Advanced Data Analysis:
SEC, SEC-MALS, SEC-SAXS, Reconstructions

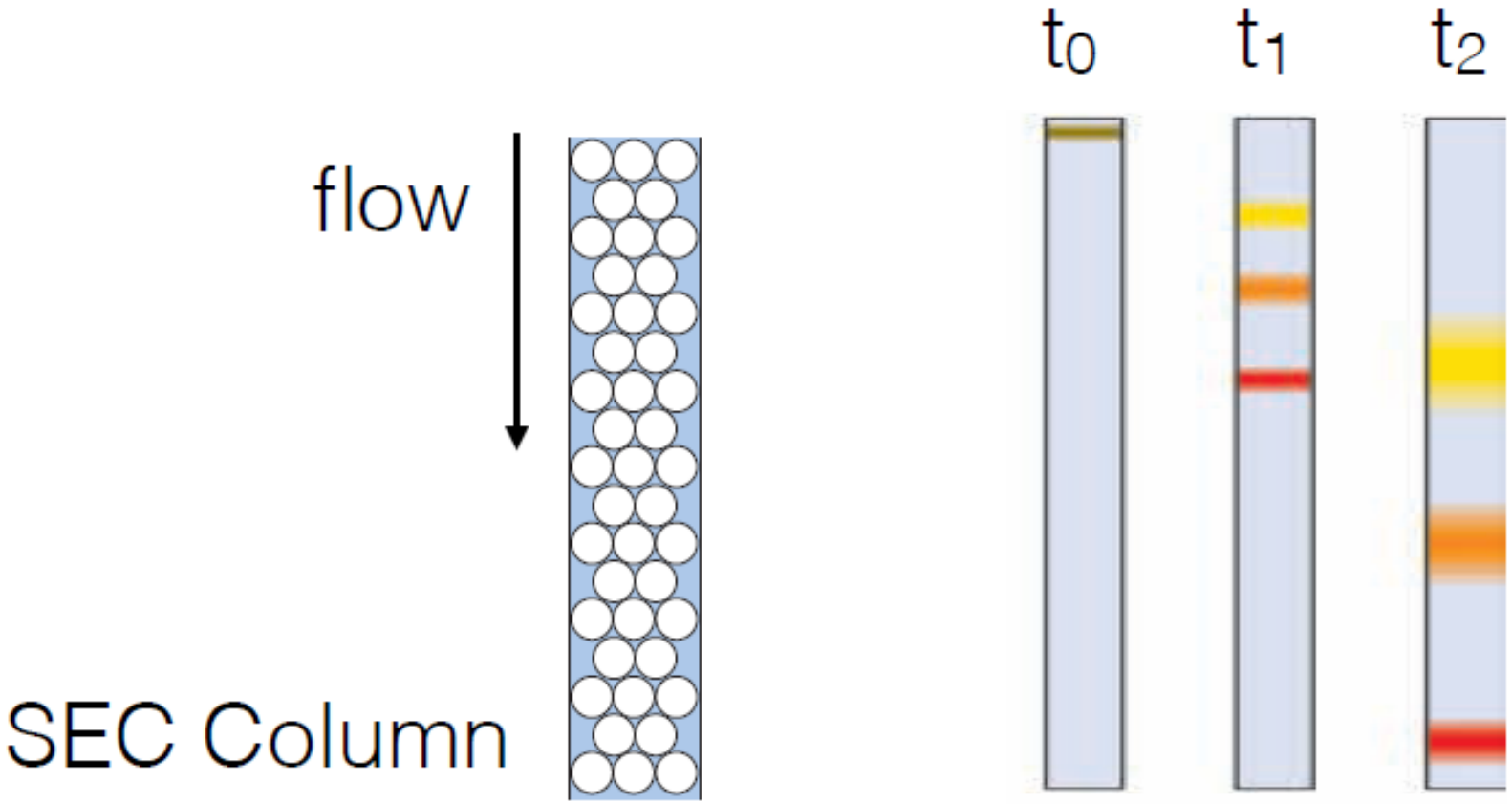
Everything BioSAXS 5
Getting Started in Biological Small-Angle X-ray Solution Scattering
Tuesday 11/5/19

Kushol Gupta
Research Asst. Professor
Department of Biochemistry and Biophysics,
Perelman School of Medicine (Univ. of Penn.)
kgupta@pennmedicine.upenn.edu

Advanced Data Analysis

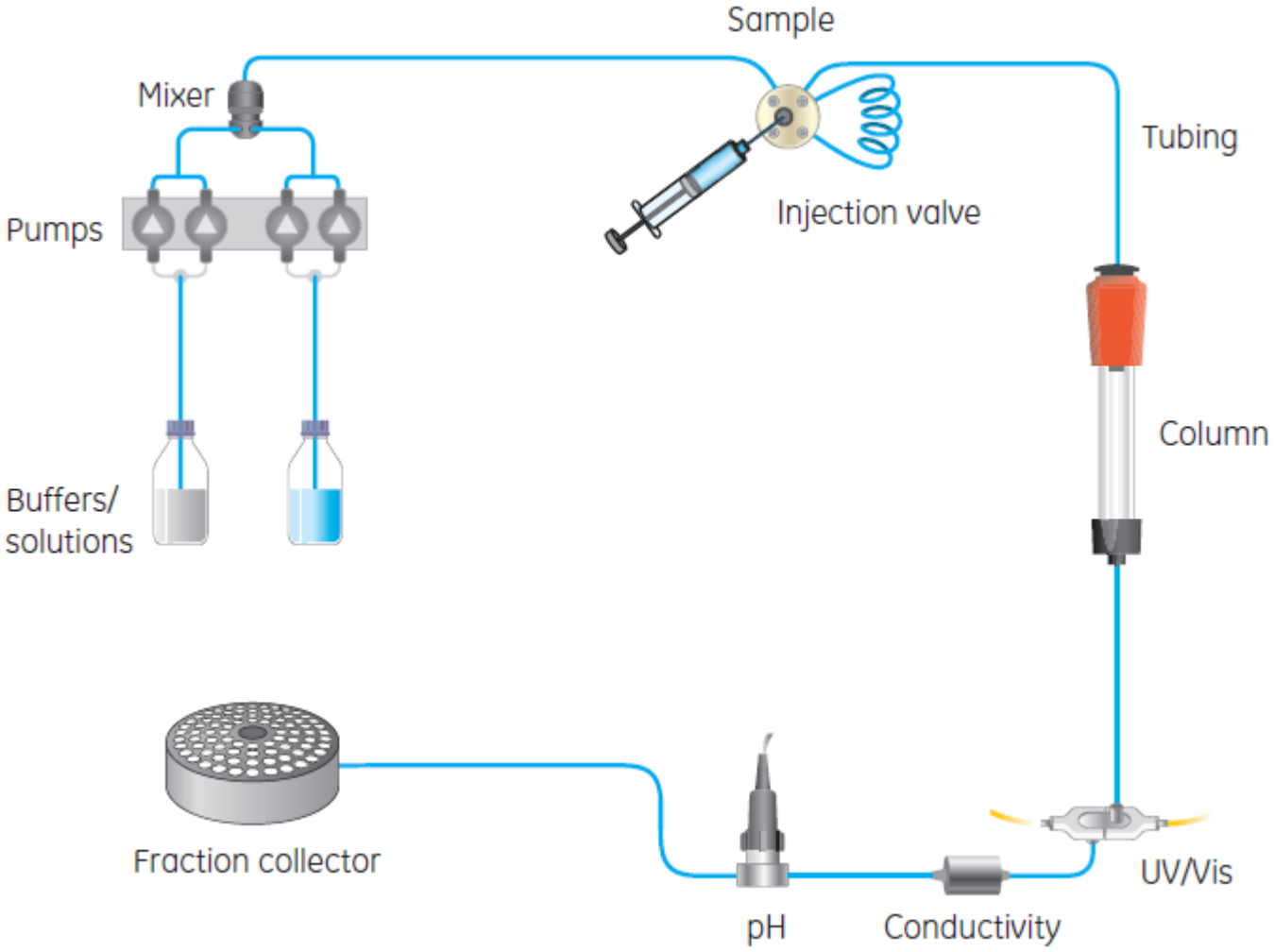
- **SEC and SEC-MALS**
- SEC-SAXS
  - SVD-EFA
- *Ab Initio* Reconstructions
  - DAMMIN/F
  - GASBOR
  - DENSS
  - MONSA

