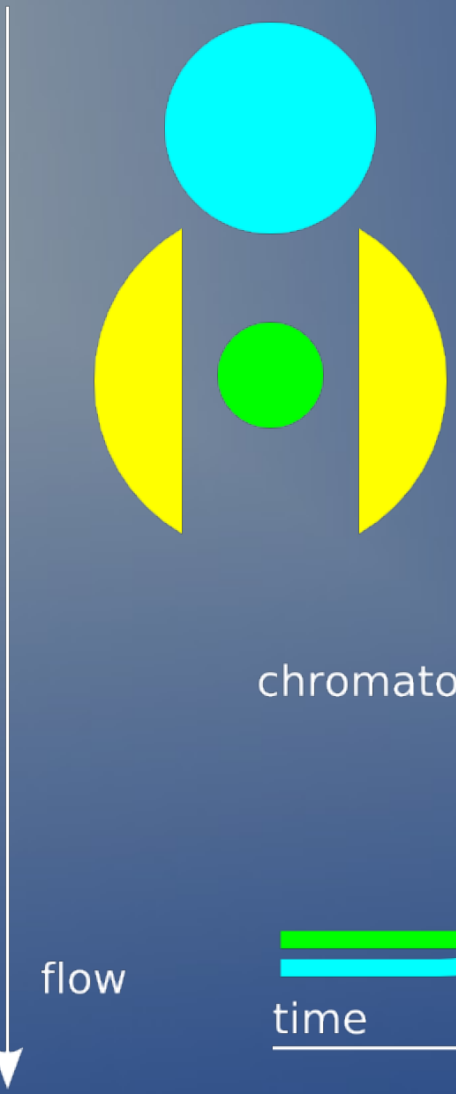
US-SOMO

- Website: somo.uthscsa.edu
- GUI based: Windows, OSX, Linux, source code (GPL).
- *Brookes et al,* UltraScan Solution Modeler: Integrated Hydrodynamic Parameter and Small Angle Scattering Computation and Fitting Tools, *ACM XSEDE 2012*
  - "The overarching goal of our software is to provide an extensible general framework for generating collections of candidate structures from an initial structure or structures, modeling candidate structures under various experimental methods and conditions, and subsequently globally fitting and screening candidate structure's models against sets of experimental data"
- *Brookes E, Demeler B, Rosano C, Rocco M.* The implementation of SOMO (SOlution MOdeller) in the UltraScan analytical ultracentrifugation data analysis suite: enhanced capabilities allow the reliable hydrodynamic modeling of virtually any kind of biomacromolecule. *Eur Biophys J. 2010 Feb;39(3):423-35.*
- *Brookes E, Demeler B, Rocco M.* Developments in the US-SOMO bead modeling suite: new features in the direct residue-to-bead method, improved grid routines, and influence of accessible surface area screening. *Macromol Biosci. 2010 Jul 7;10(7):746-53.*

HPLC-SAXS

SEC

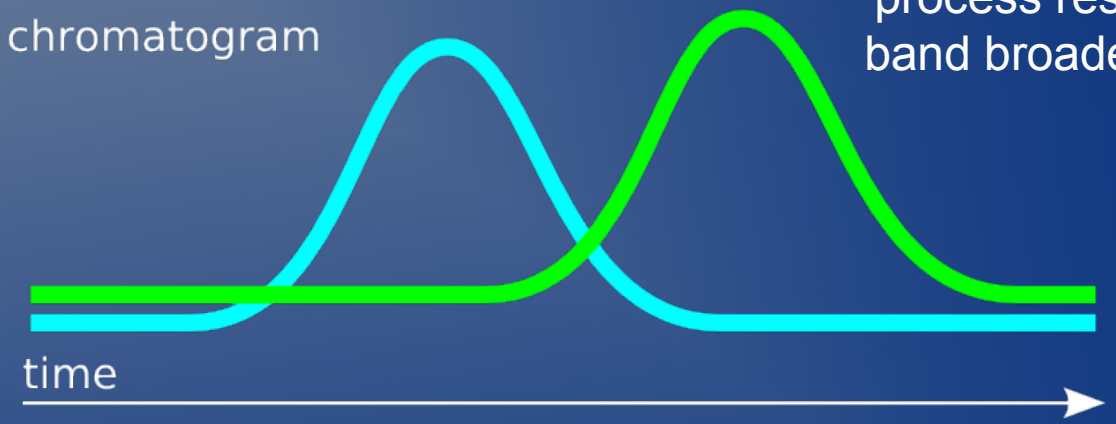




Large particles can not enter the packing material pores and are excluded. The have less volume to traverse and are excluded sooner.

Small particles can enter the packing material pores and have more volume to traverse. They elute later.

Typically Gaussian shaped profiles since diffusion is the primary physical process responsible for band broadening during separation

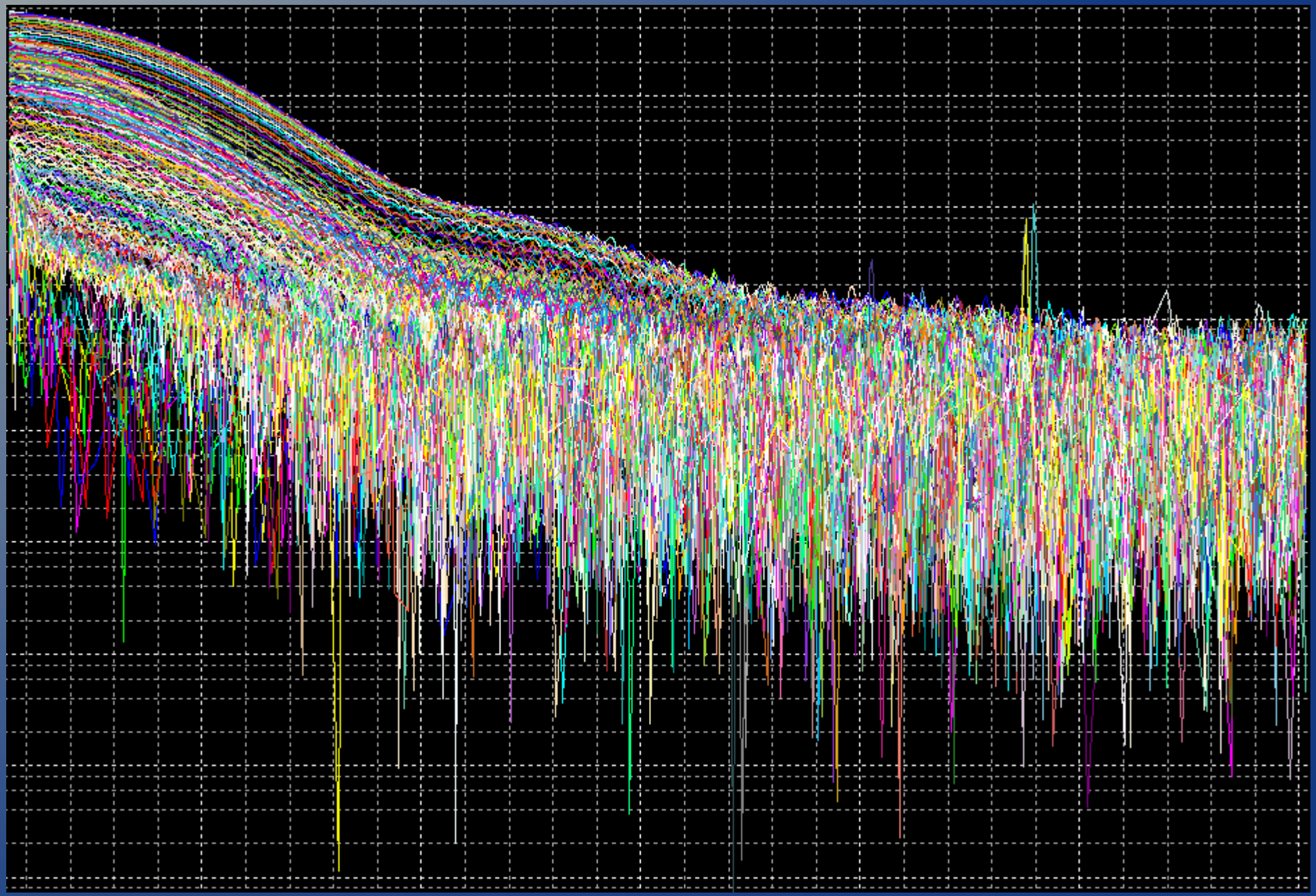


If not Gaussian, it could be:
- Interaction with the column matrix
- Superposition of multiple peaks
- Interaction between co-eluting species
- Weak self associations

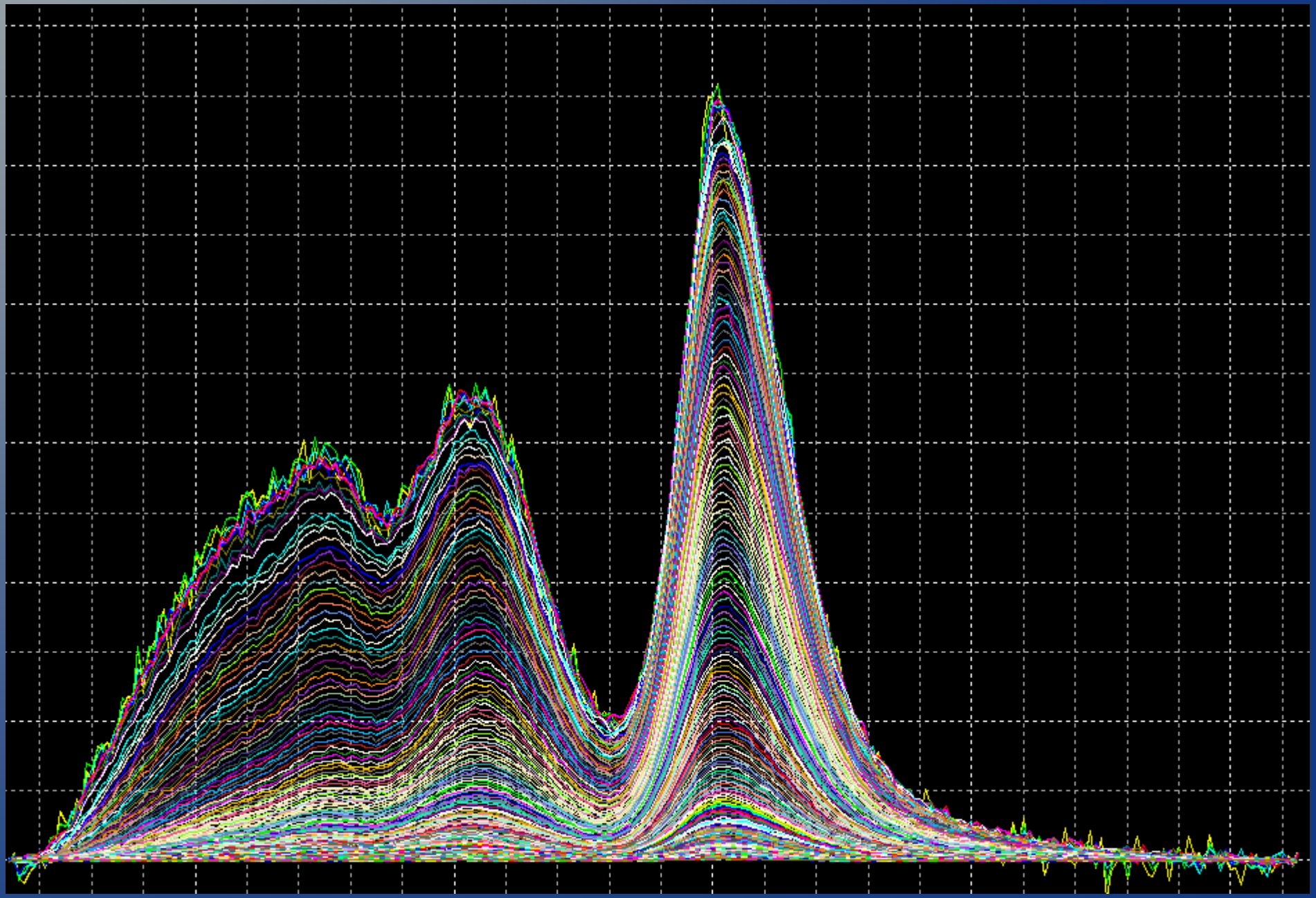
Online SEC-SAXS

- Separate immediately before measuring
- Individual peaks are more likely to be monodisperse
- First use paper, available to users who could self-manage FPLC
  - ID-18 BioCat / APS
  - *Mathew, E., Mirza, A., & Menhart, N.* **2004**. *Liquid-chromatography-coupled SAXS for accurate sizing of aggregating proteins. J. Synchrotron Rad. 11, 314-318.*
- First setup with user HPLC support
  - SWING/SOLEIL
  - *David, G. & Pérez, J.* **2009**. *Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. J. Appl. Cryst. 42, 892-900*
- Now available as primary method of analysis at multiple beamlines

