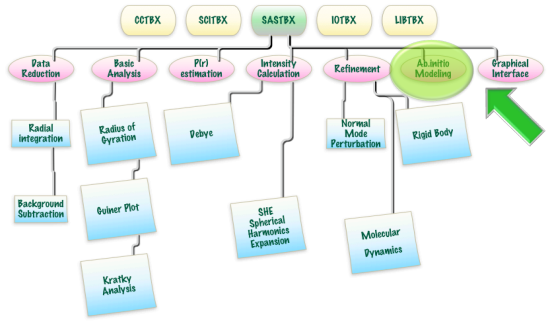
Accelerated shape discovery from SAXS data with the Small Angle Scattering ToolBox (SASTBX)

Peter Zwart

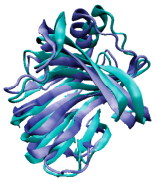
The SASTBX

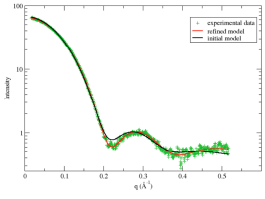
- Open source software for analyses of SAXS data
  - —http://sastbx.als.lbl.gov
  - —Extends the CCTBX
  - —Hybrid C++/phython environment
- Tools
  - —Kratky analyses
  - —I0 / Rg estimates
  - —P(r) estimates
  - —Model data calculation
  - —Model refinement
  - —Shape discovery



Model refinement

- How to modify a crystals structure so that it fits SAXS data?
  - —Iterative normal mode perturbations to model
  - —Use she (spherical harmonic expansion) for model data calculation





Zernike Polynomials

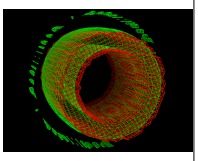
$$Z_{nlm}(\boldsymbol{r}) = R_{nl}(r) Y_{lm}(\theta,\phi) \;\; with \; \boldsymbol{r} = (r,\theta,\phi)$$

Orthonormal in unit sphere

Zernike Moments

Polynomial weights

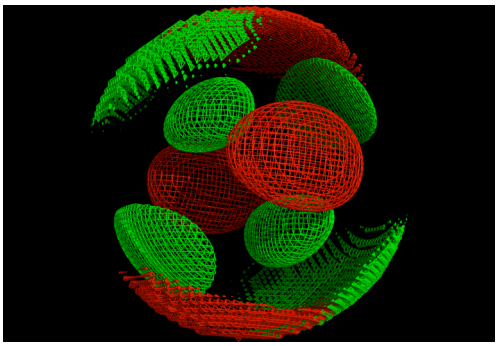
$$c_{nlm} = \frac{3}{4\pi} \int_{|\boldsymbol{r}|<1} \rho(\boldsymbol{r}) Z^*_{nlm} d\boldsymbol{r}$$



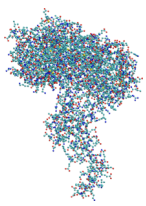
Weighted Sum of basis functions

$$\widehat{\rho}(\boldsymbol{r}) = \sum_{n=0}^{n_{max}} \sum_{l=0}^{n} \sum_{m=-l}^{m=l} c_{nlm} Z_{nlm}(\boldsymbol{r})$$

Real space representation of 3DZP



Model Representation



C_{nlm}

- F_{nl}
- $F_{nn'l}$
- $H_{nn'}$

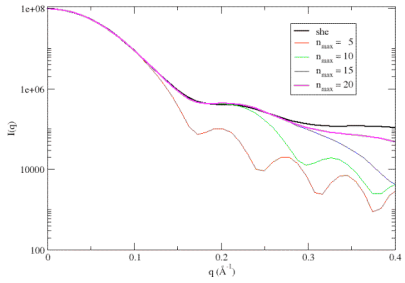
{(x,y,z)} 1000x

{C_{nlm}} 100x

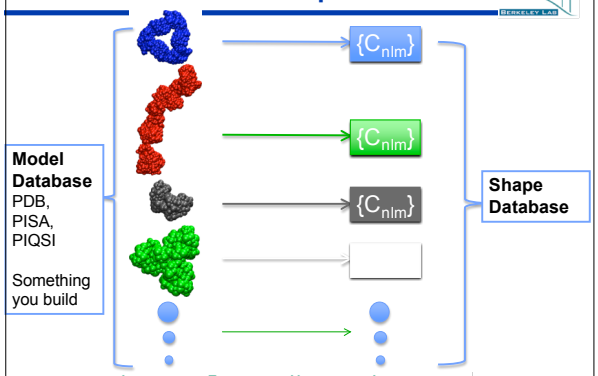
Intensity Calculation

$$I(q) = \sum_n \sum_{n'} \sum_l B_{nl}(qr_{\max}) B_{n'l}(qr_{\max}) k_{nl} k_{n'l} \sum_m c_{nlm} c^*_{n'lm}$$