Principles of SEC



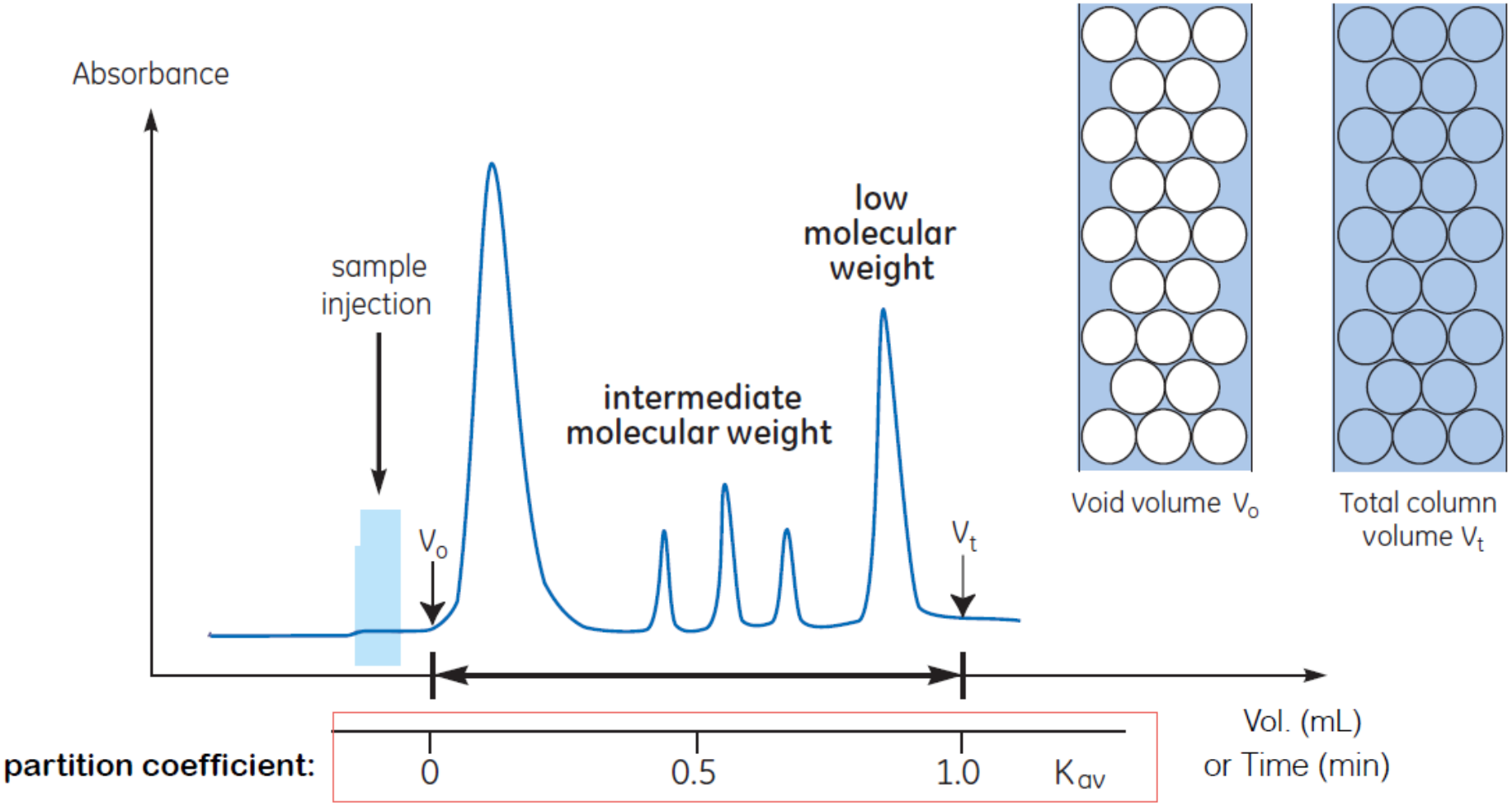
"Size-Exclusion Chromatography: Principles and Methods," GE Healthcare Life Sciences

