US-SOMO Cluster Methods: Year One Perspective

Emre Brookes
Raminderjeet Singh
Marlon Pierce
Suresh Marru
Borries Demeler
Mattia Rocco

23 July 2013

XSEDE 2013
San Diego, California

Outline

- Background
  - Software
  - Science Gateway
- Community
  - Workshops
  - Followups
- Science Enabled
  - Salicylidene acylhydrazide (Byron et al)
  - Intrinsic Disorder of Single stranded DNA binding protein (Scott et al)
  - Smae-Diablo (Vachette et al)
  - Fibrinogen studies (Rocco et al)
  - Parsimonious models (Brookes)
- Future

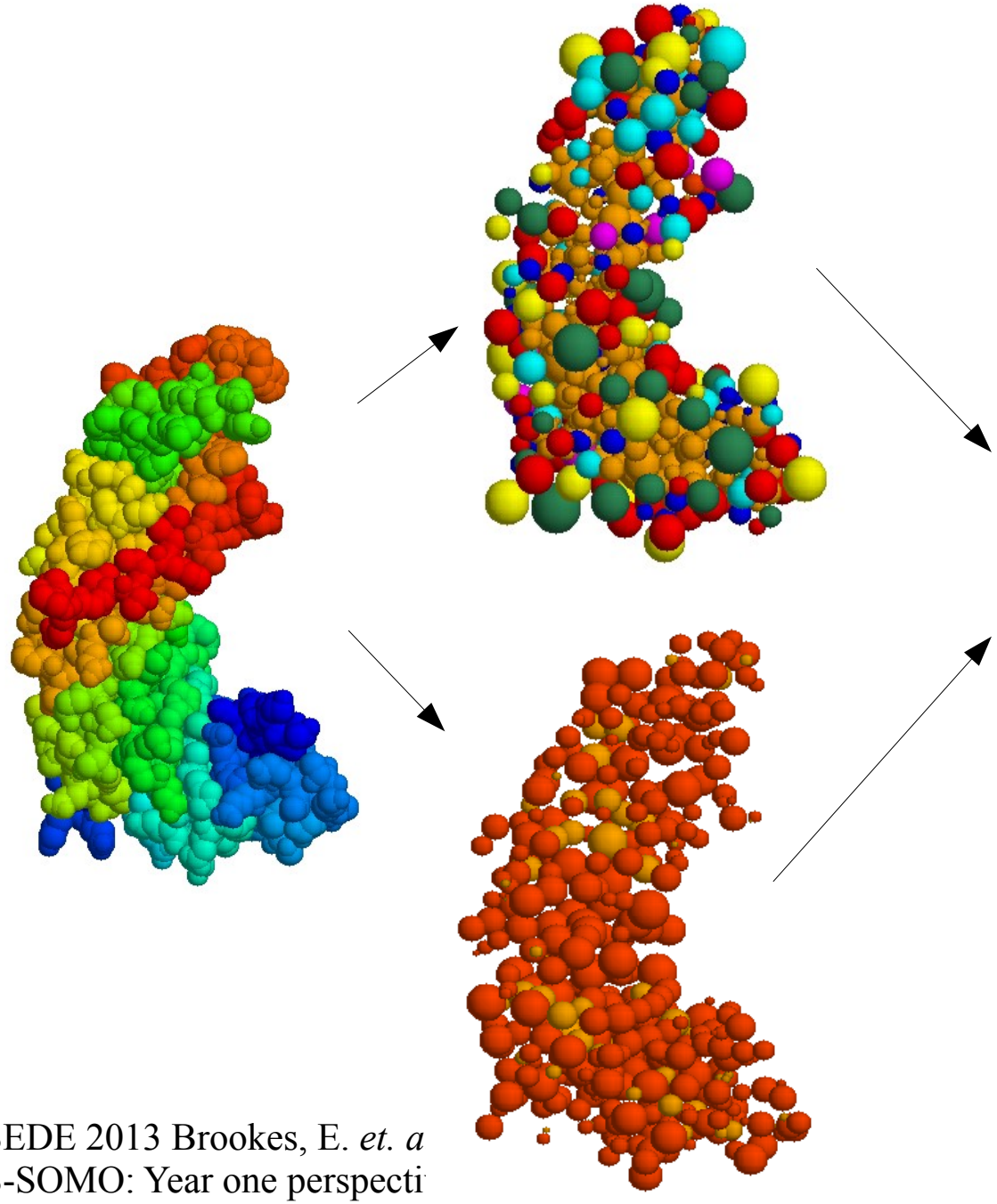
US-SOMO

- Website: somo.uthscsa.edu
- GUI based: Windows, OSX, Linux, source code (GPL).
  - C++, Qt4
- *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.*

US-SOMO

- Bead model generation from atomic structures
- Hydrodynamic parameter calculation
- SAS
- Batch & cluster computations
  - Cluster submission directly from GUI
    - Apache Airavata middleware
    - IU quarry node
- DMD on cluster for expanding conformation space
  - Run on UTHSCA's Alamo cluster