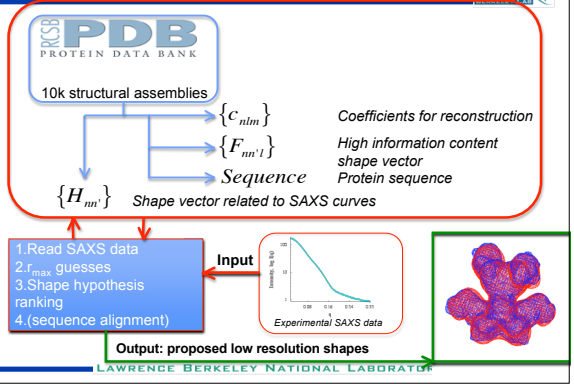
$$I(q) = \sum_n \sum_{n'} B_{nn}(qr_{\max}) B_{n'n'}(qr_{\max}) H_{nn'}$$



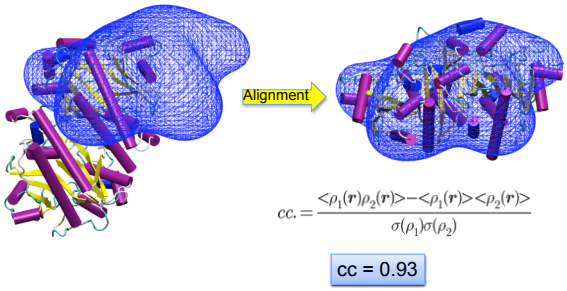
Build Shape Database



Rough shape classification from SAXS data

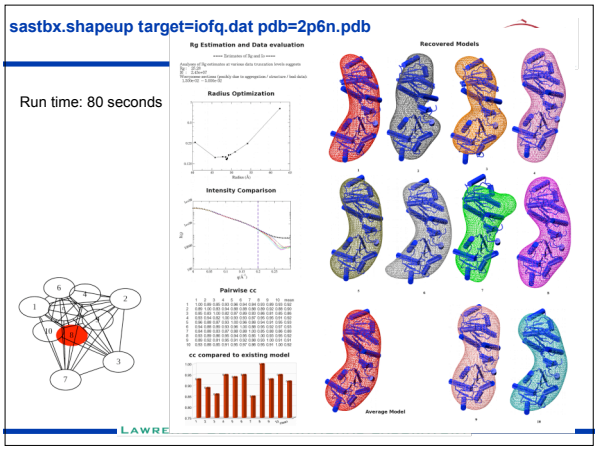


Model Comparison

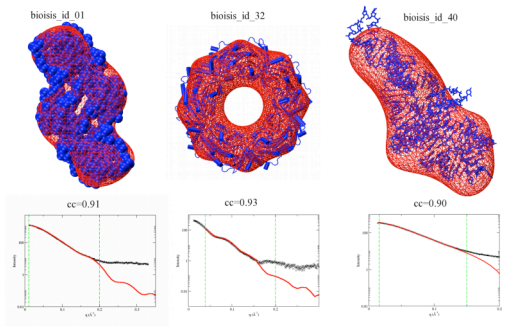


sastbx.shapeup target=iofq.dat pdb=2p6n.pdb

Run time: 80 seconds



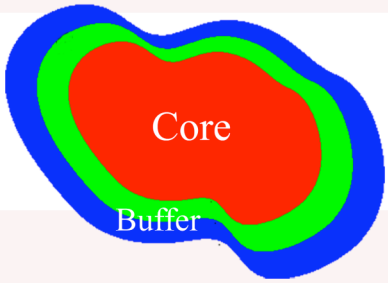
Bioisis data



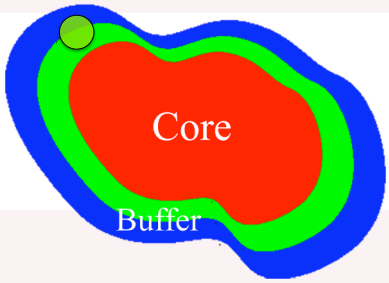
Features

- **Fast and Automated**
  - —**Providing preliminary results in 2 minutes**
  - —**real time feedback to experiments**
- **Flexible approach**
  - —**Customized shape database can enable even faster shape reconstruction and enhance more relevant results**
  - —**Not limited to biomolecules**