Principles of SEC



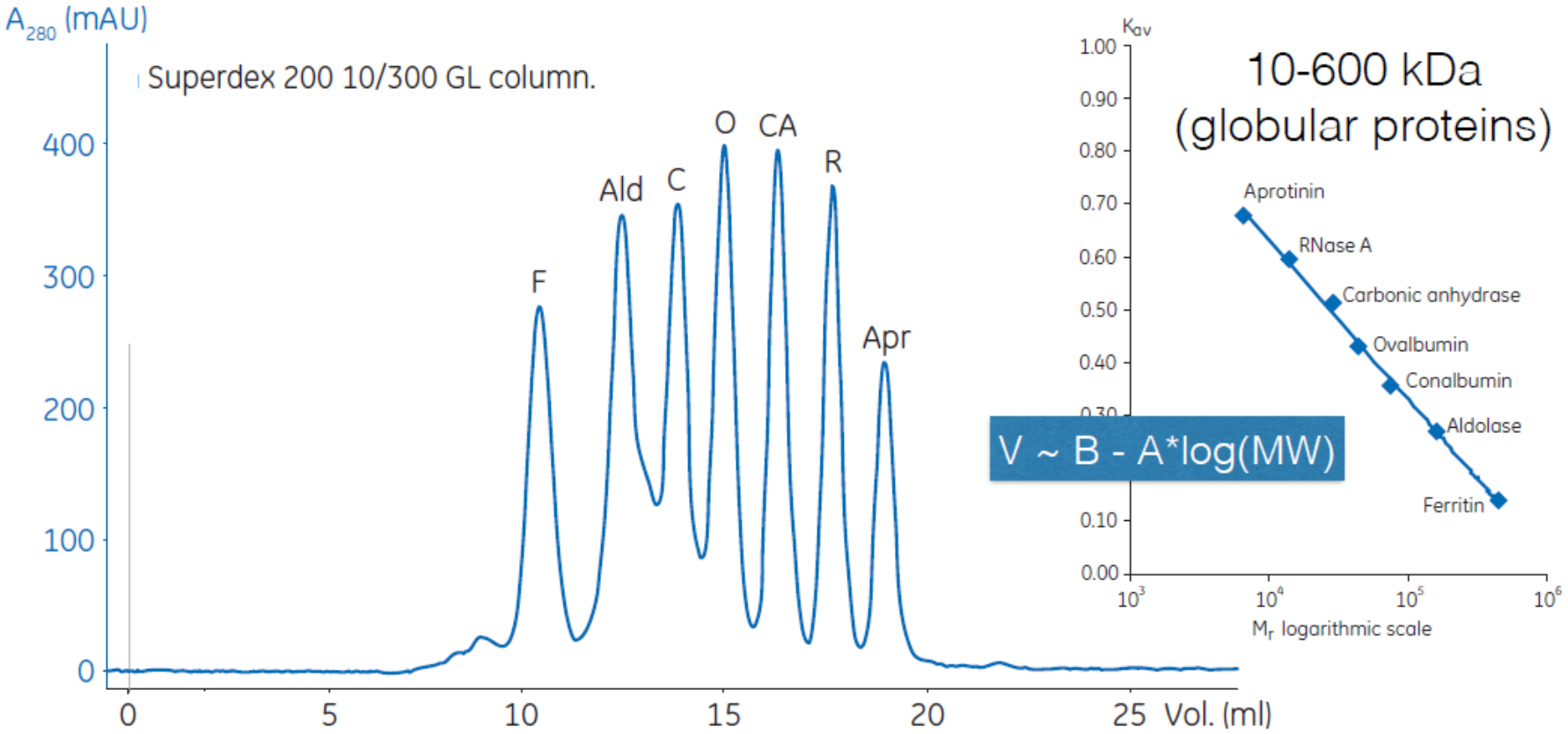
"Size-Exclusion Chromatography: Principles and Methods," GE Healthcare Life Sciences

Principles of SEC



"Size-Exclusion Chromatography: Principles and Methods," GE Healthcare Life Sciences

Principles of SEC



"Size-Exclusion Chromatography: Principles and Methods," GE Healthcare Life Sciences

Principles of SEC

Biochem. J. (1987) **243**, 399–404 (Printed in Great Britain)

The use of gel chromatography for the determination of sizes and relative molecular masses of proteins

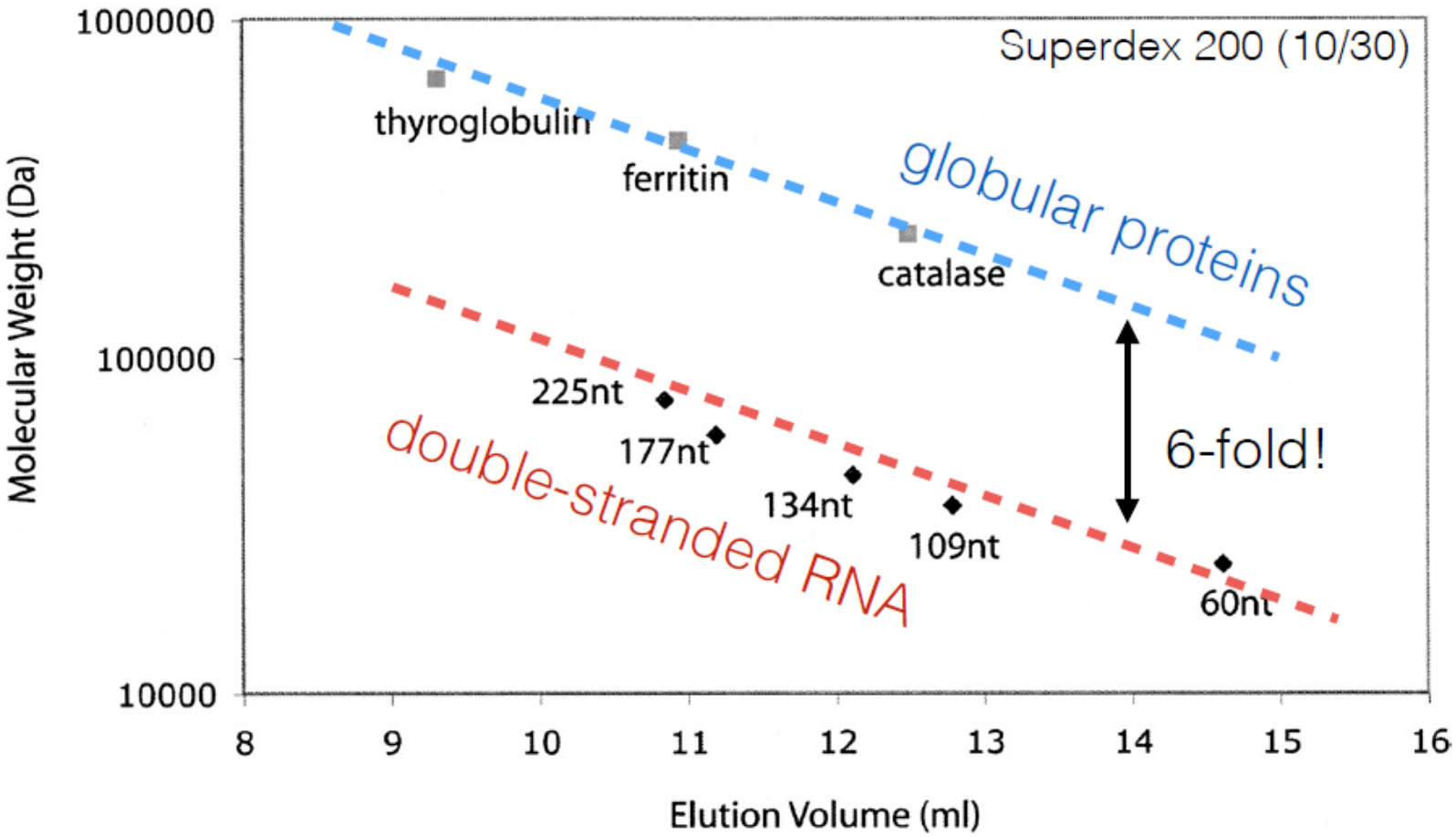
Interpretation of calibration curves in terms of gel-pore-size distribution