US-SOMO / Hydrodynamic Parameter Calculations

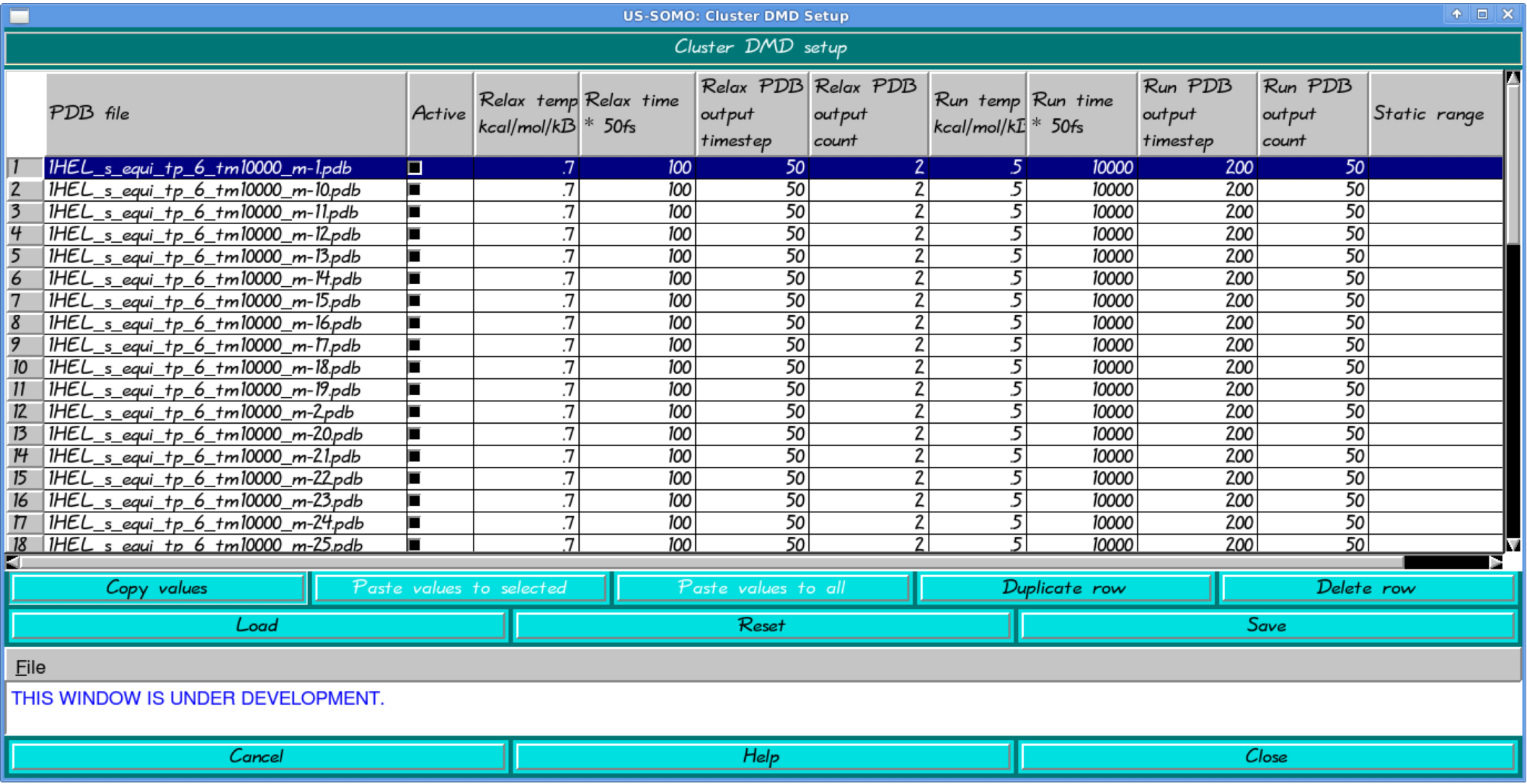


SOMO Hydrodynamic Results (Water at 20&C) (Density 1.00194 cP. Viscosity 0.998234 g/ml)	
Model:	3GKO_no_hetatm-a2b
Total Beads in Model:	439
Used Beads in Model:	369
Molecular Mass:	3.3437e+04 Da
Part. Specif. Volume:	0.733 cm^3/g
Sedimentation Coefficient s:	2.72e+00 S
Tr. Diffusion Coefficient D:	7.40e-07 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

XSEDE 2013 Brookes, E. *et. a*
US-SOMO: Year one perspecti

US-SOMO / DMD

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



SAXS/SANS Plotting Functions

PDB Filename: 1HEL_s_equi_tp_6_tm10000_m-9

Load Atom Definition File: somo.atom

Load Hybridization File: somo.hybrid

Load SAXS Coefficients File: somo.saxs_atoms

SAXS/SANS I(q) Plotting Functions:

SAXS ◇ Full ◇ H3 ◇ Fast ◇ FoXS ◇ Crysol

SANS ◇ Full ◇ H3 ◇ Fast ◇ Cryson

File suffix: fd

Compute SAXS Curve

Load SAXS Curve | Load Plotted | Set Grid | Clear SAXS Curve | Legend

Load GNOM File | Search | Guinier Analysis

☐ Create standard output files

☐ Guinier plot q^2 range:

☐ Standard ☐ Kratky plot q range: 1

P(r) vs. r Plotting Functions:

Bin size (Angstrom): 2

◇ Raw ◇ SAXS ◇ SANS ☐ Normalize

☐ Residue contrib. range (Angstrom): Display

Compute P(r) Distribution 100%

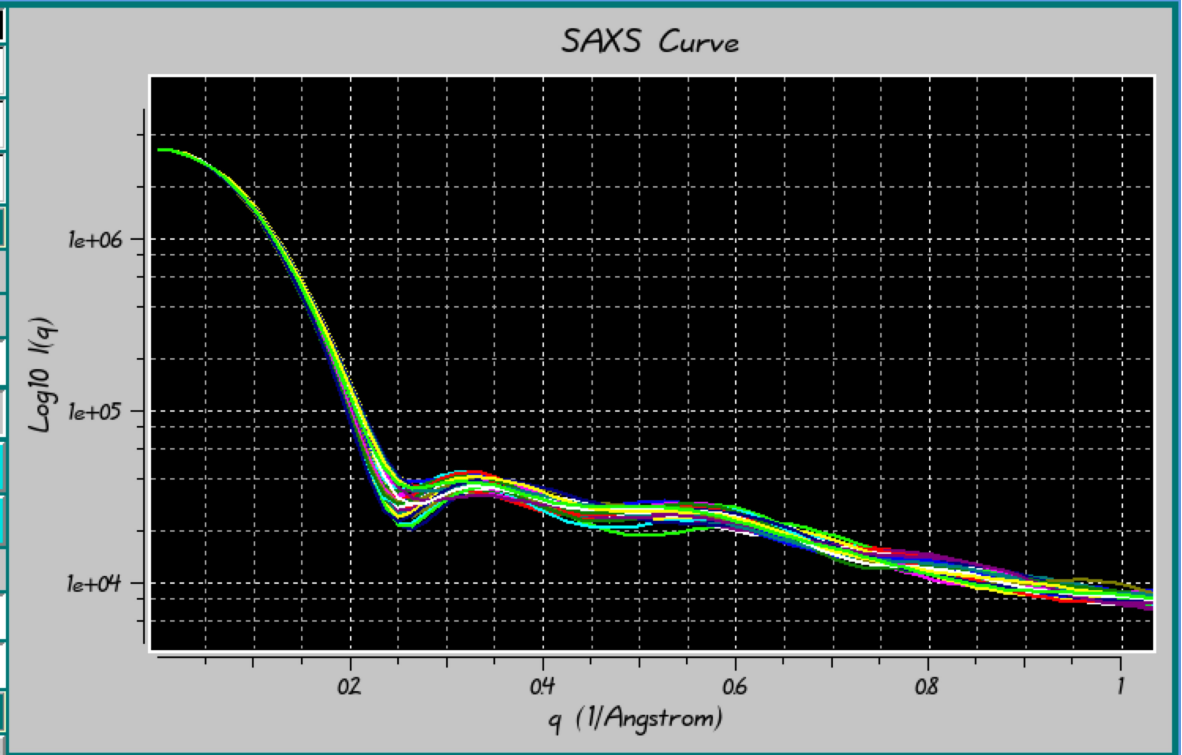
Load P(r) Distribution | Load Plotted P(r) | Clear P(r) Distribution | Legend