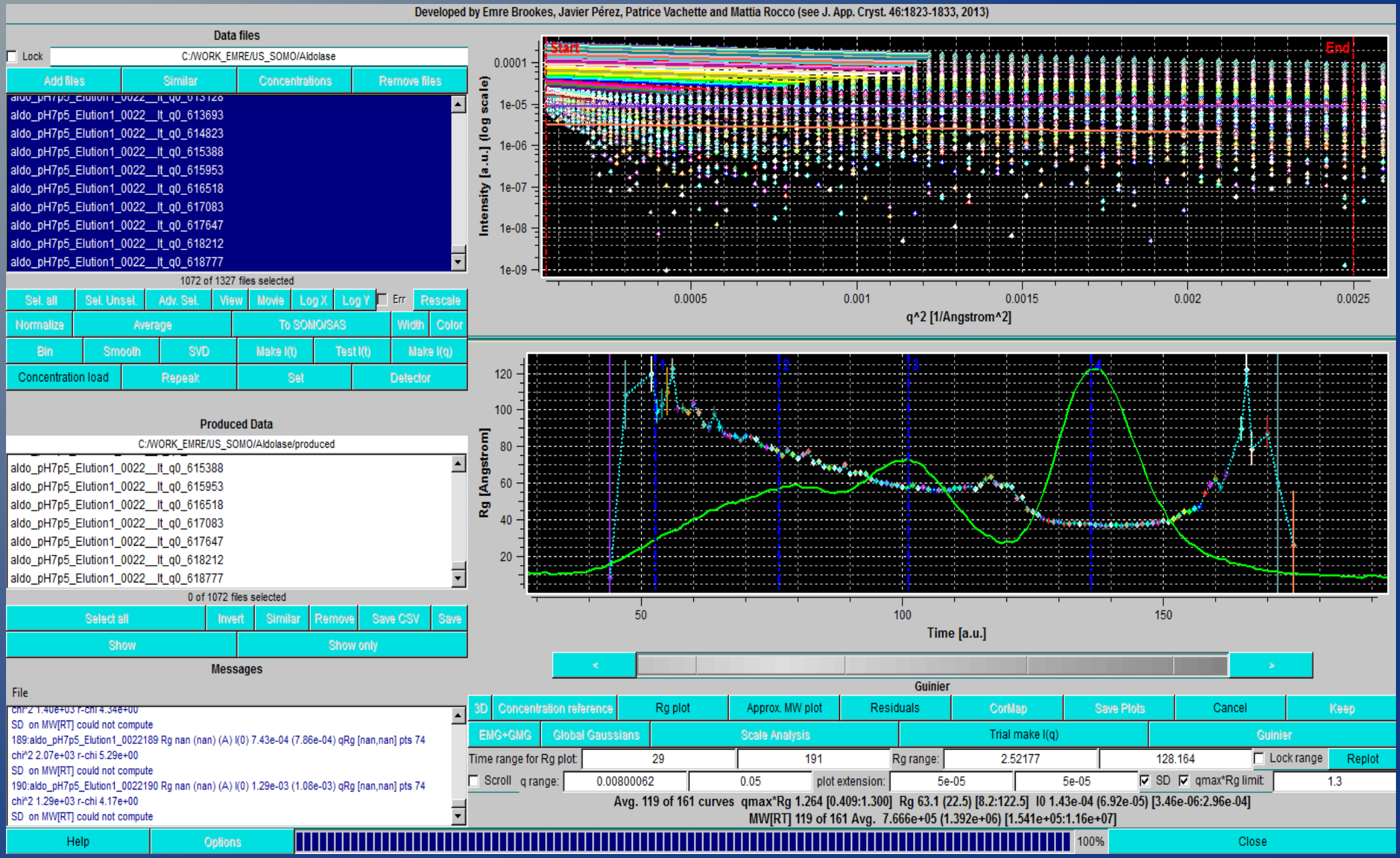
Set of I(q), each a specific time or frame, Aldolase



Set of I(t), each a specific q, Aldolase



US-SOMO / HPLC-SAXS / Aldolase / Test I(q) global



Gaussian Decomposition Procedure

- Optionally use SVD to create a Truncated SVD set and/or determine expected number of Gaussians (eq. species)
- Fit 1 typical I(t) with Gaussians or skewed Gaussians (EMG, GMG, EMG+GMG currently supported)
- Globally fit Gaussians to all I(t)
    - Global fit applies physical constraints
        - Common centers
        - Common width
        - Common skew (for skewed Gaussians)
    - Fit uses all available data and considers physical constraints

US-SOMO / HPLC-SAXS / Aldolase / SVD analysis

Data files

Source

Original data
- I(q)
- I(t)

Clear | Replot

To HPLC window | Color

Process

Compute SVD | Stop

Singular value list:

0.00894405	2.29656e-05	1.6003e-05	1.2561e-05	1.02547e-05	8.30653e-06	6.66662e-06
0.00160828	2.26205e-05	1.59418e-05	1.24927e-05	1.02044e-05	8.28492e-06	6.60606e-06

Plot SVs | Save SVs | TSVD reconstruction

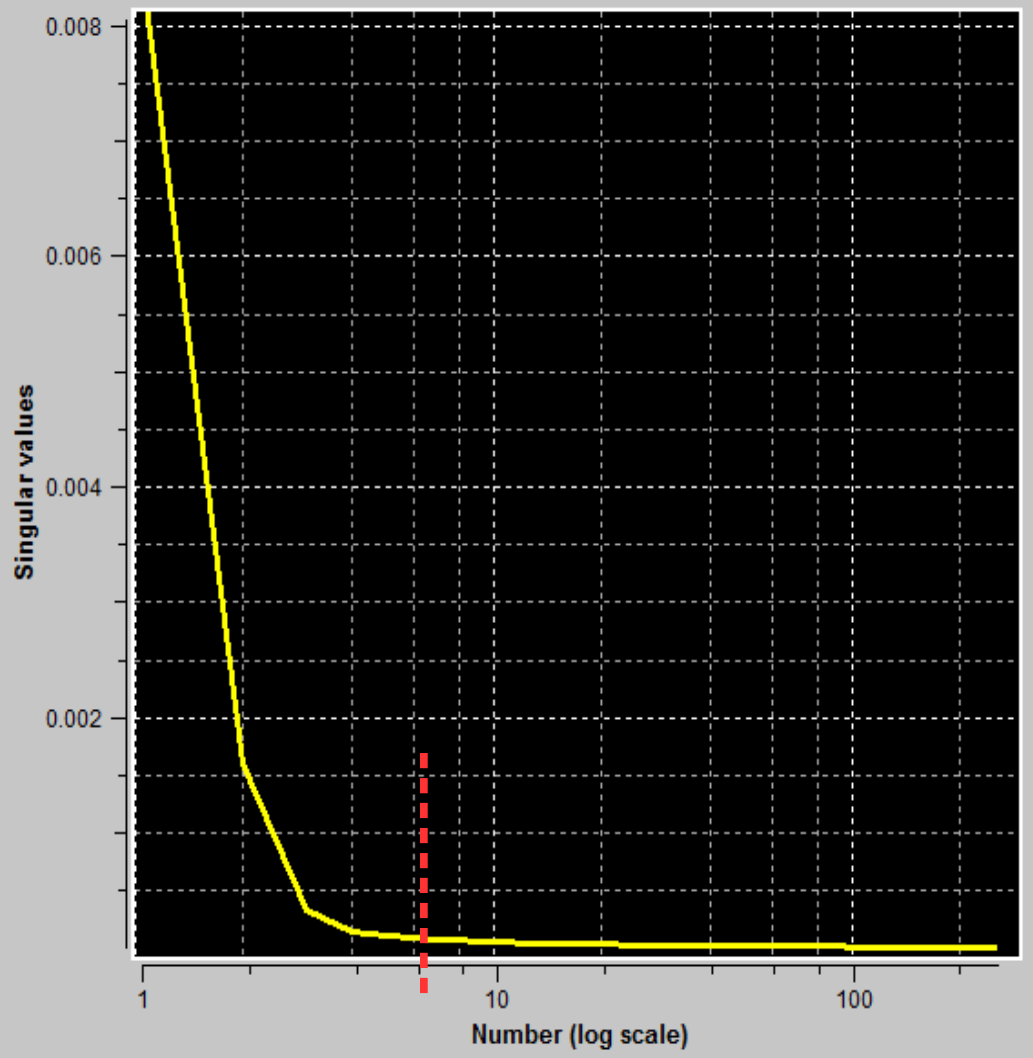
Individual TVSD recon. | Incremental TVSD recon.

Plot RMSDs | Save RMSDs

Messages

File

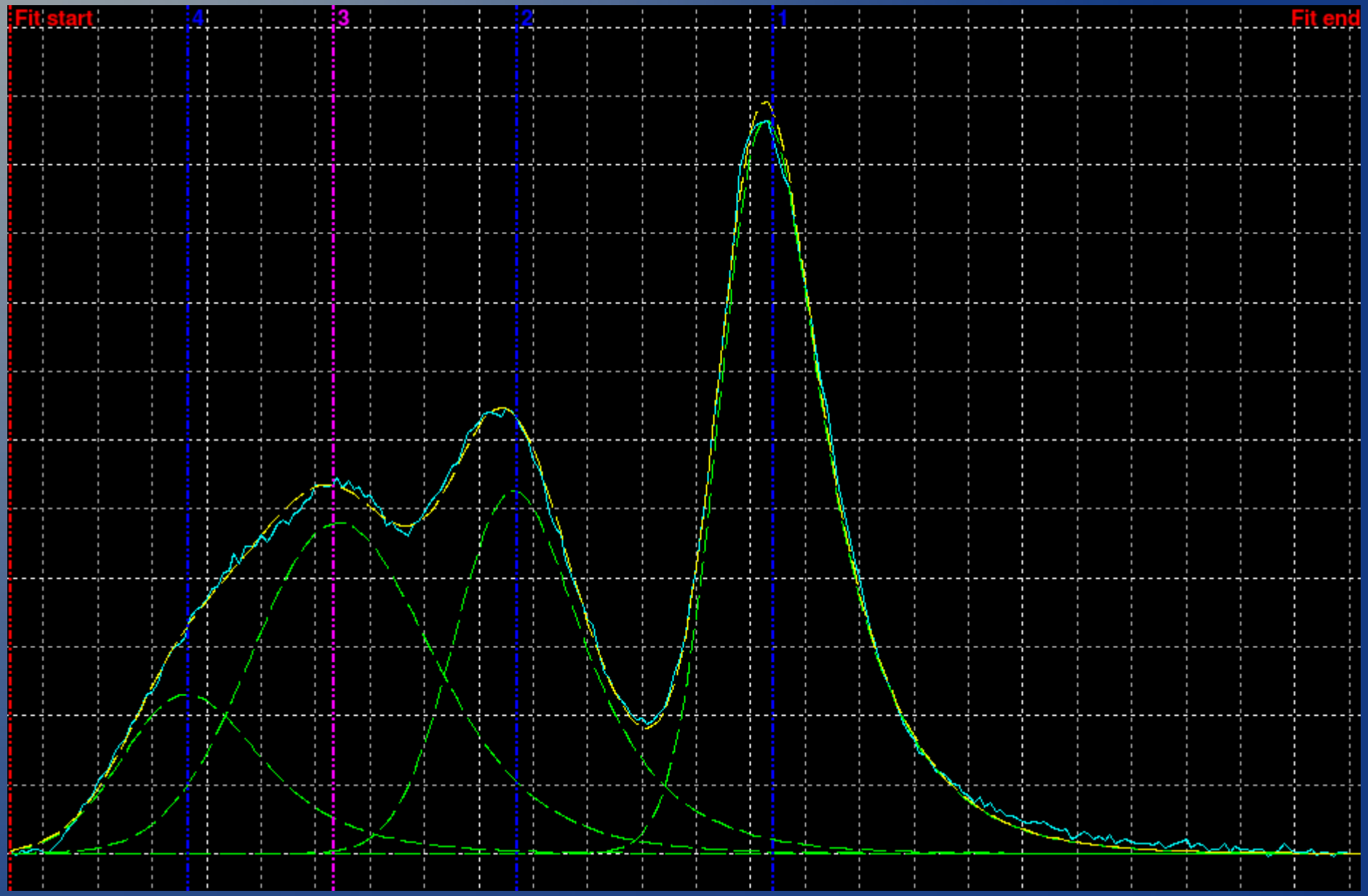
Making I(t) for source Original data
Done making I(t) for source Original data
SVD: matrix F created, computing SVD



Show I(t) | X | Y

Help | Close

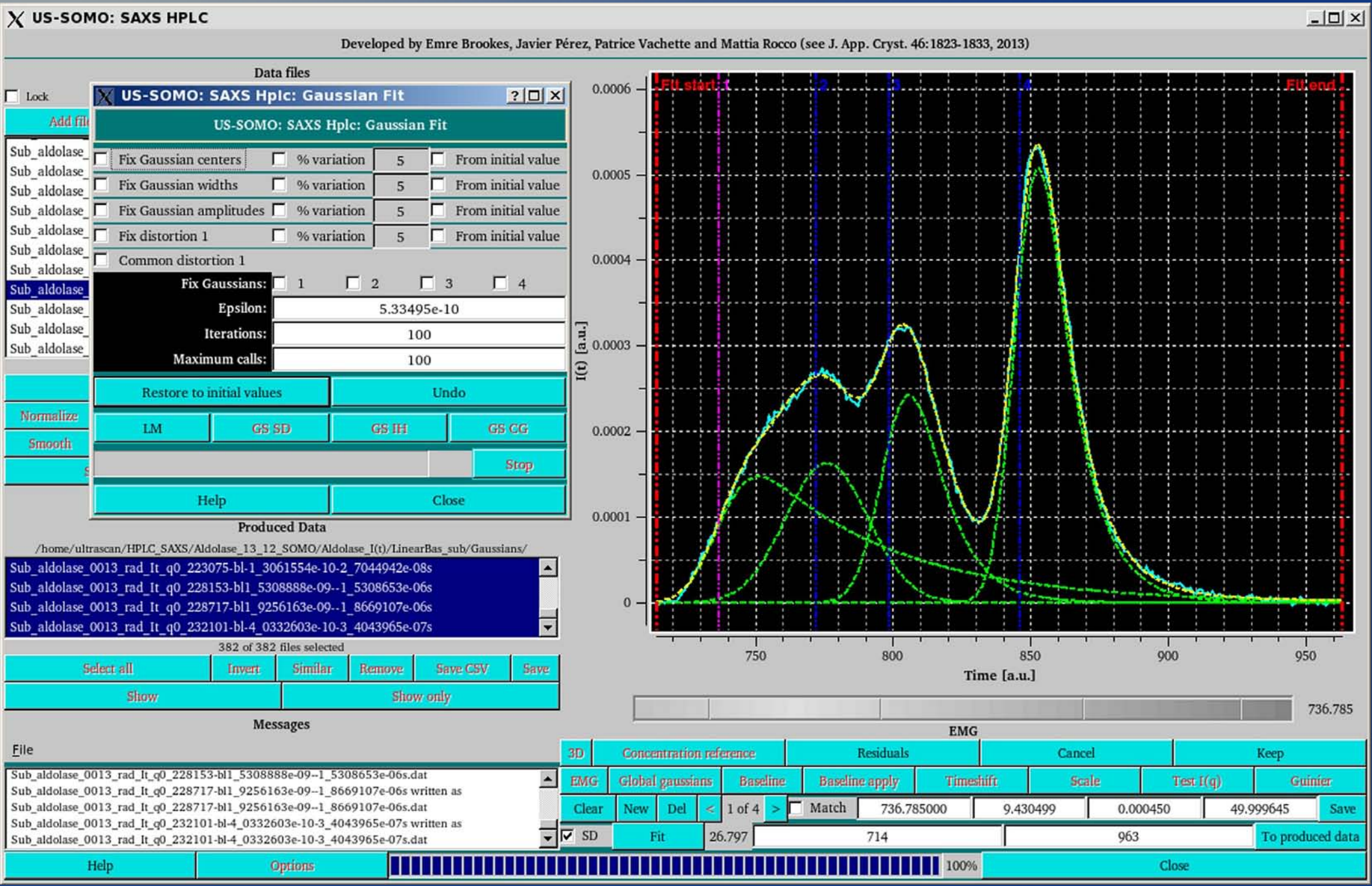
Set of I(t), each a specific q



Intensity (linear)

time or frame (linear)

US-SOMO / HPLC-SAXS / Aldolase / Why common distortion?