Marc LE MAIRE,*§ Alexandre GHAZI,† Jesper V. MØLLER‡ and Lawrence P. AGGERBECK*‖

*Centre de Génétique Moléculaire, Laboratoire propre du Centre National de la Recherche Scientifique, Associé à l'Université Pierre et Marie Curie (Paris VI), 91190 Gif-sur-Yvette, France, †Laboratoire des Biomembranes (UA 1116), Bat. 433, Université Paris-Sud, 91405 Orsay, France, and ‡Institute of Medical Biochemistry, Aarhus University, DK 8000 Aarhus, Denmark

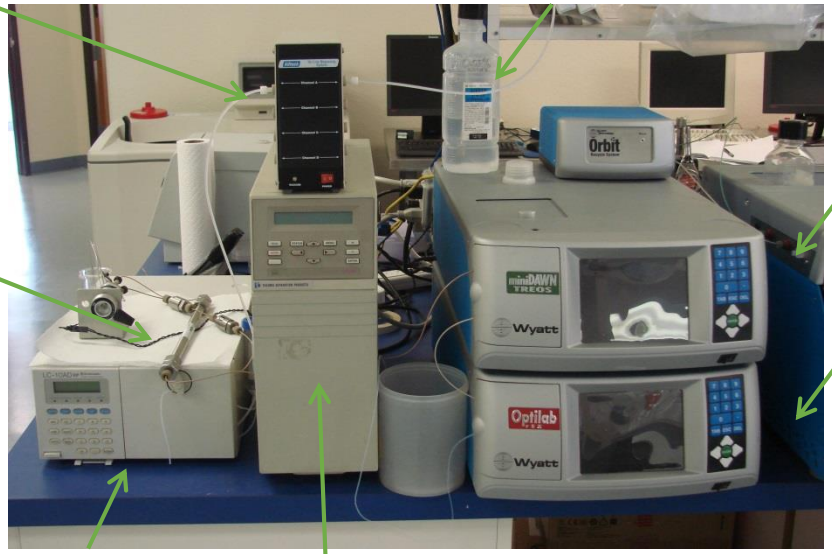
Protein	Source	R_s (nm)
Cobalamin	–	0.85
Cytochrome *c*	Horse heart	1.7
Ribonuclease II-A	Bovine pancreas	1.75
Myoglobin	Horse skeletal muscle	1.9
Haemoglobin	Human	2.4
Ovalbumin	Chicken egg	2.8
Alkaline phosphatase	*Escherichia coli*	3.3
Albumin	Bovine serum	3.5
Transferrin (iron-free)	Human	3.6
Aldolase	Rabbit muscle	4.6
Catalase	Ox liver	5.2
Aspartate transcarbamylase	*Escherichia coli*	6.0
Ferritin	Horse spleen	6.3
β-Galactosidase	*Escherichia coli*	6.9
Thyroglobulin	Ox thyroid	8.6

Principles of SEC



Kim, I., McKenna, S. A., Puglisi, E. V. & Puglisi, J. D. *Rapid purification of RNAs using fast performance liquid chromatography (FPLC)*. RNA 13, 289–294 (2007).

SEC-MALS



18+1+1=20 channels of data being collected every second!

SEC-MALS

$$\frac{Kc}{R(\theta, c)} = \frac{1}{M_w P(\theta)} + 2A_2 c$$

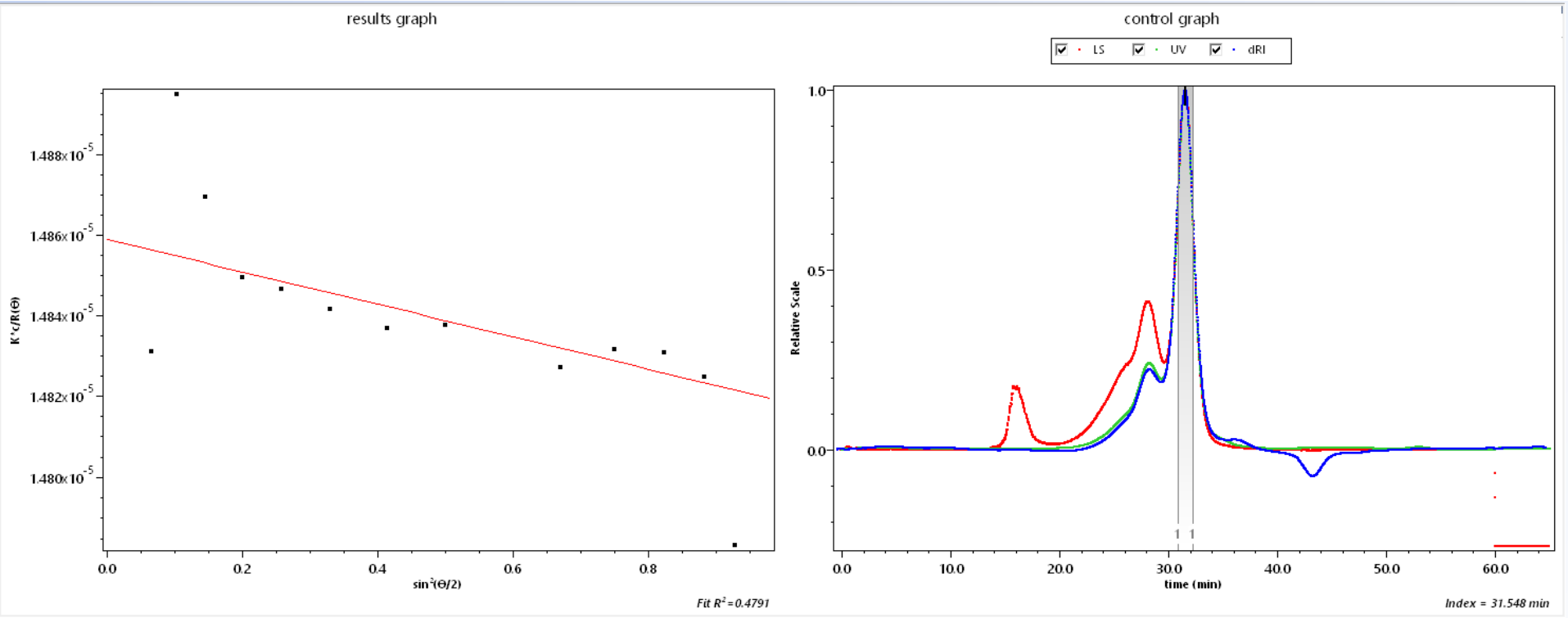
- K: constants
- c: concentration from UV or RI
- R: from measured intensity
- P: from scattering angle
- A_2: 2nd virial coefficient (fit)