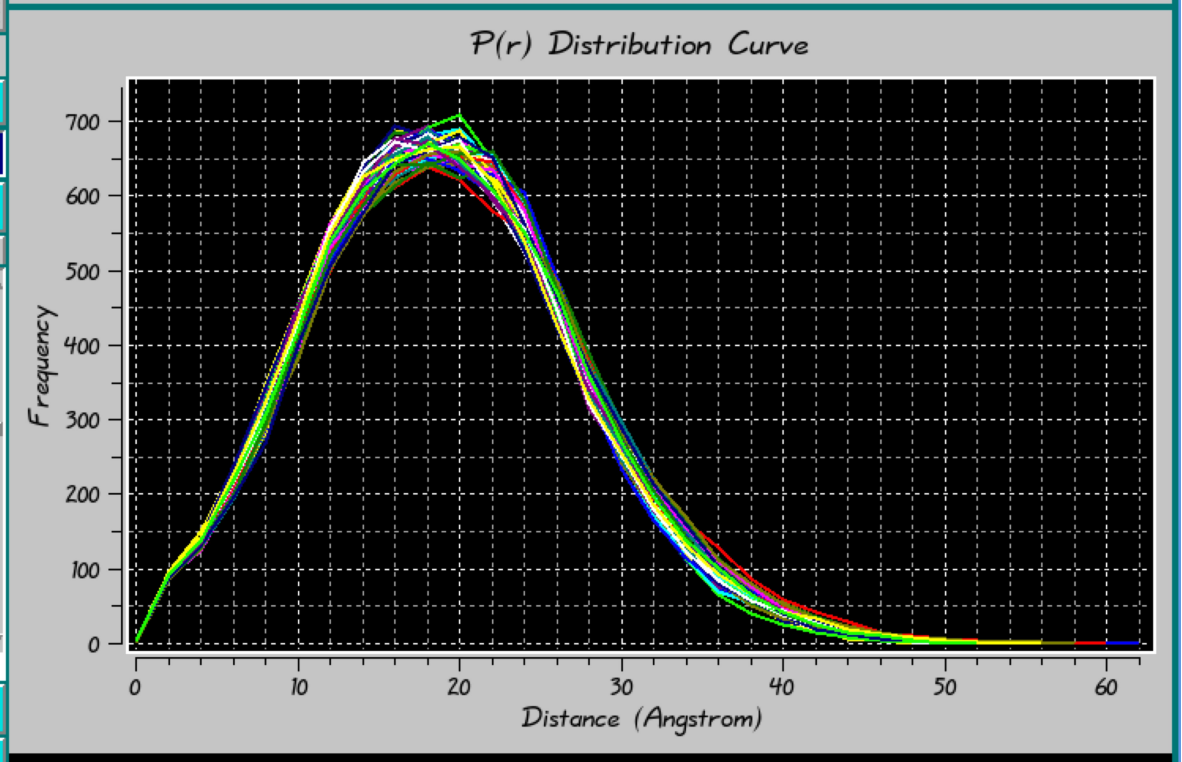
File

Model: 50"
results_iqq-fd.csv "/root/ultrascan/somo/cluster/results/dmd1/1HEL_s_equi
Model: 6"
results_iqq-fd.csv "/root/ultrascan/somo/cluster/results/dmd1/1HEL_s_equi
Model: 7"
results_iqq-fd.csv "/root/ultrascan/somo/cluster/results/dmd1/1HEL_s_equi
Model: 8"
results_iqq-fd.csv "/root/ultrascan/somo/cluster/results/dmd1/1HEL_s_equi
Model: 9"
I(q) plot done