Gaussian functions with distortions

- Exponentially modified Gaussian (EMG):

$$y = \frac{a_0}{2a_3}\exp\left(\frac{a_2^{\,2}}{2a_3^{\,2}} + \frac{a_1 - x}{a_3}\right)\left[\operatorname{erf}\left(\frac{x-a_1}{\sqrt{2}\,a_2} - \frac{a_2}{\sqrt{2}\,a_3}\right) + \frac{a_3}{|a_3|}\right]$$

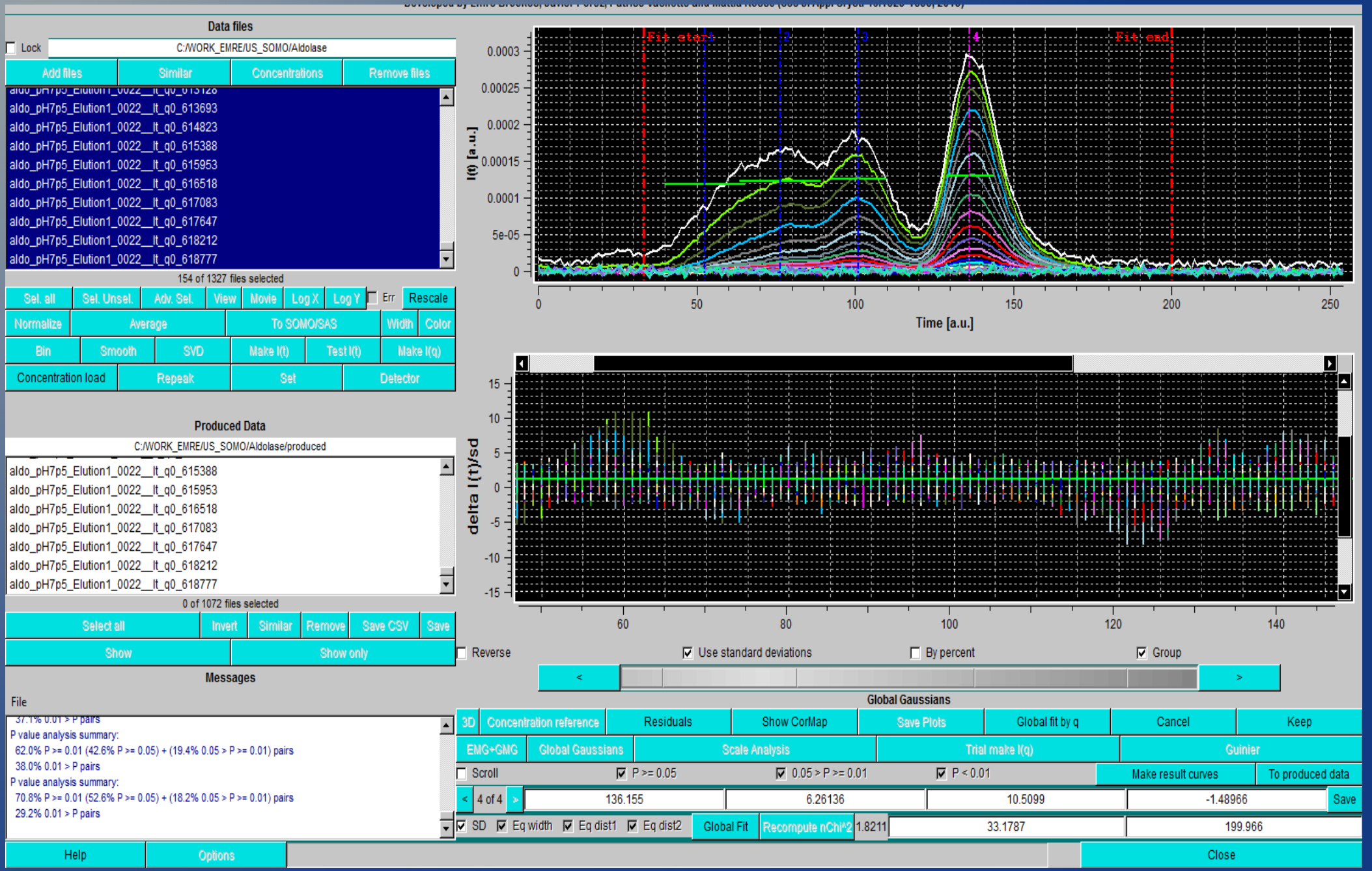
- Half-Gaussian modified Gaussian (GMG):

$$y = \frac{a_0 \exp\left(-\frac{1}{2}\frac{(x-a_1)^2}{a_3^{\,2}+a_2^{\,2}}\right)\left[1+\operatorname{erf}\left(\frac{a_3(x-a_1)}{\sqrt{2}a_2\sqrt{a_3^{\,2}+a_2^{\,2}}}\right)\right]}{\sqrt{2\pi}\sqrt{a_3^{\,2}+a_2^{\,2}}}$$

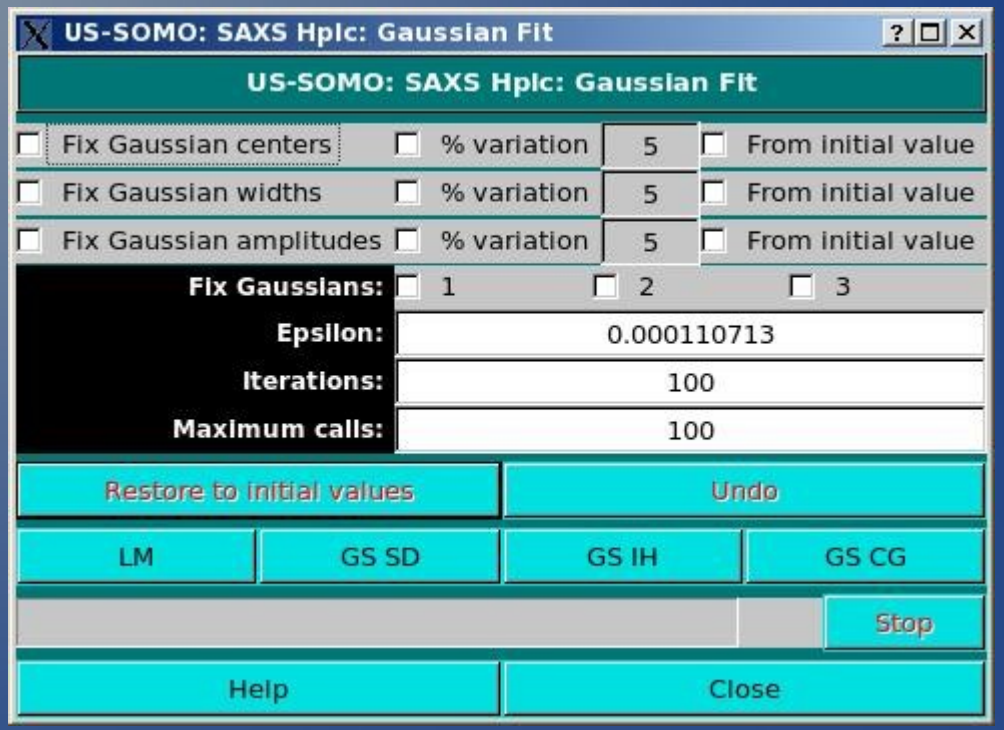
- EMG+GMG:

$$y = \frac{a_0}{4a_3}\exp\left(\frac{2a_1a_3 - 2a_3x + a_2^{\,2}}{a_3^{\,2}}\right)\operatorname{erfc}\left(\frac{a_1a_3 - a_3x + a_2^{\,2}}{\sqrt{2}a_2a_3}\right) + \frac{a_0}{2\sqrt{2\pi}\sqrt{a_2^{\,2}+a_4^{\,2}}}\exp\left(-\frac{1}{2}\frac{(a_1-x)^2}{a_2^{\,2}+a_4^{\,2}}\right)\operatorname{erfc}\left(\frac{a_4(a_1-x)}{\sqrt{2}a_2\sqrt{a_2^{\,2}+a_4^{\,2}}}\right)$$

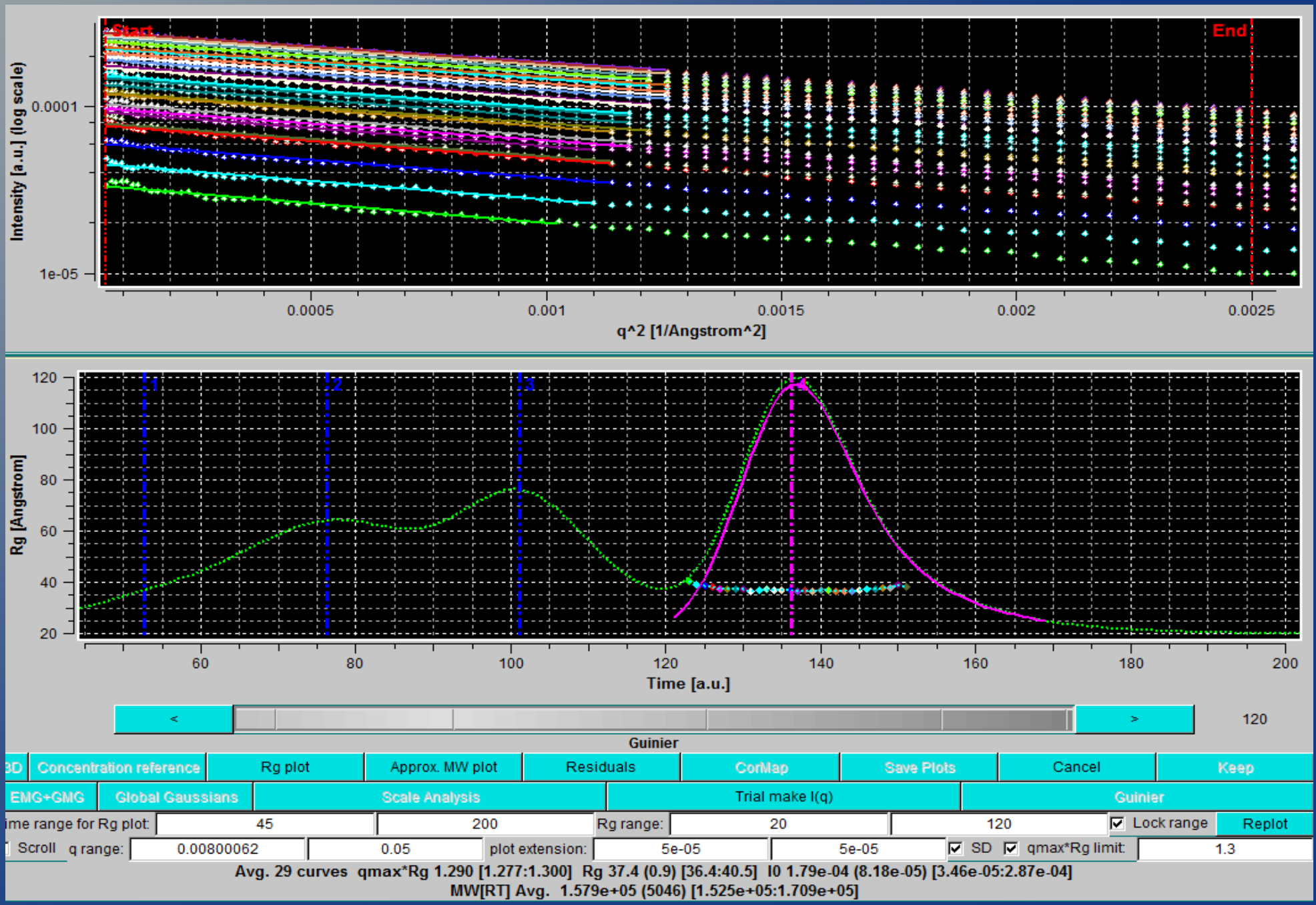
US-SOMO / HPLC-SAXS / Aldolase / Final global EMG+GMG fitting