$$M_w = \frac{\sum c_i m_i}{\sum c_i}$$

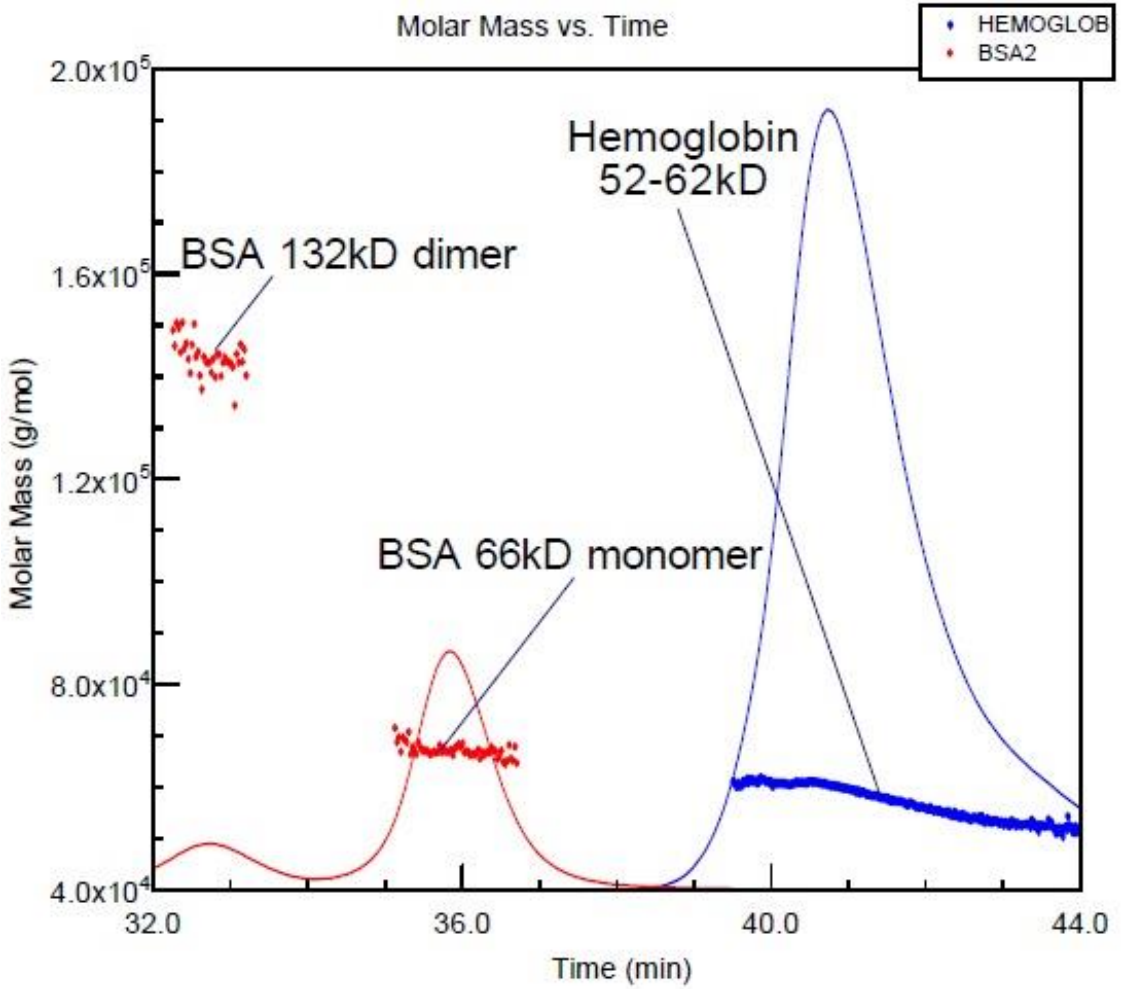
(weight-average molecular weight; mass concentrations)

- global fit of M_w values with 18 R(θ), P(θ) values
- calculation is done "on-the-fly", every few seconds of SEC column elution
- larger particles have stronger scattering signal; can use less material
- need ~20-100 uL of concentrated protein; gets diluted on column