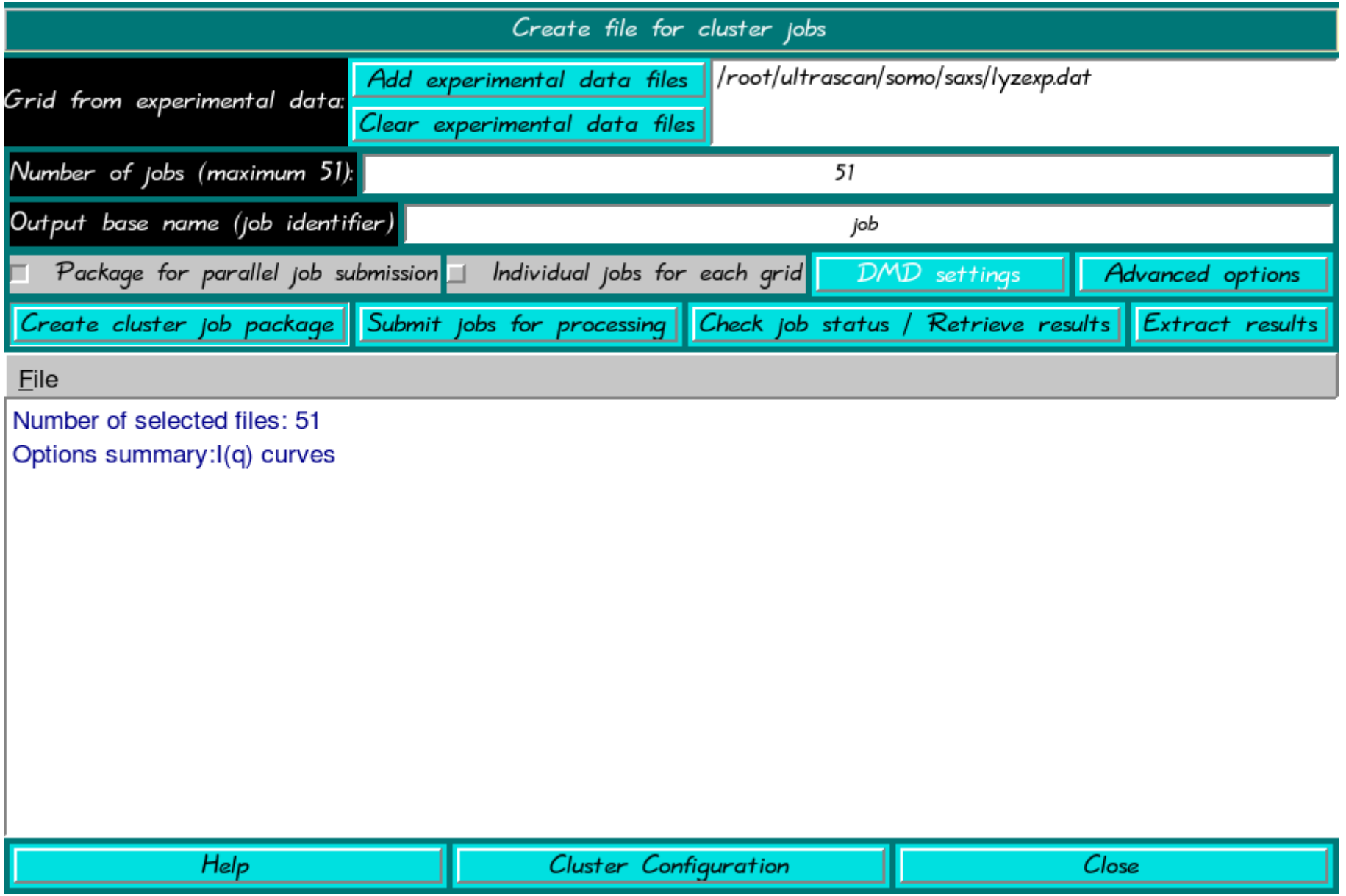
Stop | Open Options Panel

Help | Close

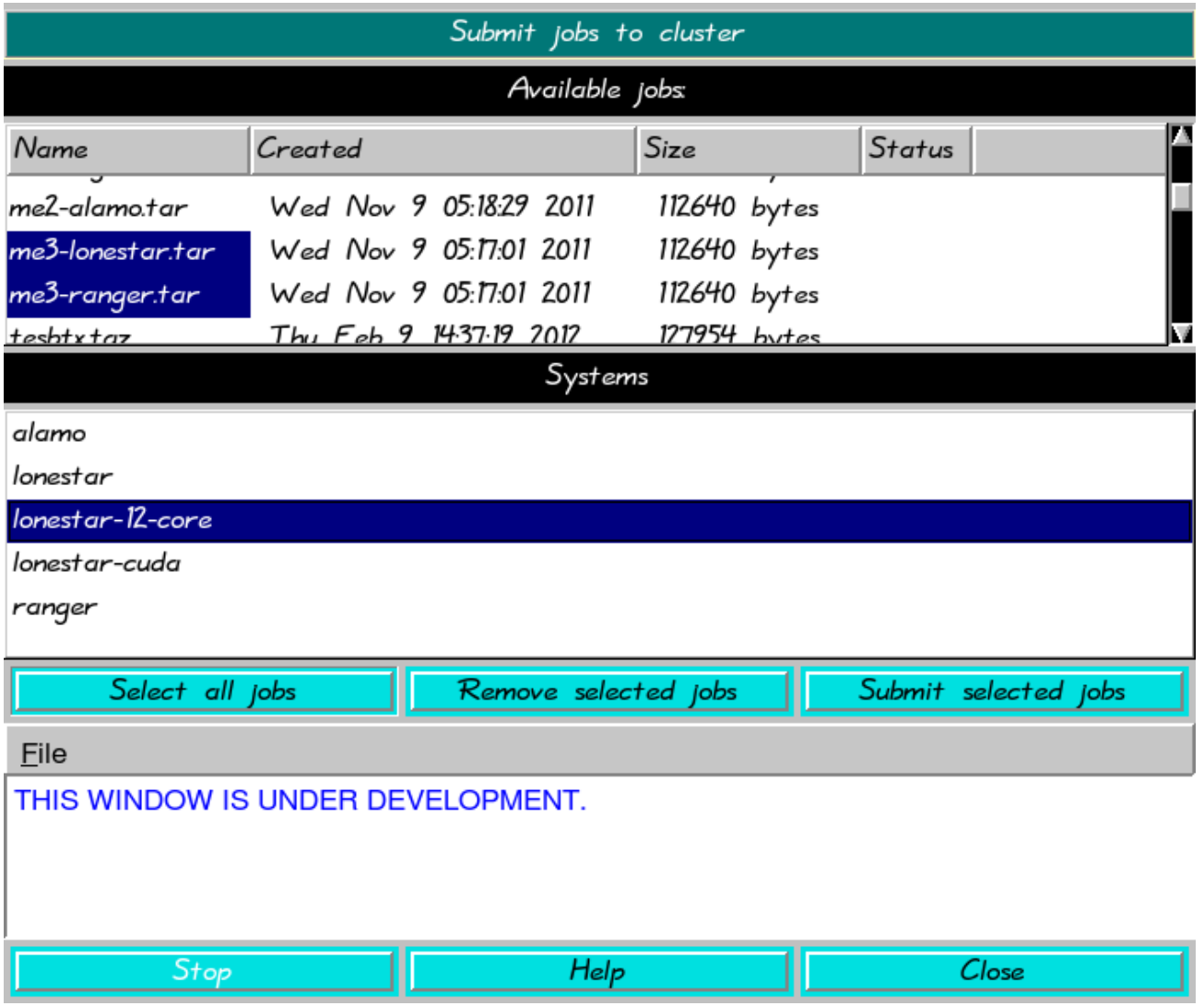




US-SOMO / Cluster Packaging



US-SOMO / Cluster Submission



US-SOMO / One possible “workflow”