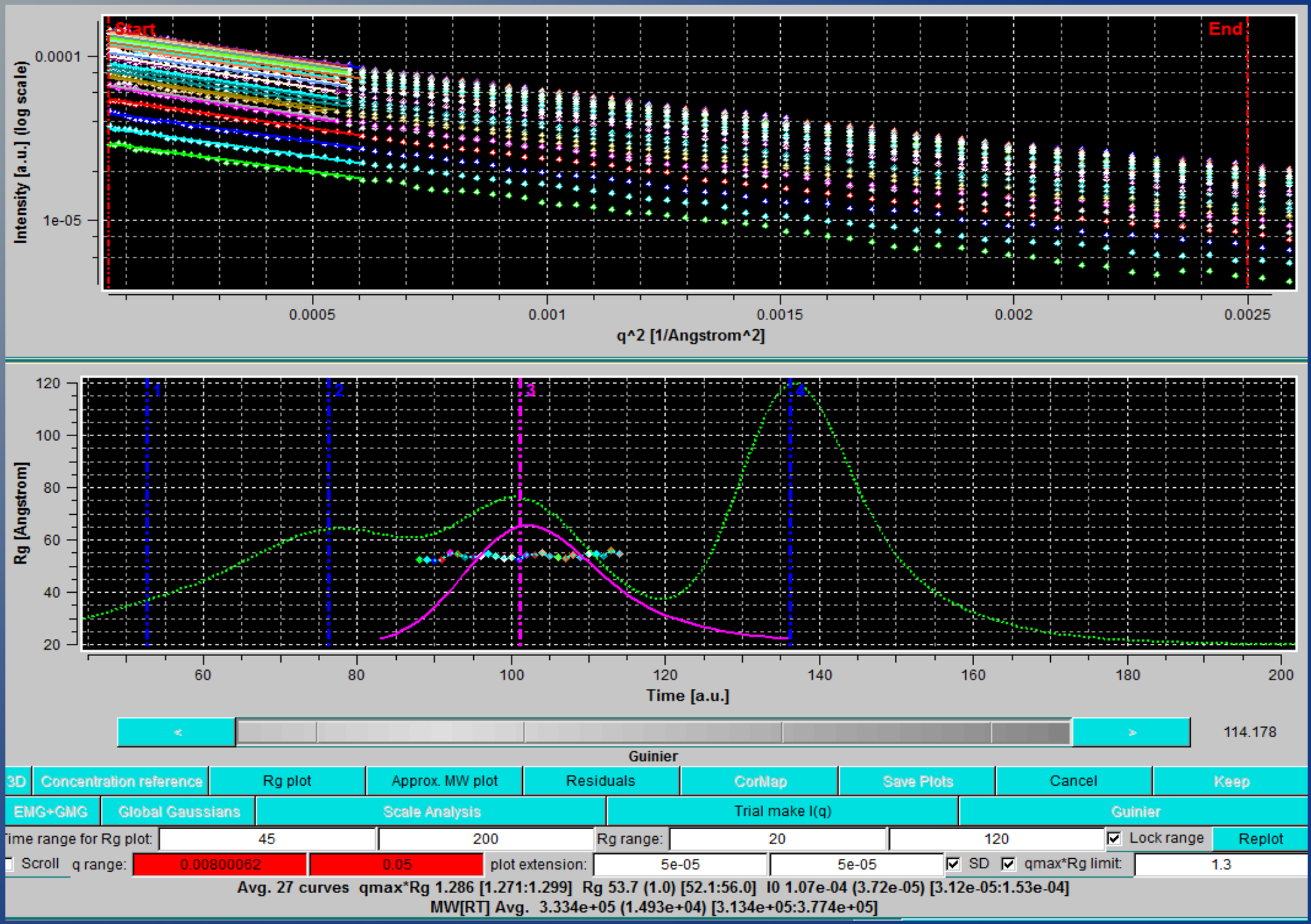
US-SOMO / HPLC-SAXS / Gaussian fitting



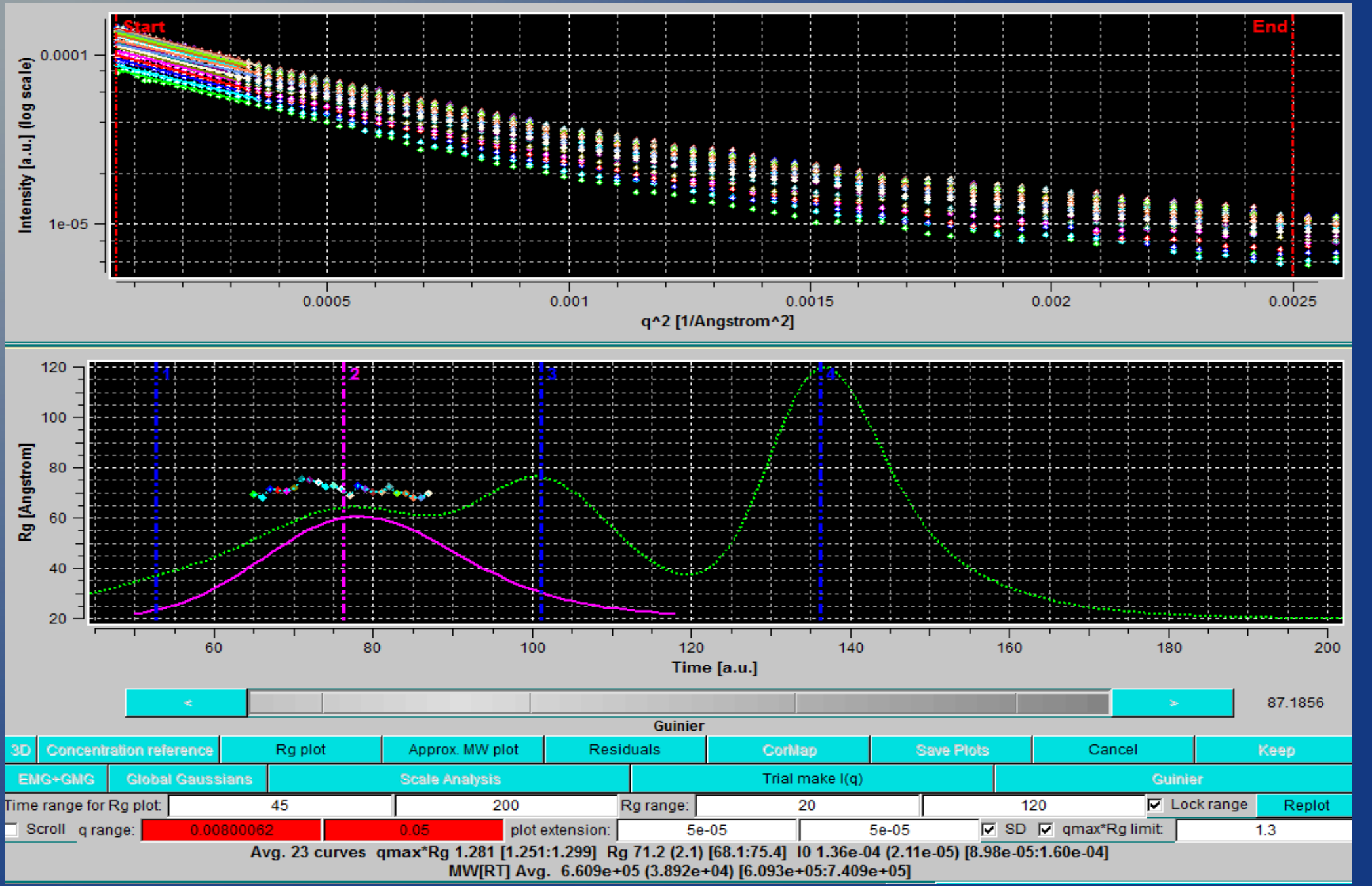
US-SOMO / HPLC-SAXS / Aldolase / Test I(q) 4th peak



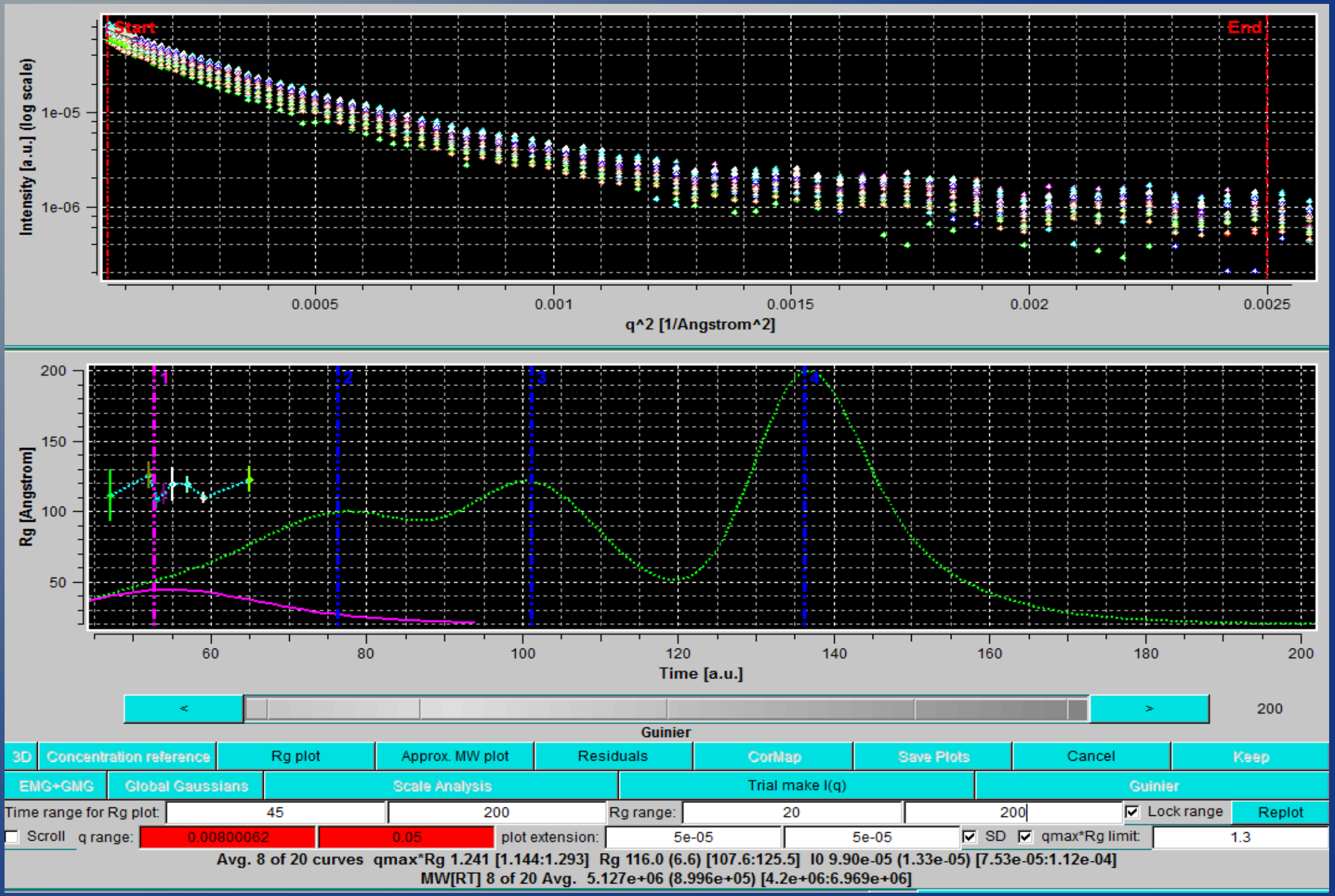
US-SOMO / HPLC-SAXS / Aldolase / Test I(q) 3rd peak