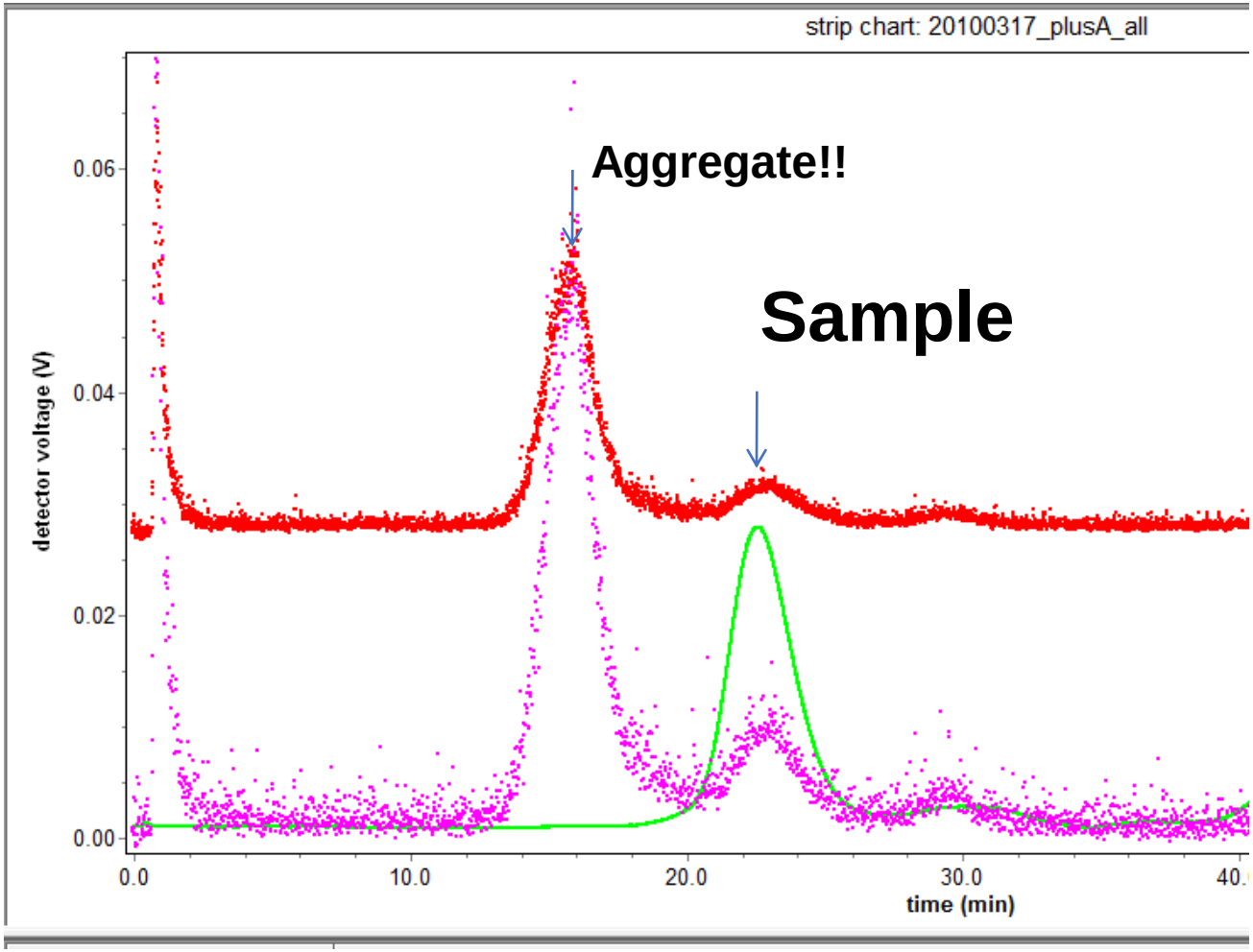
SEC-MALS


results graph
1.488x10^-5
1.486x10^-5
1.484x10^-5
1.482x10^-5
1.480x10^-5
K*c/R(Θ)
0.0
0.2
0.4
0.6
0.8
sin²(Θ/2)
Fit R²=0.4791
control graph
LS
UV
dRI
1.0
0.5
0.0
Relative Scale
0.0
10.0
20.0
30.0
40.0
50.0
60.0
time (min)
Index = 31.548 min

SEC-MALS


Molar Mass vs. Time
HEMOGLOB
BSA2
Hemoglobin
52-62kD
BSA 132kD dimer
BSA 66kD monomer
Molar Mass (g/mol)
2.0x10^5
1.6x10^5
1.2x10^5
8.0x10^4
4.0x10^4
32.0
36.0
40.0
44.0
Time (min)

SEC-MALS

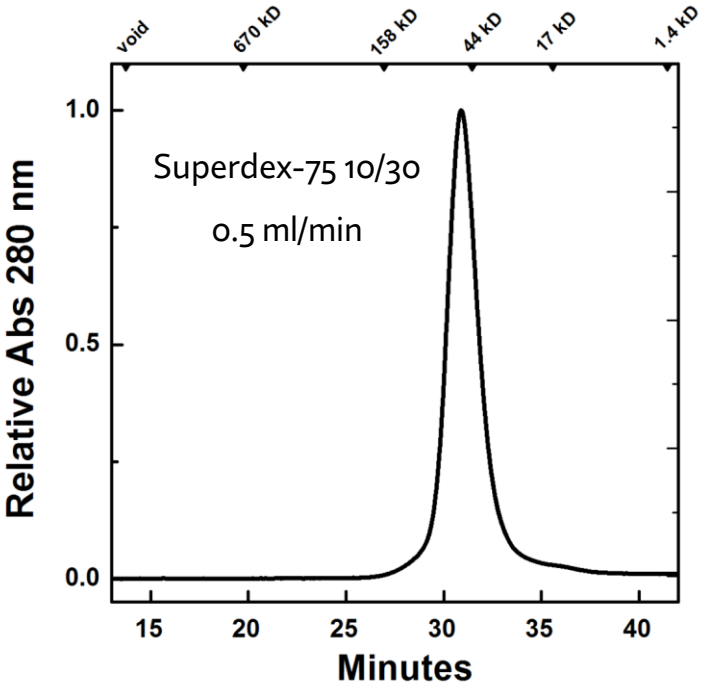


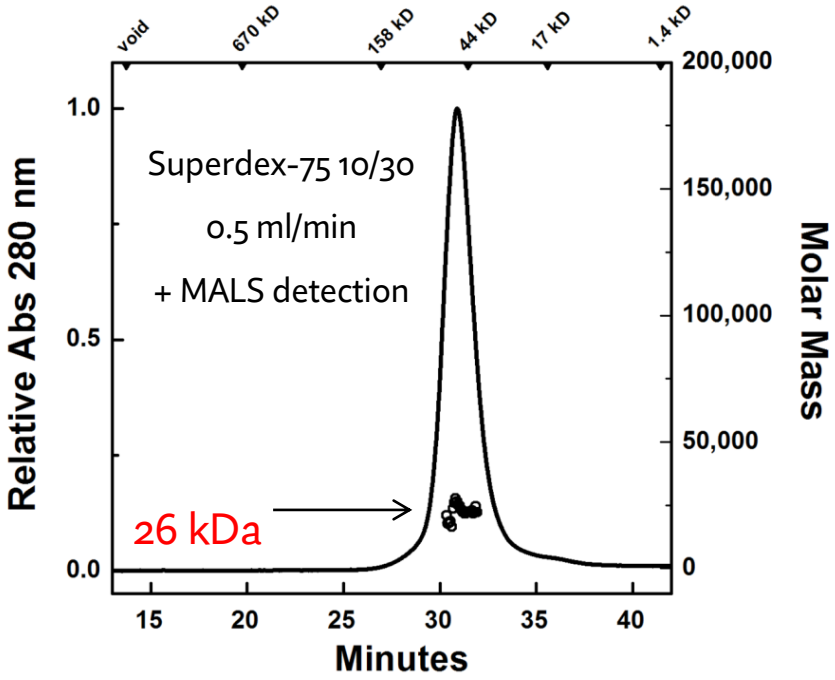
Example: Preparation of Group II Intron from Lactococcus Lacti.

SEC-MALS

What is M for this protein?

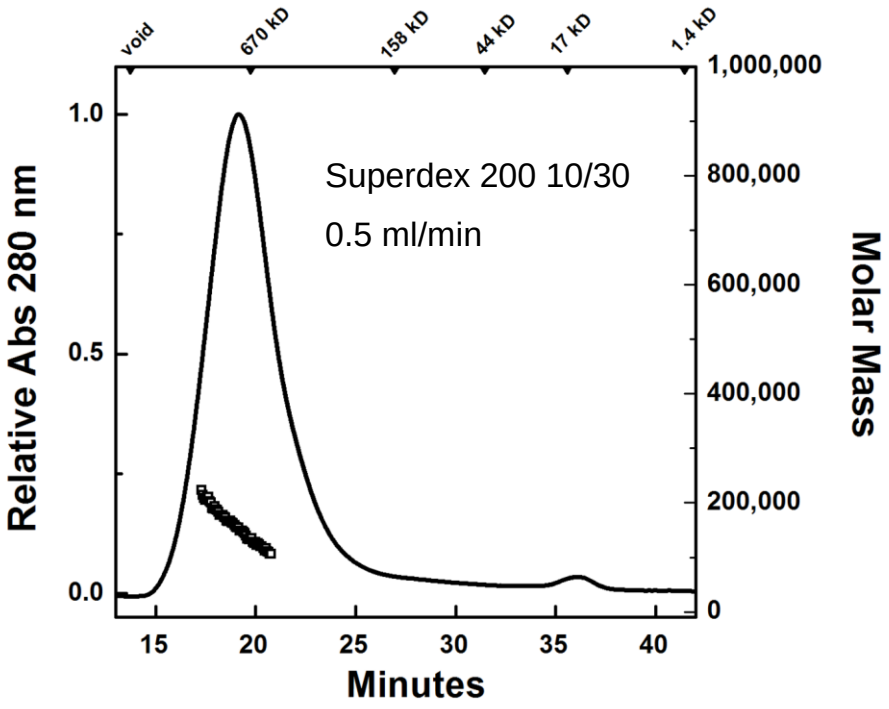
If M=26 kDa, what is the oligomeric state? Monomer





SEC-MALS

60 kDa heterodimer



Gupta et al 2015, JBC

You can often establish that your protein is monomeric from a SEC experiment alone. You can not conclude anything about your protein's oligomeric state from SEC alone!