- Starting with model(s)
- Expand conformation space
- Compute hydrodynamic parameters &/or SAXS curves
- Compare to experimental data
- Identify likely models
- Process locally or on clusters

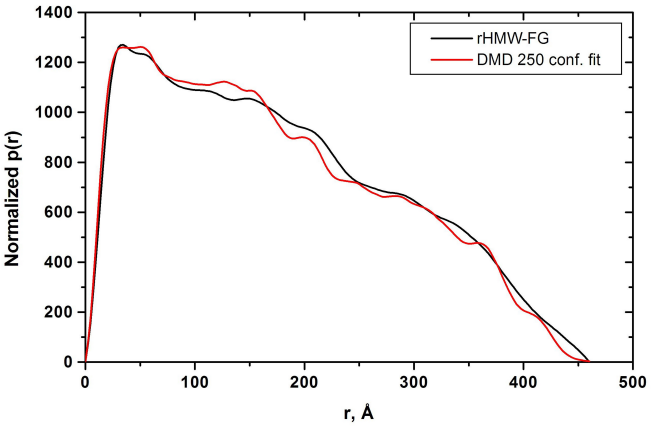
Workshops

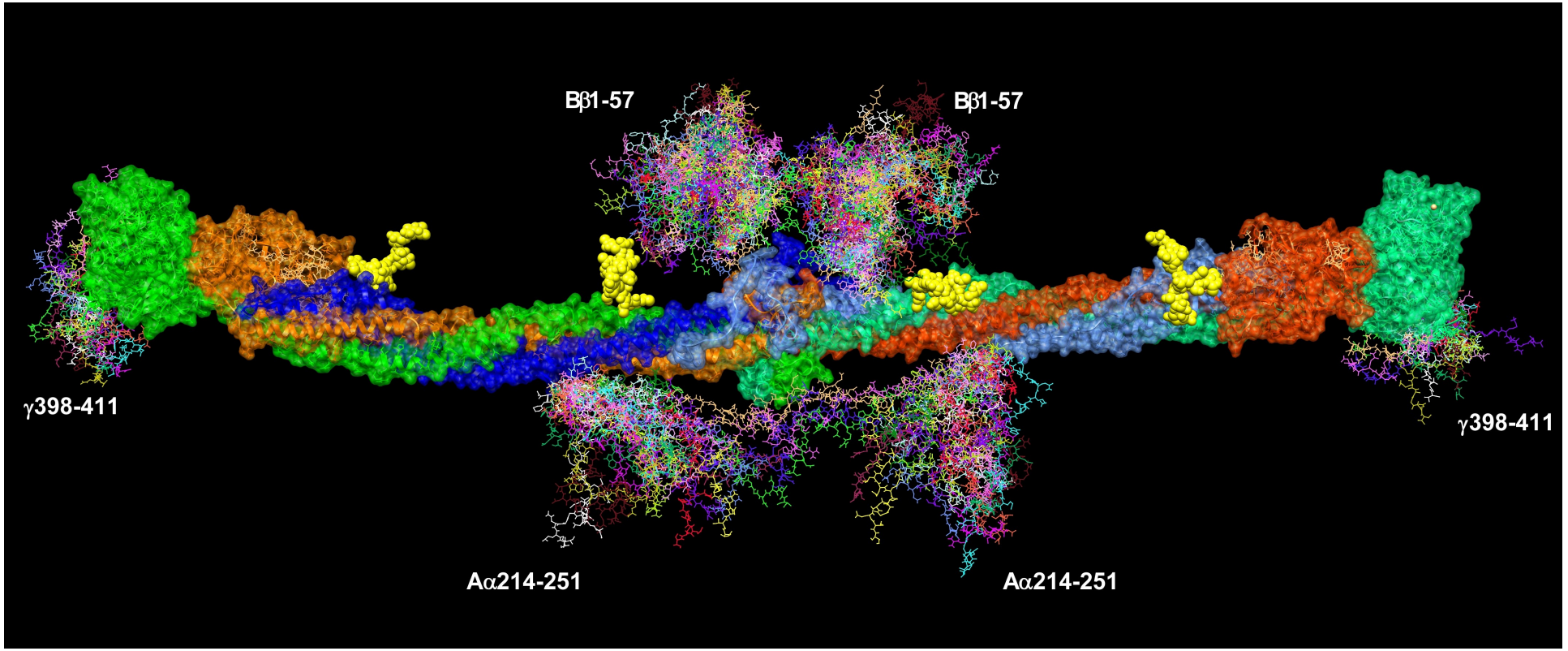
- 5 workshops so far
- 2x International Analytical Ultracentrifugation Symposium
- Harwell Research Complex, Oxfordshire
- La Trobe University, Melbourne
- UltraScan annual workshop, San Antonio

Event	Date	Attendees	Duration	Cluster usage
IAUCS #1	26 March '12	12	8 h	Demo only
IAUCS #2	27 March '12	18	8 h	Demo only
HRC	5 October '12	6	6 h	Demo only Follow up usage
LU	15-16 November '12	17	12 h	Yes, by all attendees
UltraScan	3-4 June '13	19	12 h	Yes*

Science Enabled / Fibrinogen

- Mattia Rocco et al.
- Fibrinogen is an important component of the coagulation cascade, as well as a major determinant of blood viscosity and blood flow
- A centrosymmetric dimer made by 3 pairs of chains
- US-SOMO/DMD simulations of the conformational variability for comparison to experimental data