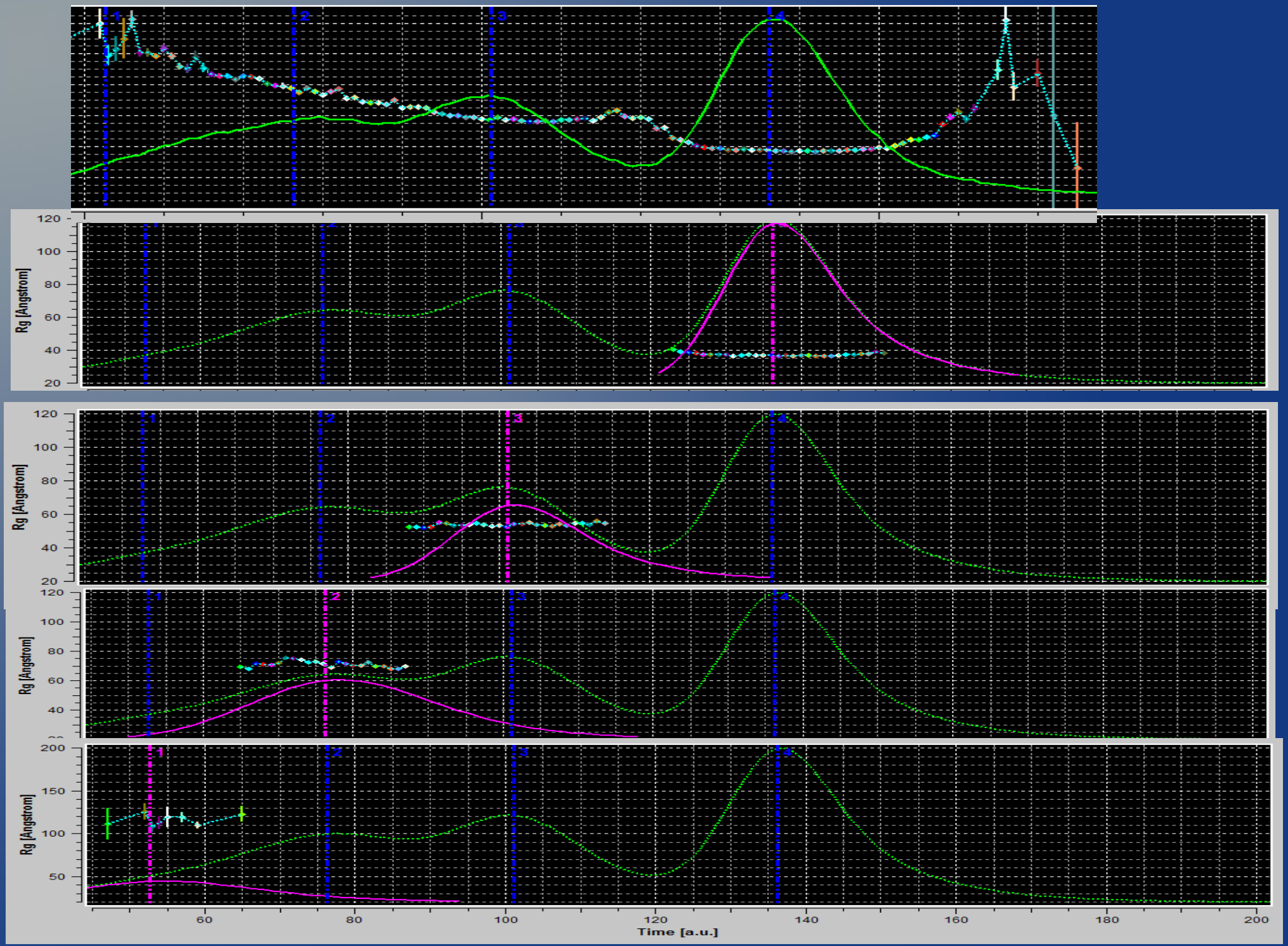


US-SOMO / HPLC-SAXS / Aldolase / Test I(q) 2nd peak

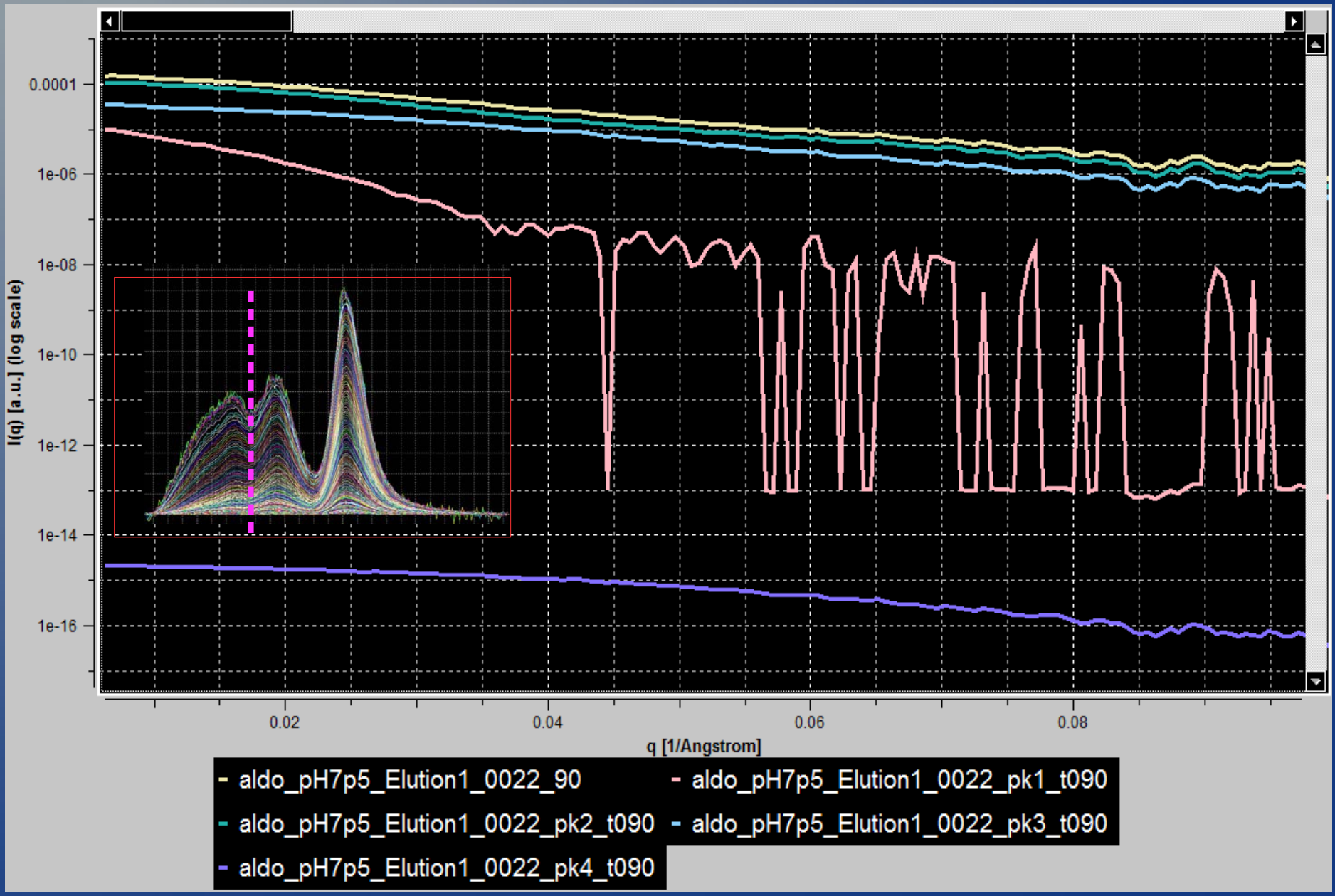


US-SOMO / HPLC-SAXS / Aldolase / Test I(q) 1st peak

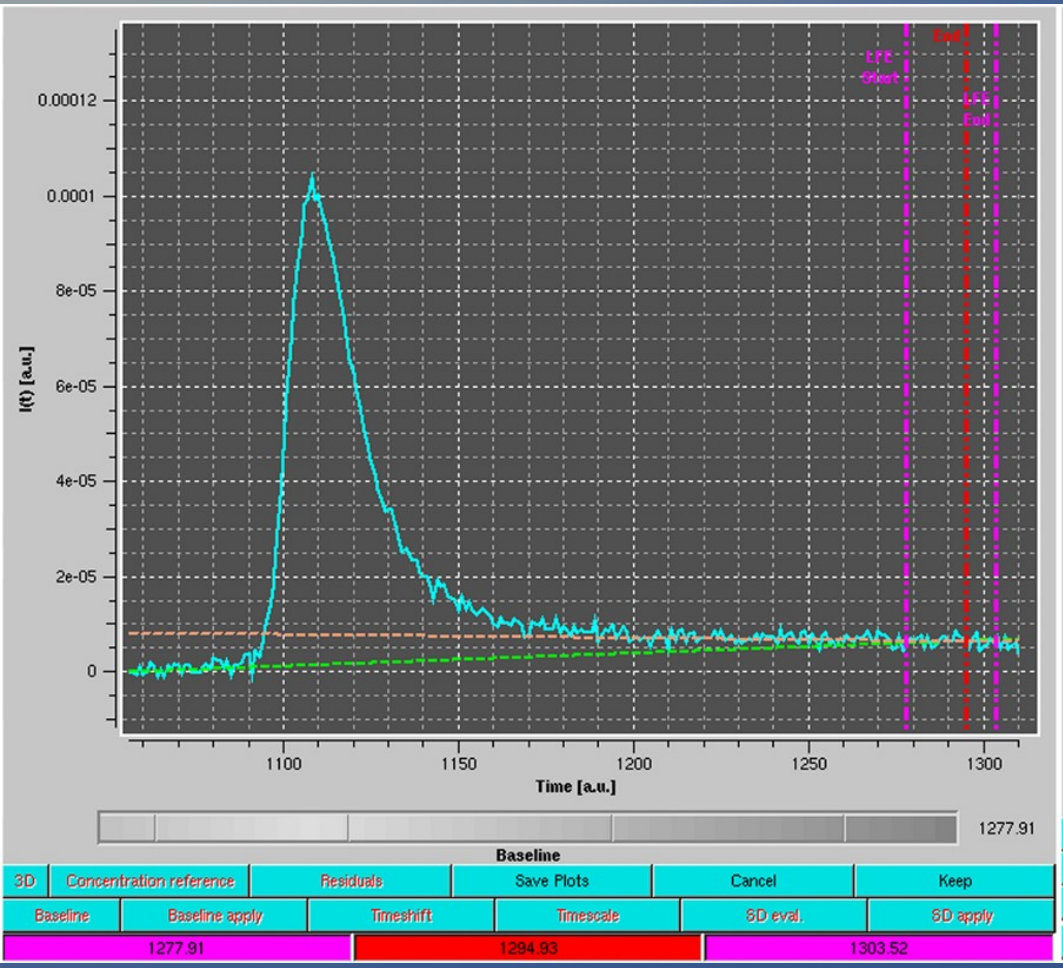


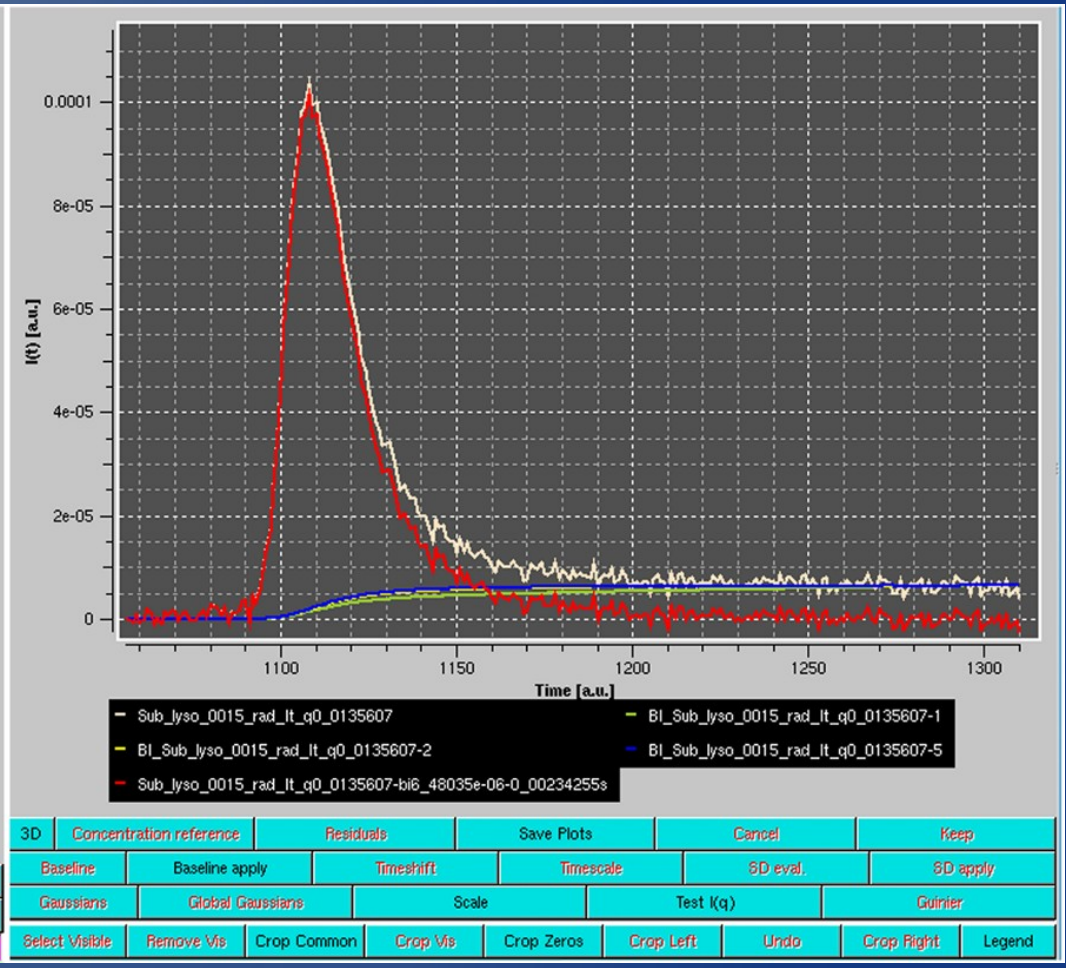

Rg [Angstrom]
Time [a.u.]

US-SOMO / HPLC-SAXS / Aldolase
Frame #90: decomposition



Physically based capillary fouling removal





- Better to not have fouling
- If you do, this may help

Physically based capillary fouling removal

$$I_{BL}(q) = \sum_{k=1,\, m} \frac{I(q,t_{sk})}{m}$$

1. Set the initial baseline to zero: $B_0(q,t) = 0$

2. Loop $i = 0, \ldots,$ maximum-iterations

3. Compute the total intensity above the baseline:

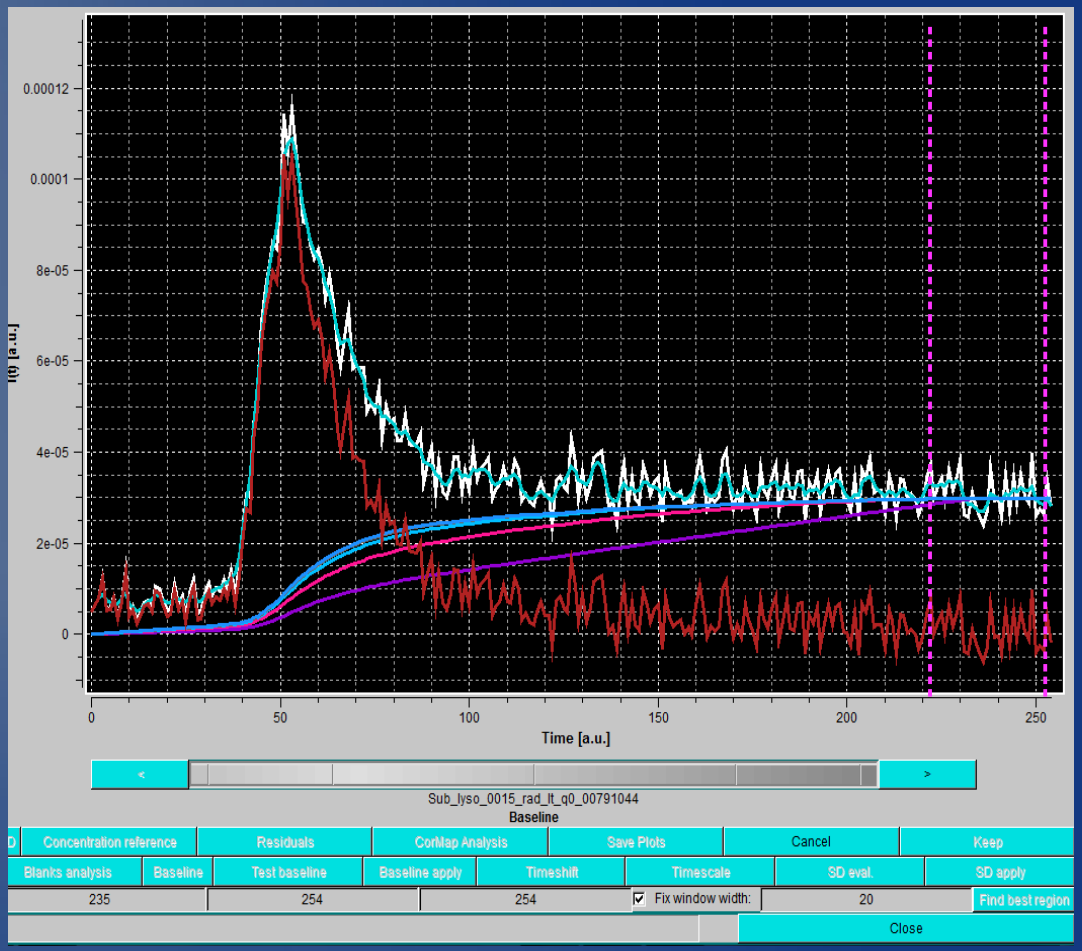
Let $I_{TOT_i}(q) = \sum_{k=1}^{n-m} I(q,t_k) - B_i(q,t_k)$