Advanced Data Analysis

- SEC and SEC-MALS
- **SEC-SAXS**
  - **SVD-EFA**
- *Ab Initio* Reconstructions
  - Calculating Profiles
  - DAMMIN/F
  - GASBOR
  - DENSS
  - MONSA

SEC-SAXS:Suggested Reading

• Meisburger SP, Taylor AB, Khan CA, Zhang S, Fitzpatrick PF, Ando N. Domain Movements upon Activation of Phenylalanine Hydroxylase Characterized by Crystallography and Chromatography-Coupled Small-Angle X-ray Scattering. J Am Chem Soc. 2016 May 25;138(20):6506-16. doi: 10.1021/jacs.6b01563. Epub 2016 May 12. PMID: 27145334

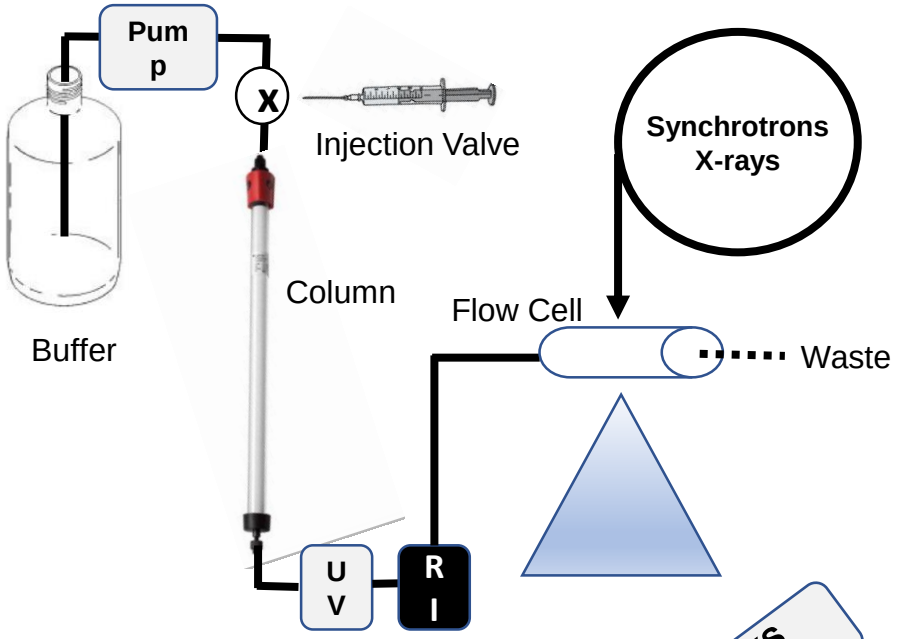
• Mathew, E., Mirza, A. & Menhart, N. Liquid-chromatography-coupled SAXS for accurate sizing of aggregating proteins. J Synchrotron Radiat 11, 314–318 (2004).

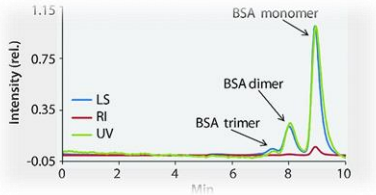
• Malaby, A. W. et al. Methods for analysis of size-exclusion chromatography–small- angle X-ray scattering and reconstruction of protein scattering. J Appl Crystallogr 48, 1102–1113 (2015).

• Brookes, E. et al. Fibrinogen species as resolved by HPLC-SAXS data processing within the UltraScan Solution Modeler (US-SOMO) enhanced SAS module. J Appl Crystallogr 46, 1823–1833 (2013).

• Hopkins, J.B., Gillilan, R.E., Skou, S.BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. J Appl Cryst, J Appl Crystallogr 50, 1545–1553 (2017).