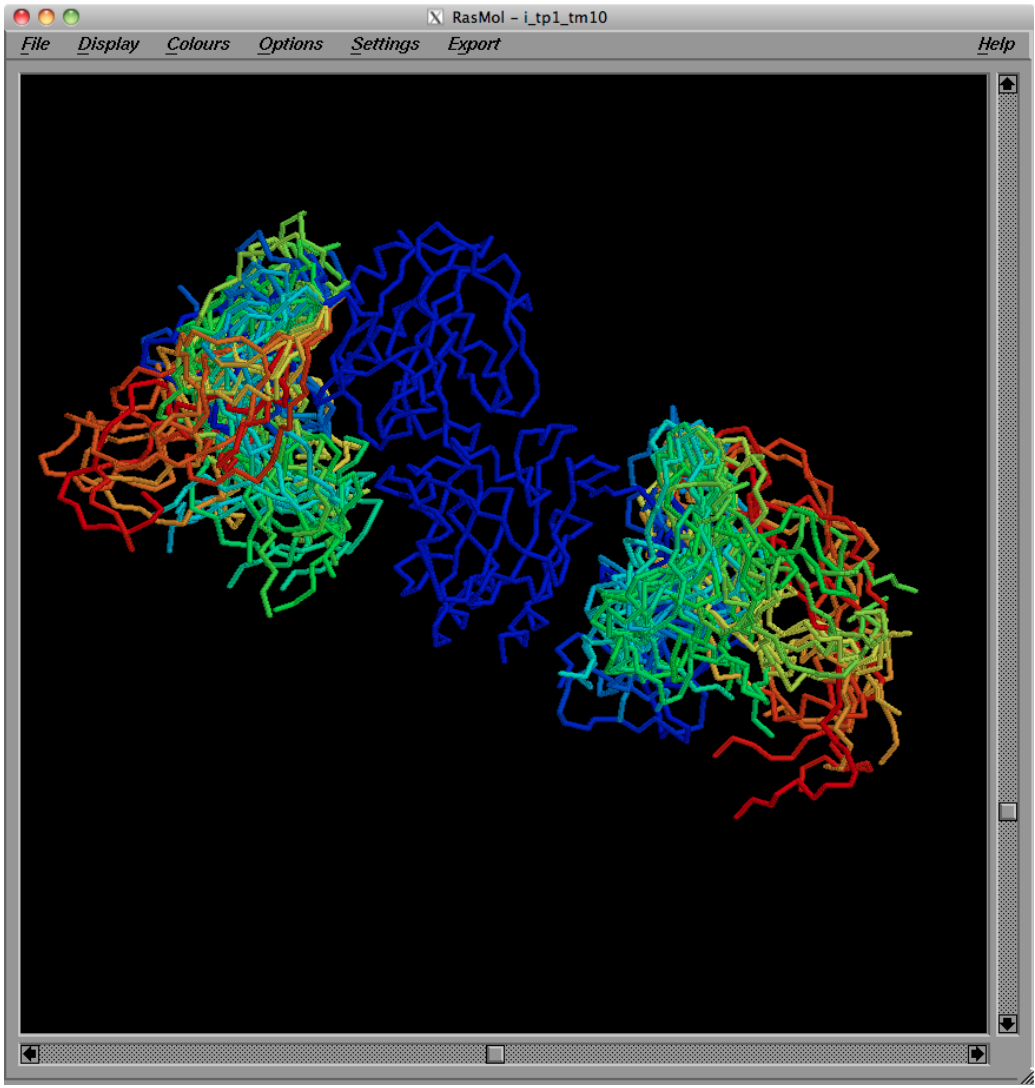


XSEDE 2013 Brookes, E. *et. al.*
US-SOMO: Year one perspective

Images credit: Mattia Rocco

Science Enabled / Salicylidene acylhydrazides

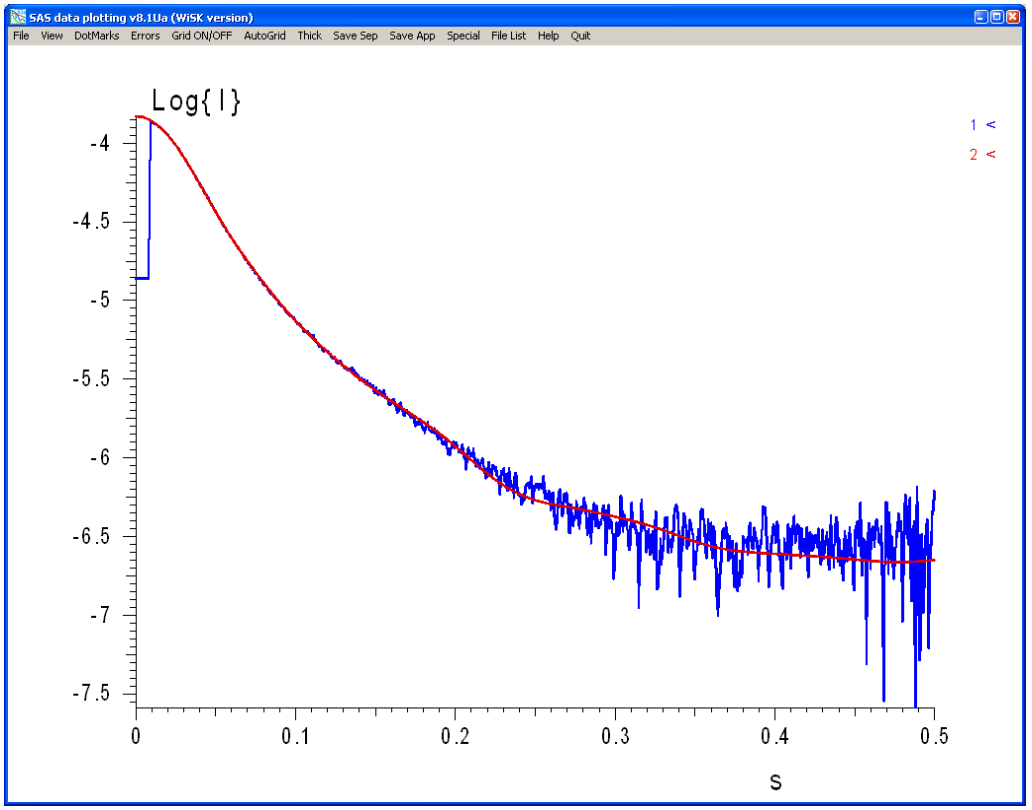
- Kate Beckman et al,
  University of Glasgow
- Salicylidene acylhydrazides inhibit virulence of E. coli O157
- Tpx
- US-SOMO used for hydrodynamic parameter calculations & DMD