4. Compute $\alpha_i(q) = \dfrac{I_{BL}(q)}{I_{TOT_i}(q)}$

5. if $\left|\alpha_i(q) - \alpha_{i-1}(q)\right| < \varepsilon$, terminate early

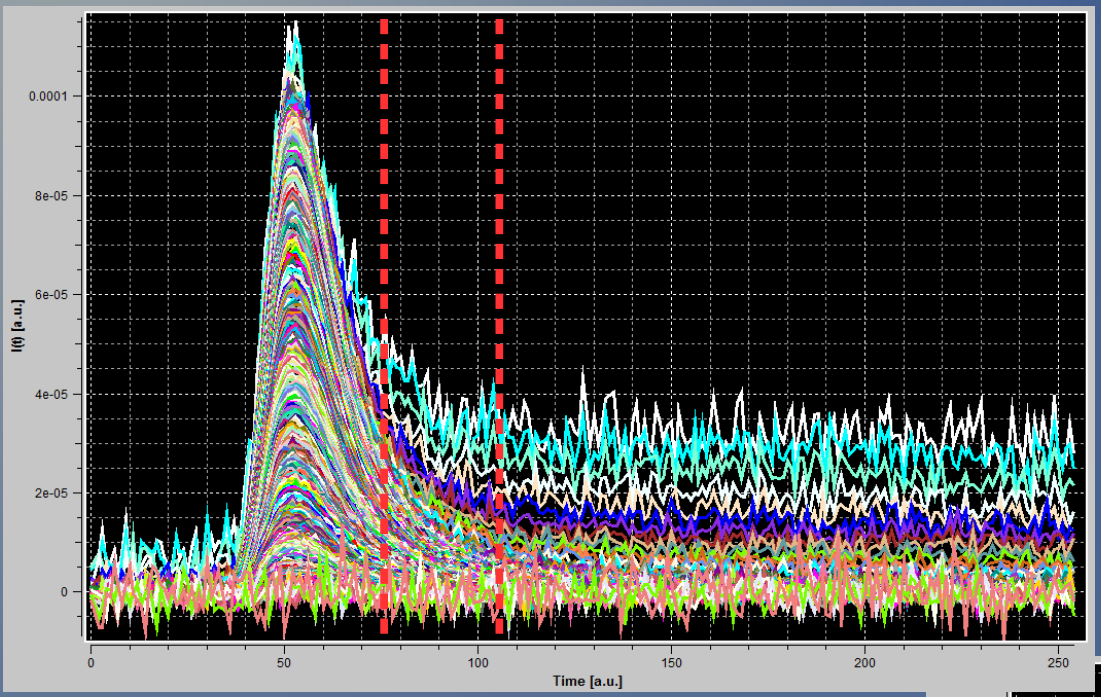
6. $D_i(q,t) = \alpha_i(q)\left[I(q,t) - B_i(q,t)\right]$

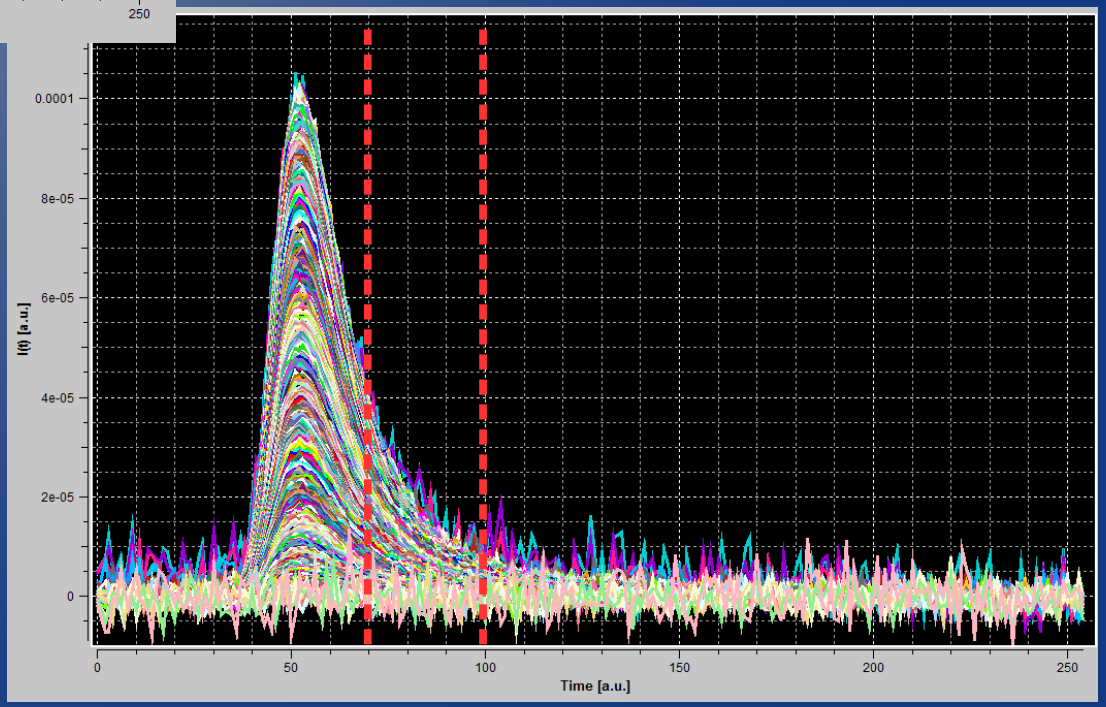
7. $B_{i+1}(q,t) = \sum_{t'=1}^{t} D_i(q,t')$



Integral Baseline Correction

Physically based capillary fouling removal



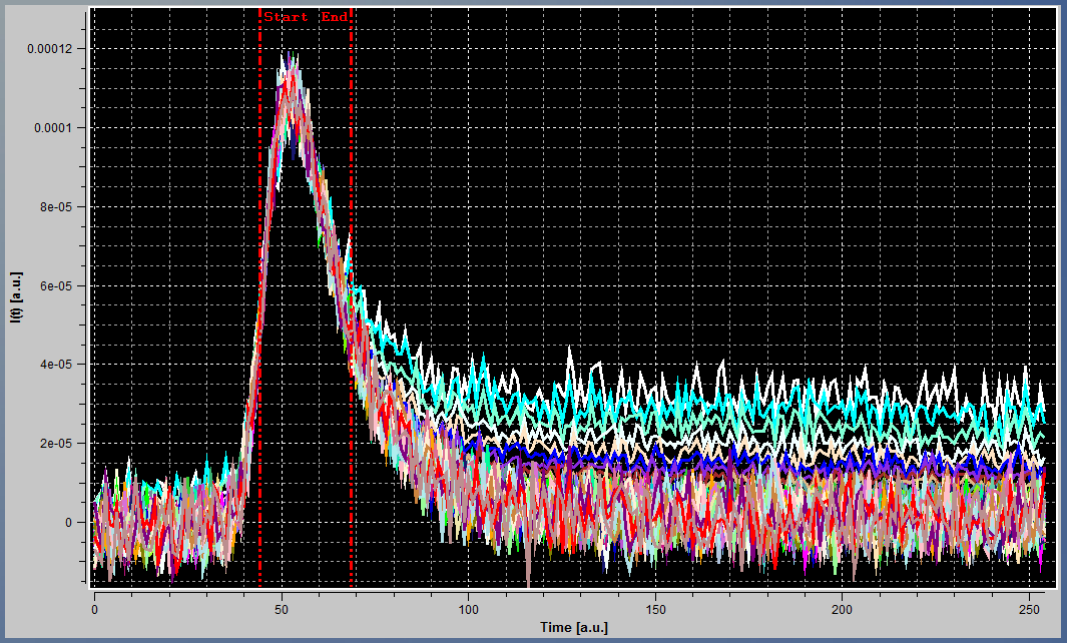


Effect of capillary fouling removal

Original data
(frames 50-110)

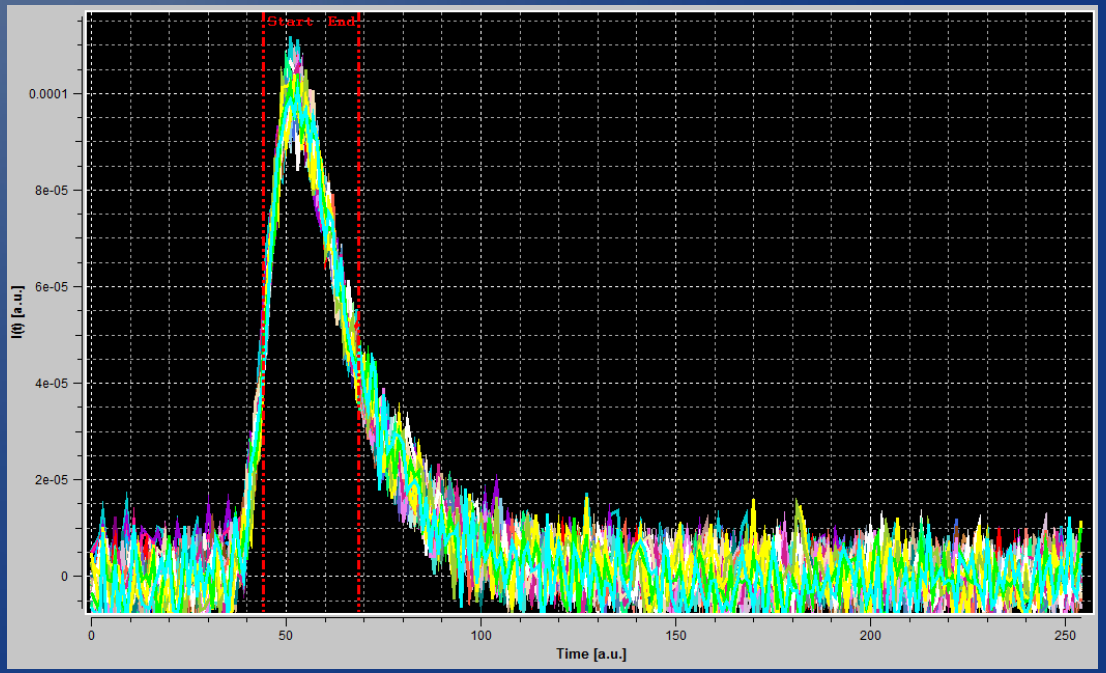
Baseline
corrected
Data
(frames 50-110)

Effect of capillary fouling removal: scaling



Original data
($q < 0.2\ A^{-1}$)
scaling to MAX value in the range of [45-65]

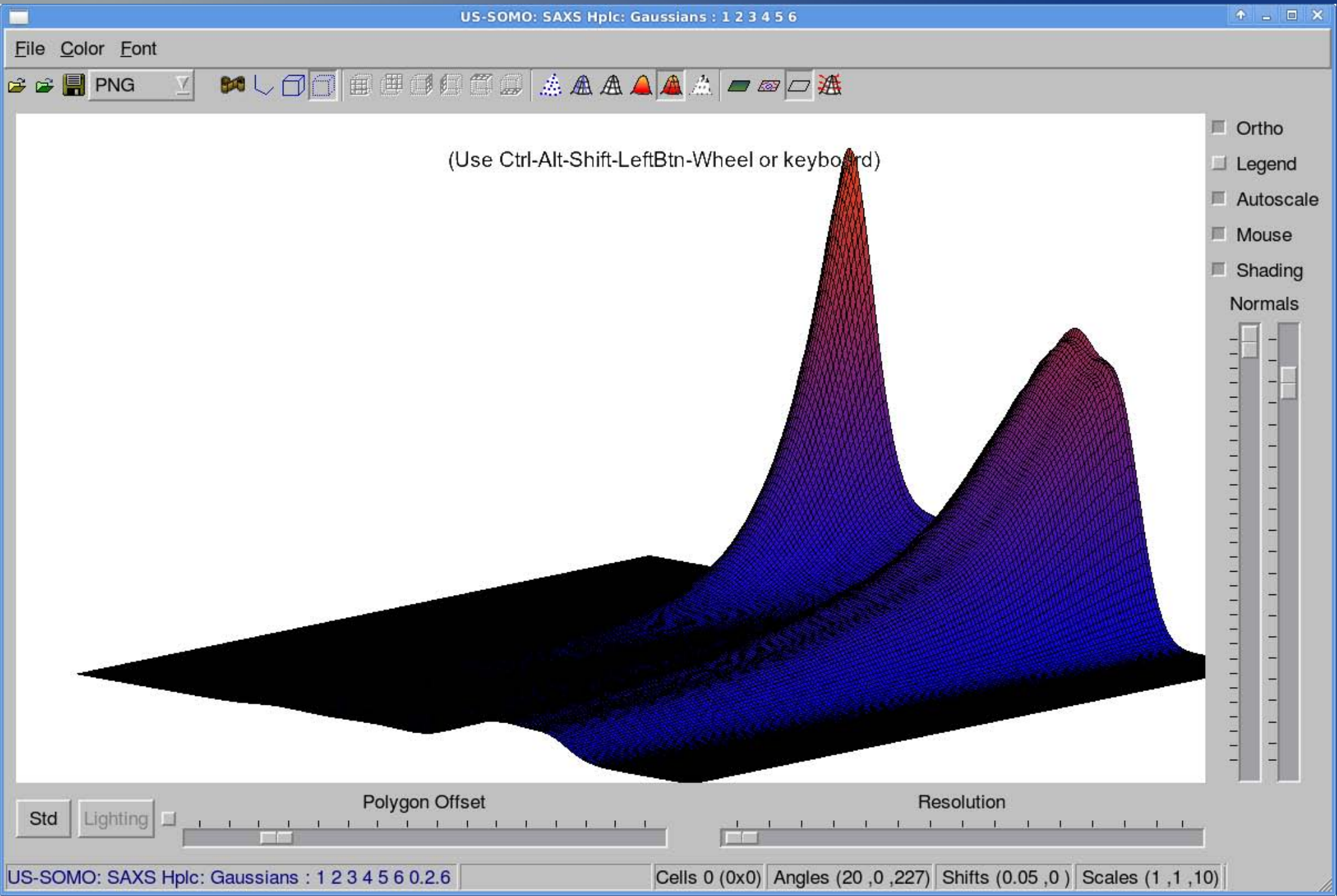
BL-corrected data
($q < 0.2\ A^{-1}$)
scaling to MAX value in the range of [45-65]