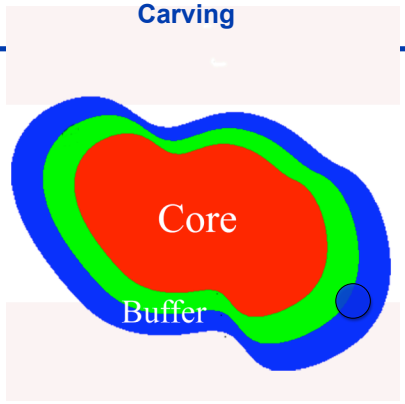
Refinement



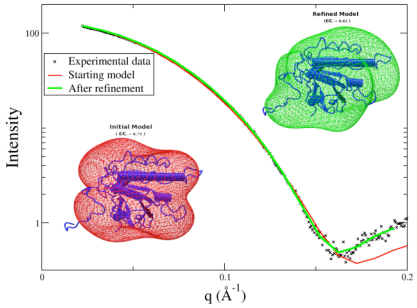
Expansion



Carving



Local Refinement

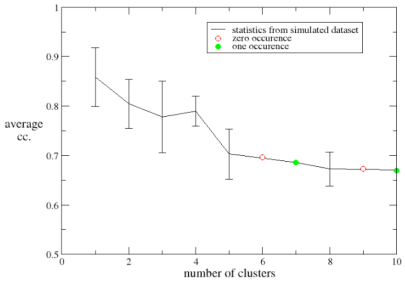


Challenging Questions

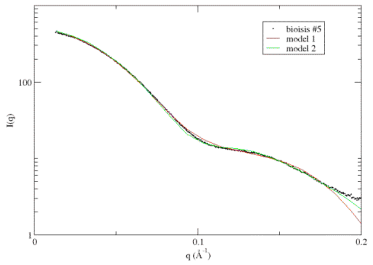
- Number of Zernike polynomes
  - —Speed vs Accuracy (and q_range)
- q_range
  - —q_start; q_stop
- Model Quality
  - —Compare to available model
  - —Check consistency of the top models

Model Quality Measure

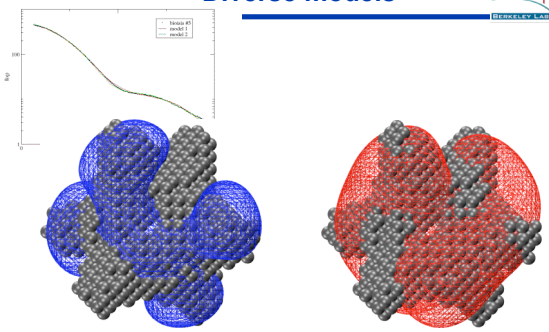
- Model Quality is positively correlated to the consistency



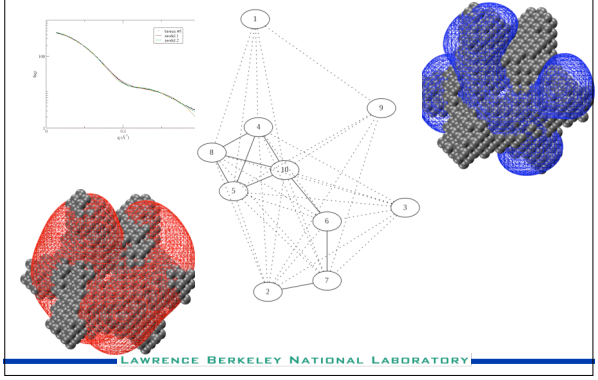
Diverse models



Diverse models



Diverse models

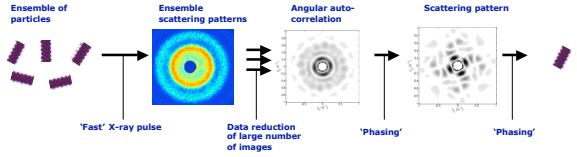


Summary

- Zernike Moments can represent biomolecule models
- SAXS intensity can be calculated from zernike moments in an efficient manner
- Model can be efficiently recovered from Databases
- The process is fast and automated
- The model quality can be gauged by comparing to existing models or from model consistency

Outlook

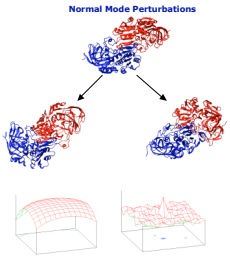
- Increasing information content in SAXS data can be achieved by taking ultra-fast snapshots, before molecules reorient themselves
- Fluctuation X-ray Scattering (fXS)



Beyond Small Angle Scattering: Exploiting angular correlations
Saldin et al, (2010) Phys Rev B, 81, 174105.

Outlook

- The fXS experiment results in an estimate of the auto correlation function of the scattering pattern of a single particle.
- fXS data has orders of magnitudes higher information content as compared to SAXS data.
- fXS can in principle be used for ab-initio shape reconstruction as well
  - The 2D case has recently shown to be feasible
  - The 3D case is feasible as well, but await implementation
- All that can be done with SAXS/WAXS can be done with fXS data.



Acknowledgements

Haiguang Liu
Ralf Grosse-Kunstleve
Paul Adams
Alex Hexemer
Eric Schaible
Robert Rambo

Dilano Saldin
John Spence

$$$ LBNL

Question/Suggestion, send to HGLIU@LBL.GOV

jc/ALSaerial/11-96