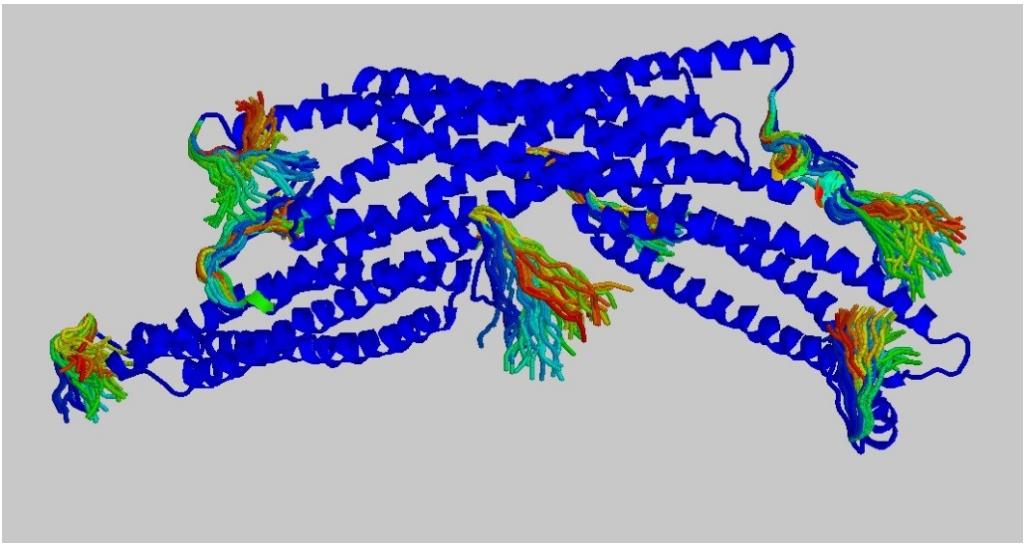


Images credit: Olwyn Byron

Science Enabled / Smac-Diablo

- Group of M. Milani, University of Milano
- Smac-Diablo a dimeric protein involved in apoptosis (programmed cell death)
- chain: 192 residues, MM= 21.8 kDa
- Final refinement of model by addition of N- and C-termini using US-SOMO/DMD



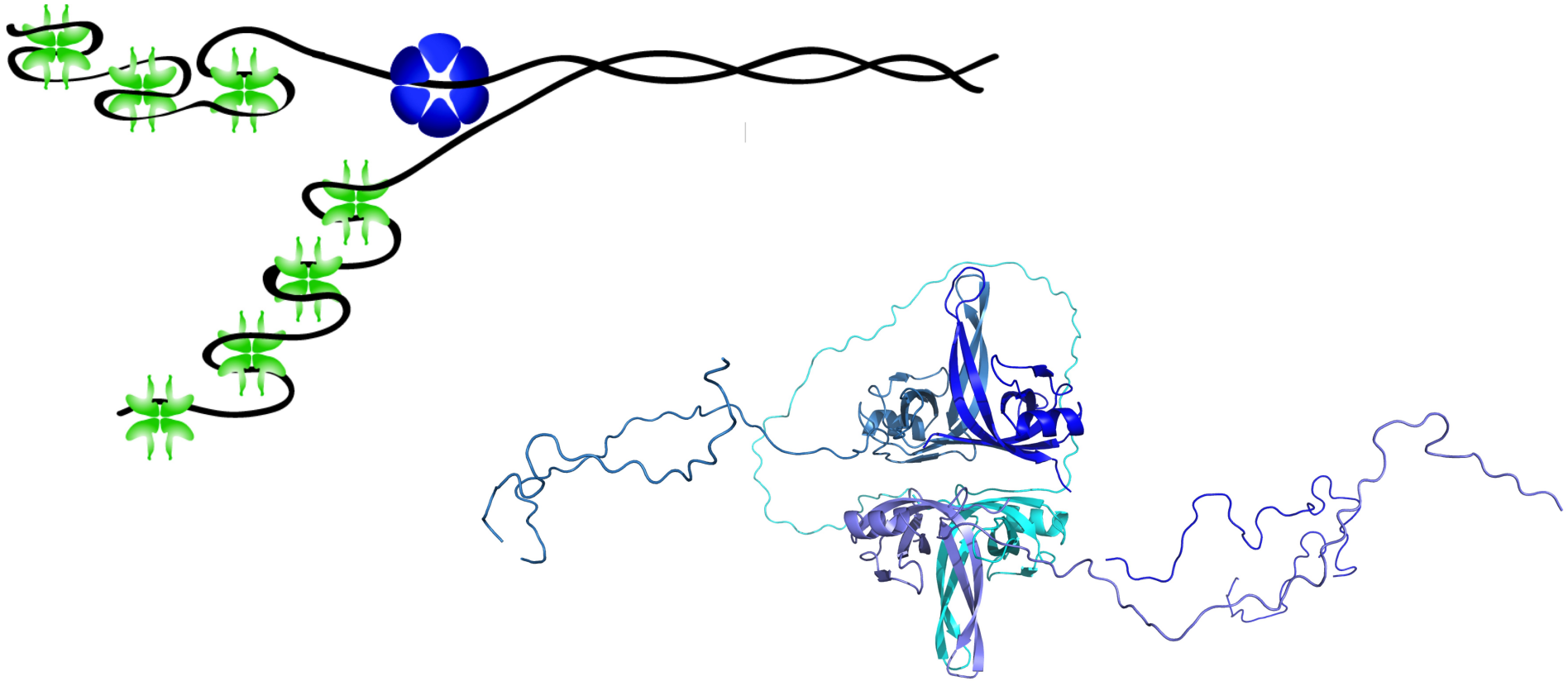


XSEDE 2013 Brookes, E. *et. al.*
US-SOMO: Year one perspective

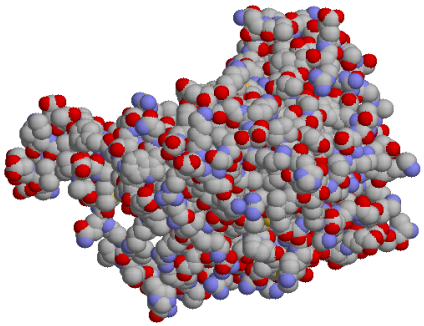
Images credit: Patrice Vachette

Science Enabled / Single Stranded DNA Binding Protein (SSB)

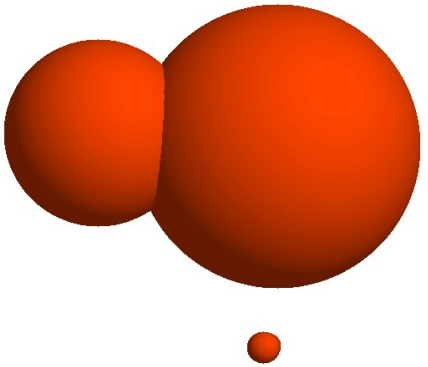
- Mathew Green et al, University of Nottingham
- SSB binds individual strands of DNA
- Critical role in DNA metabolism: Replication, recombination & repair
- Intrinsically disordered
- US-SOMO/DMD used to create conformations to screen against SAXS data