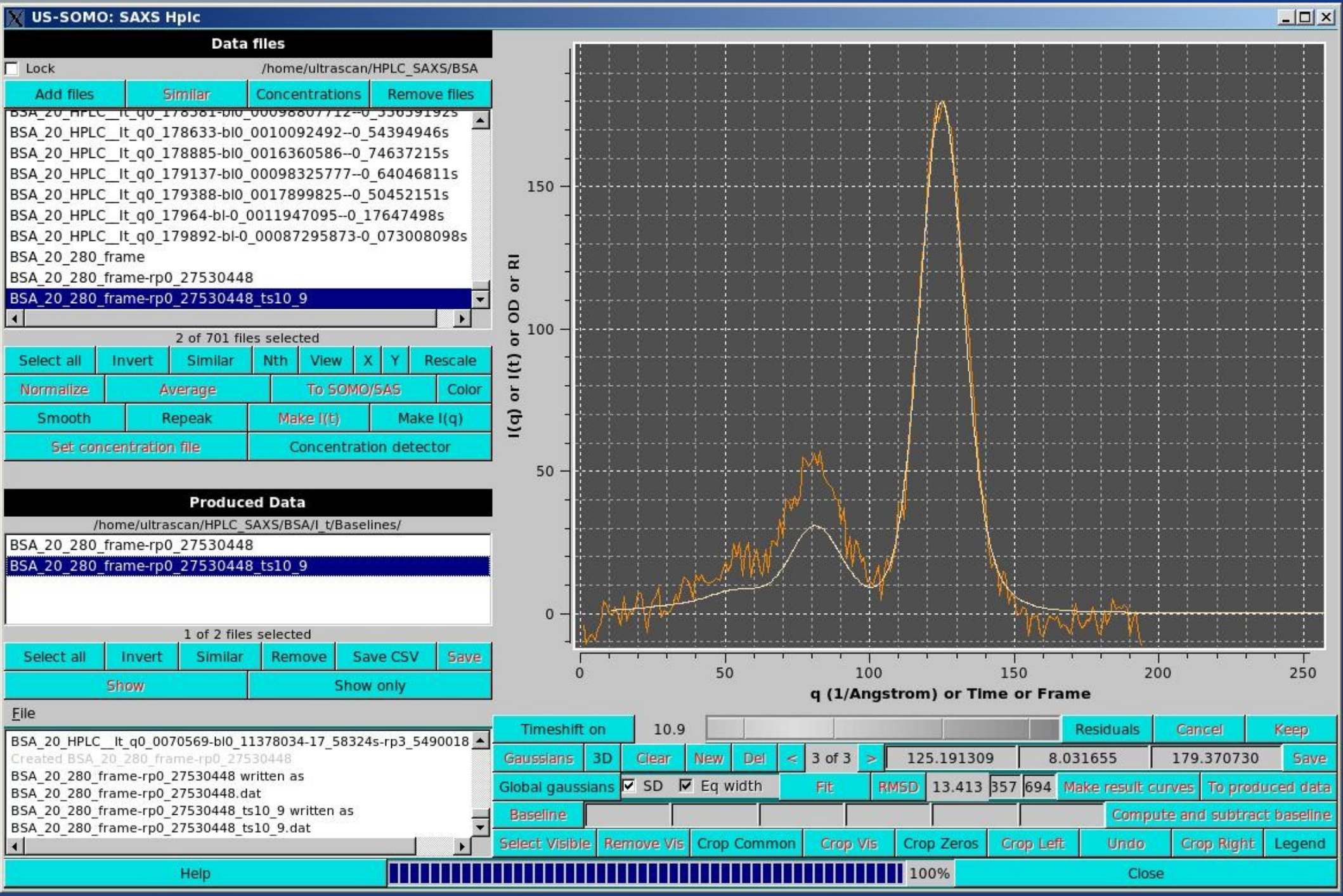
Takeaways

- Use SEC-SAXS for biological macromolecules
- If you have true baseline separation, excellent, you should probably be ok simply taking the peak data, but global Gaussian decomposition will use all of the data
- If you do not have true baseline separation, be very careful and you should use these techniques

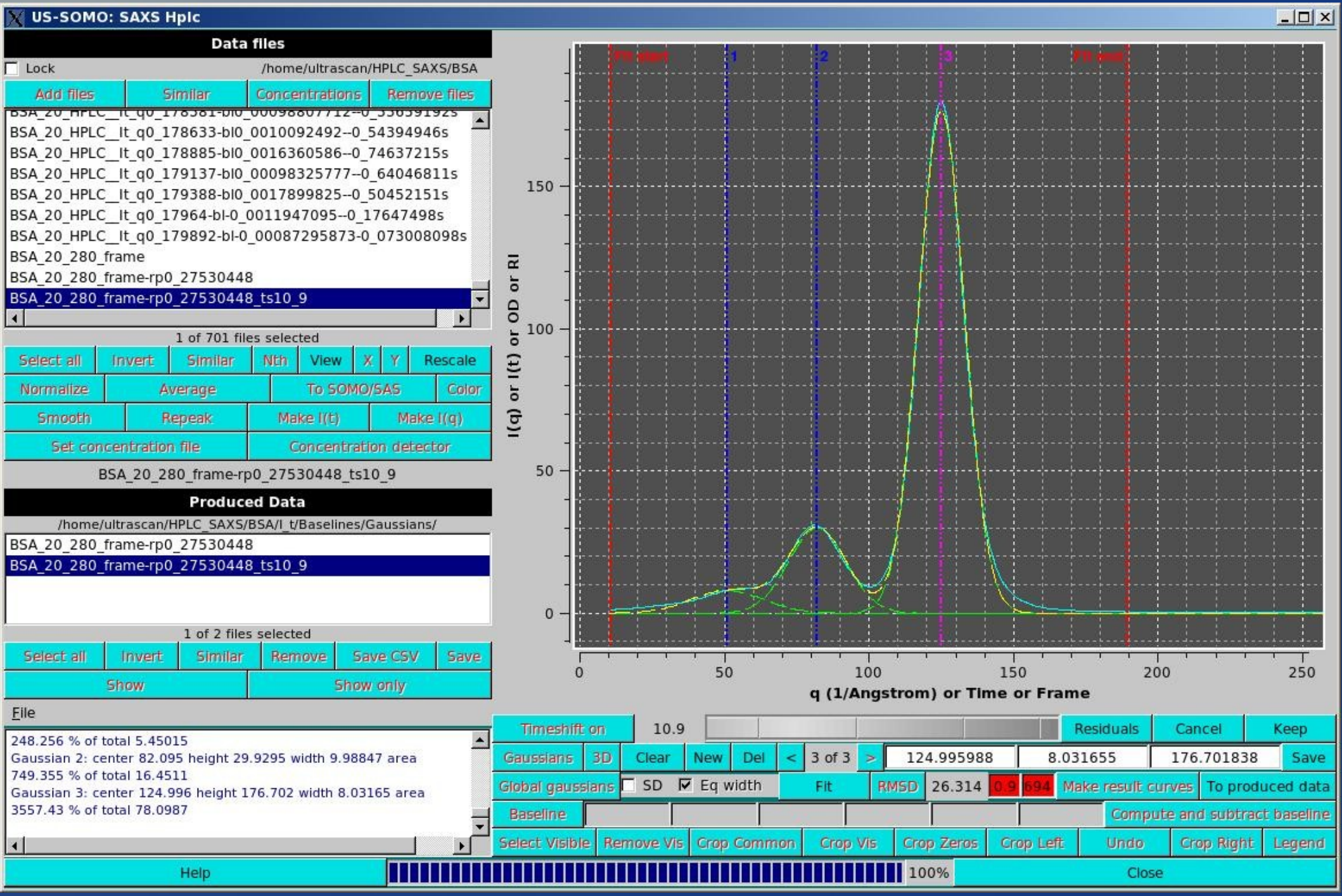
US-SOMO / HPLC-SAXS / 3D View



US-SOMO / HPLC-SAXS / Concentration curve



US-SOMO / HPLC-SAXS / Concentration curve



MW Computations

- Parameters to convert UV or RI signals to concentration
  - Detector calibration constants
- Simultaneously fit Gaussians to concentration curve
- Each Gaussian is the curve of a specific species, so each Gaussian could conceivable have its own extinction coefficient and psv

$I_{\mathrm{exp}}^{\mathrm{std}}(0)$ is the $I(0)$ of a standard (reference) scatterer (a.u.)

$I_{\mathrm{abs}}^{\mathrm{std}}(0)$ is the theoretical value of $I(0)$ for the standard scatterer (cm^{-1})

$I_{\mathrm{exp}}(0)$ is experimental $I(0)$ (a.u.)

$$I_{\mathrm{abs}}(0) = \frac{I_{\mathrm{exp}}(0)\, I_{\mathrm{abs}}^{\mathrm{std}}(0)}{I_{\mathrm{exp}}^{\mathrm{std}}(0)} \ (\mathrm{cm}^{-1})$$

- c concentration (mg/cm^3)
- R_e is the diffusion length of the electron (cm)
- Z is avg number of electrons per atom, A is avg # of nucleons
- m_n is the mass of a nucleon (g)
- v_2 is psv (cm^3/g)
- ρ_e is solvent e density (e/A^3)

$$I^*(0) = \frac{I_{\mathrm{abs}}(0)\, N_A}{\frac{c}{1000} R_e^2 \left[\frac{Z}{A}\frac{1}{m_n} - \bar{v}_2\, \rho_e 10^{24}\right]^2} \ (\mathrm{Da})$$

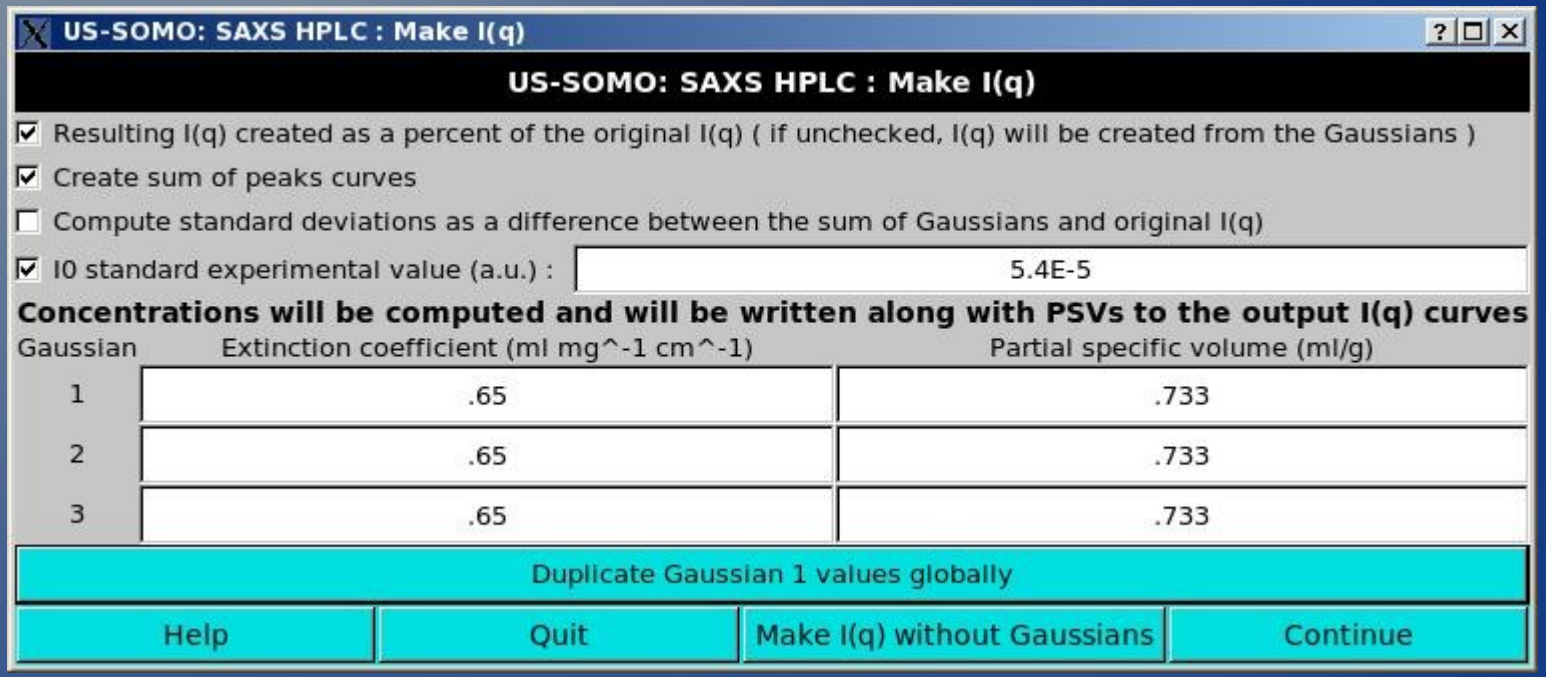
MW Computations

- Entered per I(q) curve (either Gaussian derived or general I(q))
  - c concentration (mg/ml)
    - can be computed from UV or RI curves
  - v_2 psv (ml/g)