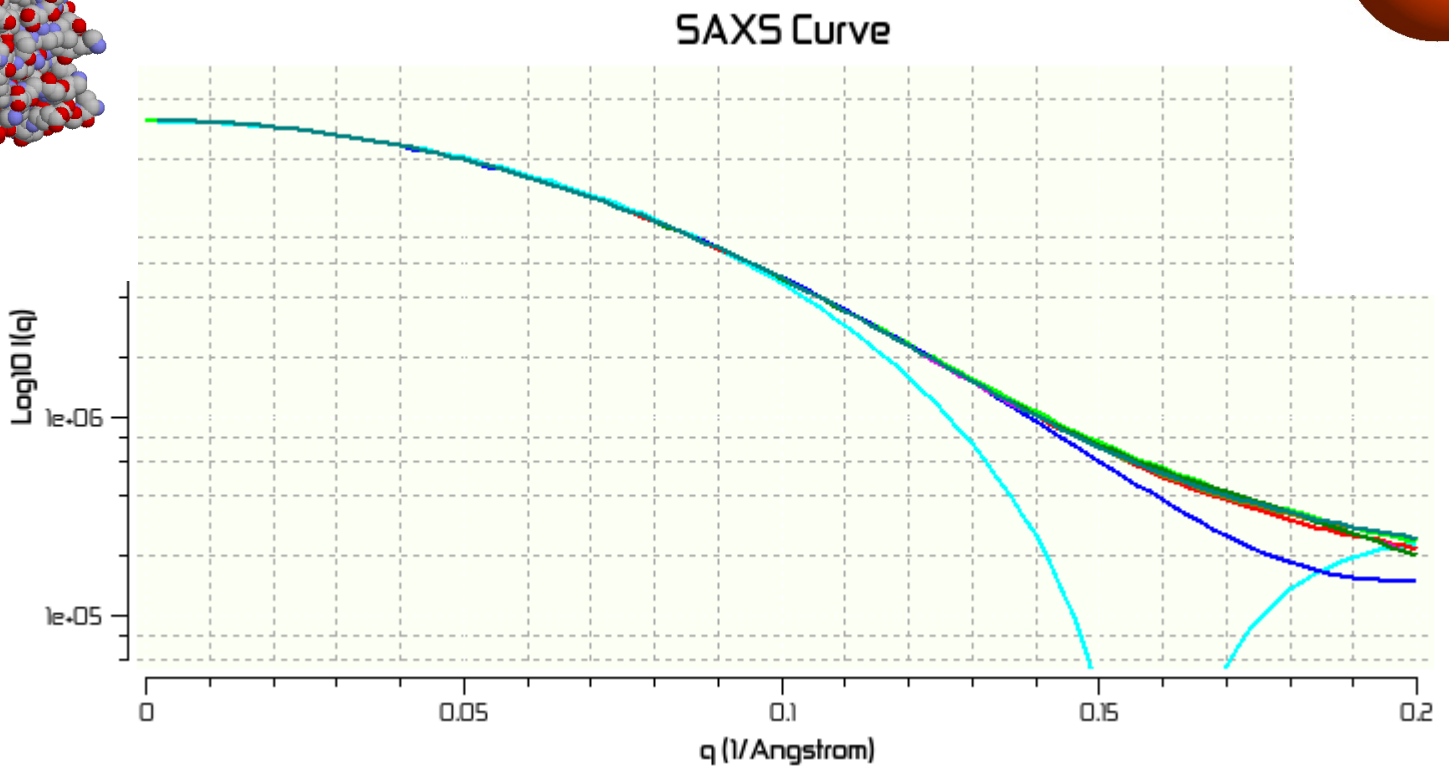


Images credit: Mathew Green



SE: Parsimonious spatial models





Model name	D(tr) [cm/sec^2]	Rg [nm]	Max extensions X [nm]	Y [nm]	Z [nm]	Axial ratios X:Z	X:Y	Y:Z
1CTS_1-db_1sa-10_1	7.530e-07	2.20	5.69	5.69	5.69	1.00	1.00	1.00
1CTS_1-db_2sa-10_1	7.480e-07	2.34	7.62	5.04	5.04	1.51	1.51	1.00
1CTS_1-db_3sa-10_1	7.410e-07	2.34	7.42	5.06	5.06	1.47	1.47	1.00
1CTS_1-db_4sa-10_1	6.910e-07	2.39	7.39	8.62	5.52	1.34	0.86	1.56
1CTS_1-db_5sa-10_1	7.430e-07	2.31	7.43	6.37	5.06	1.47	1.17	1.26
1CTS_1-db_6sa-10_1	6.900e-07	2.38	7.32	8.50	5.52	1.33	0.86	1.54
1CTS_1-db_7sa-10_1	7.390e-07	2.32	7.48	5.12	6.09	1.23	1.46	0.84
1CTS_1-so	7.098e-07	2.32	7.46	6.31	5.52	1.35	1.18	1.14

Future

- EPSRC/NSF: *SI2-CHE: CCP-SAS – Collaborative Computing consortium for advanced analysis of structural data in chemical biology and soft condensed matter*
  - UK Lead PI Steve Perkins, UCL; US Lead PI Paul Butler, NIST
    - Open source
    - SASSIE
    - Perkins code
  - Framework for scientific application development
    - generate GUI & web based versions from a single set of minimal text based description files
      - GUI required for secure data
      - needs to be simple enough for a scientist to wrap their own applications
      - Current base python, c, fortran, c++ (qt)
    - local & “cloud/grid” computations
    - big data
    - if something already exists will adopt/adapt
      - otherwise, we will have to create
  - ½ a postdoc

Acknowledgments

- To E. Brookes
  - NIH NIGMS 1K25GM090154 / PI
  - NSF OCI-1032742 (Pierce) / co-PI
  - NSF CHE-1265817 / PI
  - NSF XSEDE

Funded by a grant from the National Institute of General Medical Sciences of the National Institutes of Health