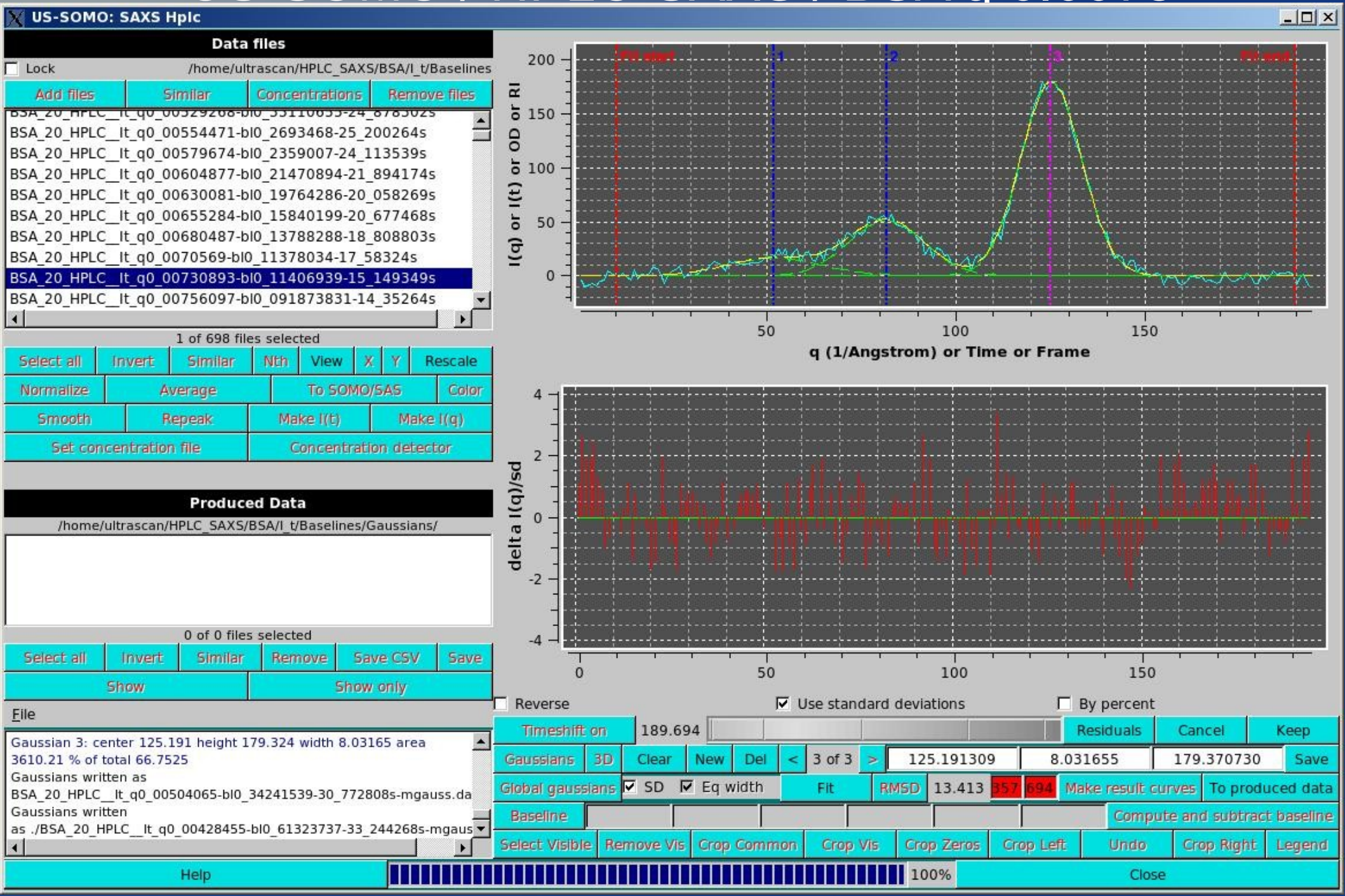
- Entered in "Gaussian" options
  - R_e the diffusion length of the electron (cm)
  - Z/A avg number of electrons per avg # of nucleons
  - m_n the mass of a nucleon (g)

- Entered in "SAS curve options"
  - ρ_e solvent e density (e/A^3)

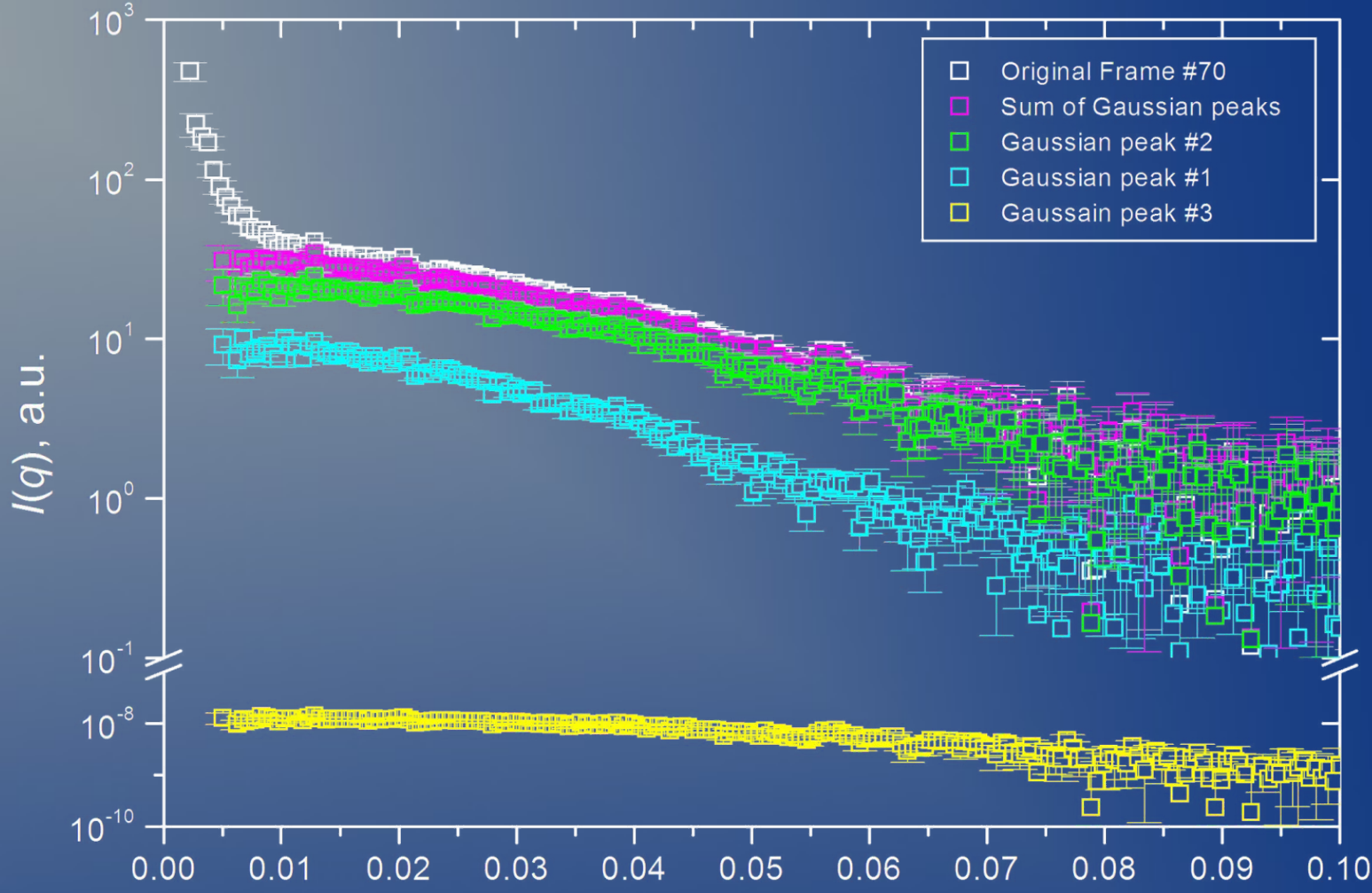
US-SOMO / HPLC-SAXS / Make I(q)



US-SOMO / HPLC-SAXS / BSA q 0.0073



US-SOMO / HPLC-SAXS / BSA



Frame, G-Pk	c, mg mL^{-1}	$[<R_g^2>_z]^{½}$, Å	$<M_w>$, g mol^{-1}	q min, Å^{-1}	q max, Å^{-1}	Fit St. Er.[a]
Frame #70, original	0.090	46.7 ± 0.8	183,277 ± 2,200	0.010081	0.028227	0.0436
Frame #70, bas. sub.	0.090	41.0 ± 2.9	145,755 ± 3,068	0.010081	0.019911	0.0435
Frame #70, G-pk #1	0.022	50.2 ± 1.0	177,944 ± 2,728	0.007813	0.028227	0.0657
Frame #70, G-pk #2	0.069	39.5 ± 1.2	134,205 ± 1,868	0.010333	0.027723	0.0493

US-SOMO / HPLC-SAXS / BSA

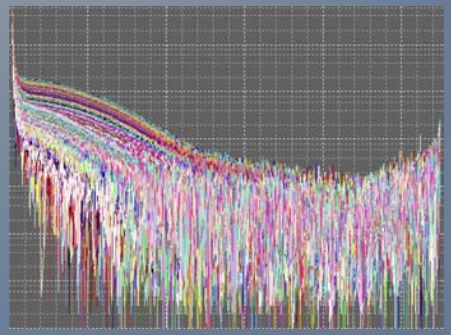
Frame, G-Pk	c, mg mL^{-1}	$[\langle R_g^2 \rangle_z]^{1/2}$, Å	$\langle M \rangle_w$, g mol^{-1}	q min, Å^{-1}	q max, Å^{-1}	Fit St. Er.[a]
Frame #70, original	0.090	46.7 ± 0.8	183,277 ± 2,200	0.010081	0.028227	0.0436
Frame #70, bas. sub.	0.090	41.0 ± 2.9	145,755 ± 3,068	0.010081	0.019911	0.0435
Frame #70, G-pk #1	0.022	50.2 ± 1.0	177,944 ± 2,728	0.007813	0.028227	0.0657
Frame #70, G-pk #2	0.069	39.5 ± 1.2	134,205 ± 1,868	0.010333	0.027723	0.0493
Frame #50, original	0.044	49.2 ± 4.5	235,472 ± 10,578	0.012350	0.021171	0.0768
Frame #50, bas. sub.	0.044	52.7 ± 2.8	177,003 ± 6,093	0.010081	0.024447	0.1054
Frame #50, G-pk #1	0.042	52.9 ± 2.8	172,525 ± 5,959	0.010081	0.024447	0.1057
Frame #81, original	0.173	41.7 ± 0.6	152,101 ± 1,415	0.014366	0.031000	0.0284
Frame #81, bas. sub.	0.173	40.3 ± 0.6	137,953 ± 954	0.010081	0.028227	0.0250
Frame #81, G-pk #2	0.160	39.0 ± 1.7	138,192 ± 2,440	0.014366	0.024447	0.0319
Frame #125, original	1.008	28.1 ± 0.3	77,147 ± 257	0.014618	0.031504	0.0098
Frame #125, bas. sub.	1.008	27.2 ± 0.4	75,189 ± 247	0.010081	0.028227	0.0116
Frame #125, G-pk #3	0.982	26.9 ± 0.3	76,557 ± 274	0.014618	0.031504	0.0106

[a] $Fit\ St.\ Er. = \left[\frac{\chi^2 \langle s.d.^2 \rangle}{DOF} \right]^{1/2}$

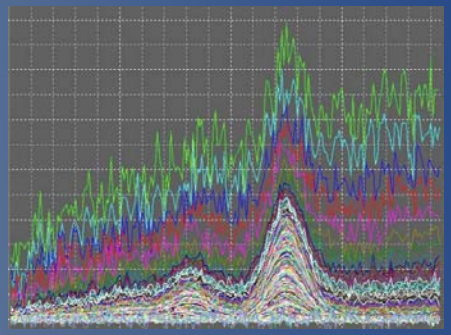
US-SOMO: HPLC-SAXS

Brookes E., Perez J, Cardinali B, Profumo A., Vachette P & Rocco M. (2013) Fibrinogen species as resolved by HPLC-SAXS data processing within the UltraScan Solution Modeler (US-SOMO) enhanced SAS module. J. Appl. Cryst. 46, 1823-1833

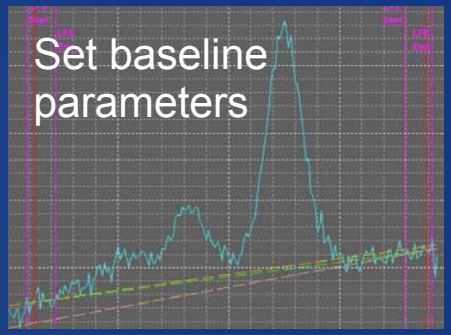
Collect HPLC-SAXS data →



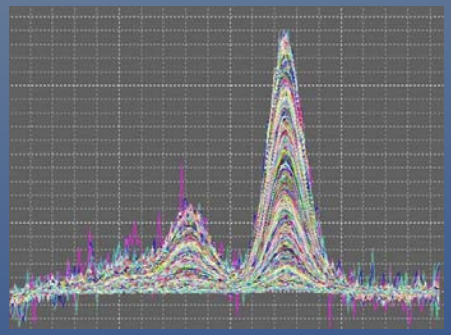
Make I(t) →



Select typical curve →

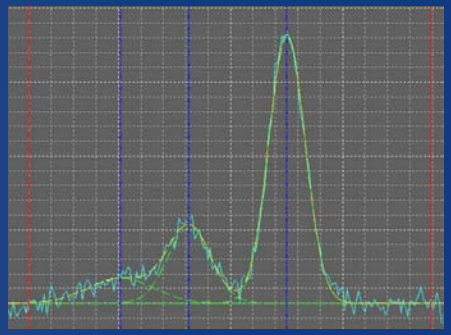


Apply b.l. params to all curves →

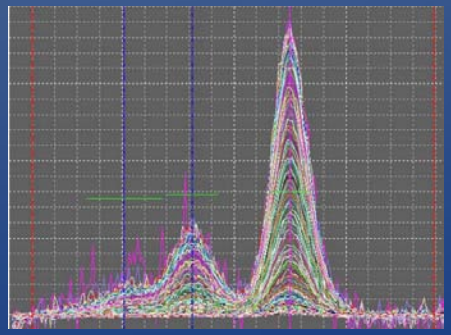


Select typical curve →

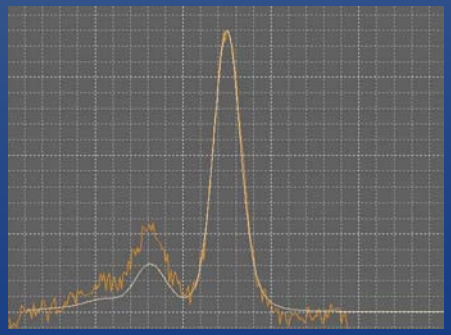
Gauss Fit →



Global Gauss Fit →



Opt. apply to conc. curve →



Make I(q) →

Set of I(q) curves for each Gaussian peak