SEC-SAXS

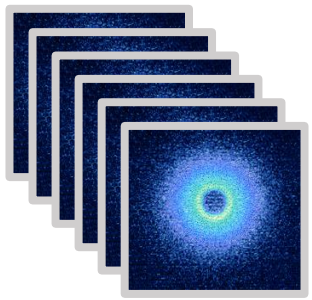


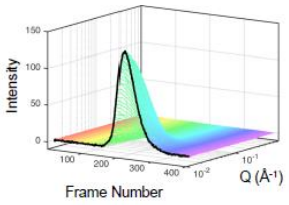


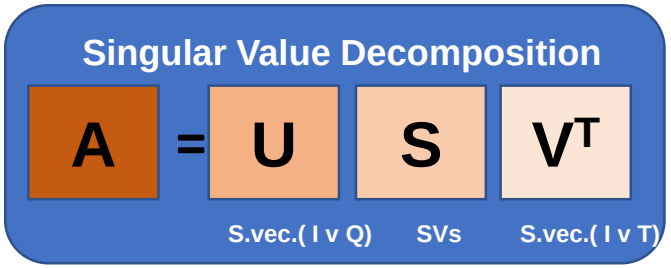
Chromatogram

WAXS Detector

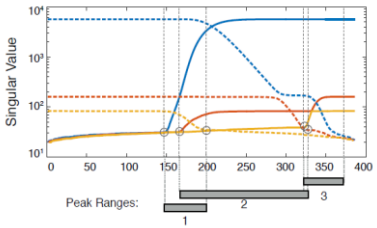
SAXS Detector



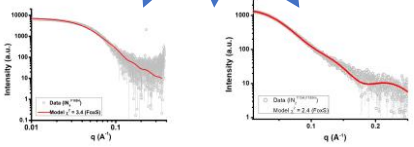




SVD-EFA: Meisburger et al 2016 JACS
RAW: Hopkins et al 2017 J App Cryst



Evolving Factor Analysis



Dimer

Tetramer

SEC-SAXS

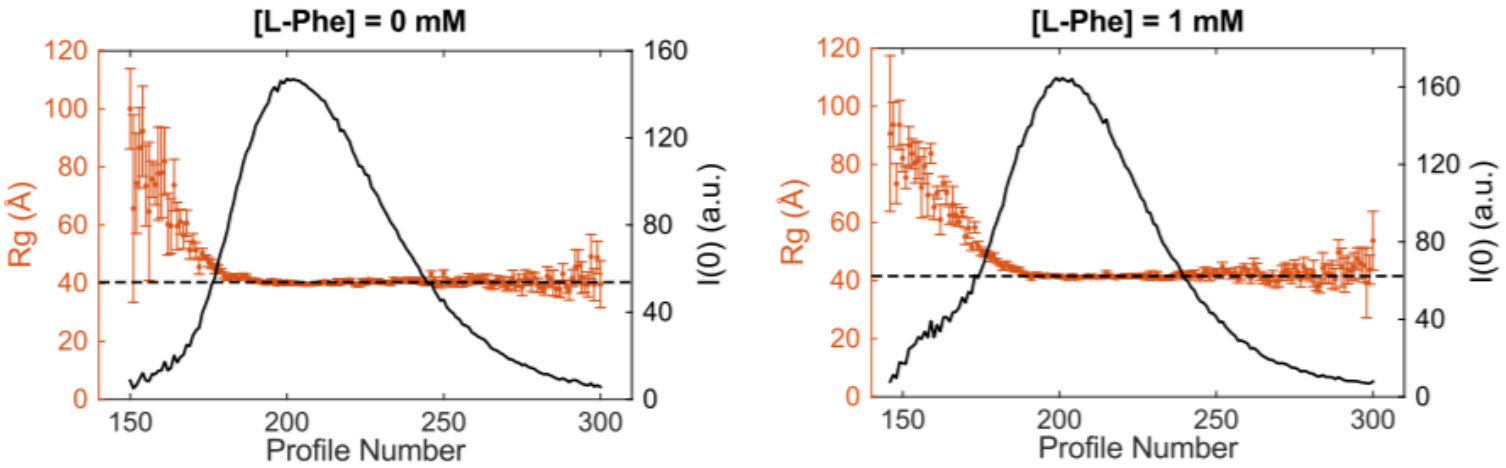


Figure S9: R_g and $I(0)$ across the elution peak of wt-PheH with and without L-Phe. Although only a single elution peak is observed in each case, the R_g values determined by Guinier analysis vary, suggesting the presence of multiple species. While the low signal-to-noise of the profiles at the beginning of the run complicates Guinier analysis, it clear that the R_g is significantly higher than 40 Å (dotted line).

Meisburger SP, Taylor AB, Khan CA, Zhang S, Fitzpatrick PF, Ando N. Domain Movements upon Activation of Phenylalanine Hydroxylase Characterized by Crystallography and Chromatography-Coupled Small-Angle X-ray Scattering. J Am Chem Soc. 2016 May 25;138(20):6506-16. doi: 10.1021/jacs.6b01563. Epub 2016 May 12. PMID: 27145334

SEC-SAXS: SVD

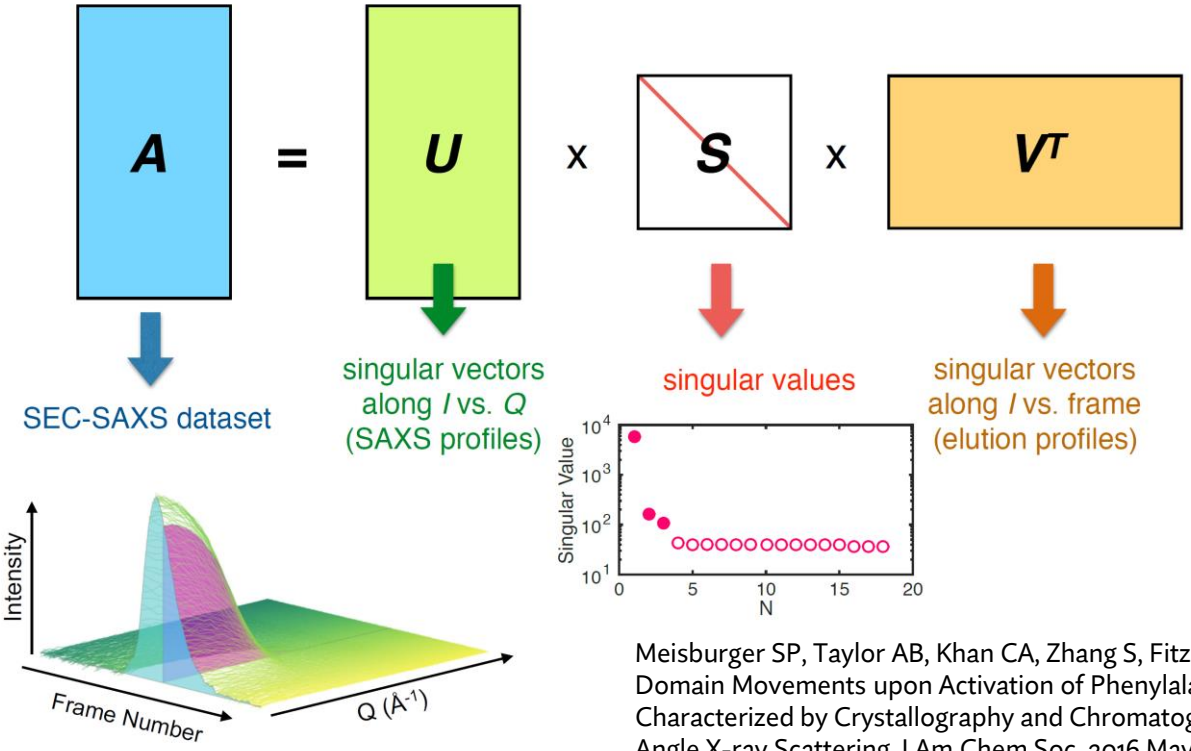
Applications for biological studies:

- ❑ Oligomerization
- ❑ Multiple assembly forms
- ❑ Temperature dependent transitions
- ❑ Ligand-dependent transitions

Polydisperse & interactive systems:

- ❑ Equilibrium oligomeric mixtures (OLIGOMER, COSMIC)
- ❑ Assembly/disassembly processes (SVDPLOT, MIXTURE)
- ❑ Natively unfolded proteins and multidomains proteins with flexible linkers (EOM, SASSIE, BILBOMD)
- ❑ SVD-EFA (RAW), US-SOMO – SEC-SAXS

SEC-SAXS: SVD



Meisburger SP, Taylor AB, Khan CA, Zhang S, Fitzpatrick PF, Ando N. Domain Movements upon Activation of Phenylalanine Hydroxylase Characterized by Crystallography and Chromatography-Coupled Small-Angle X-ray Scattering. J Am Chem Soc. 2016 May 25;138(20):6506-16. doi: 10.1021/jacs.6b01563. Epub 2016 May 12. PMID: 27145334

The number of significant singular vectors in SVD (i.e. non-random curves with significant singular values) yields the minimum number of independent curves required to represent the entire data set by their linear combinations (e.g. for mixtures). (Konarev)

SEC-SAXS: SVD-EFA

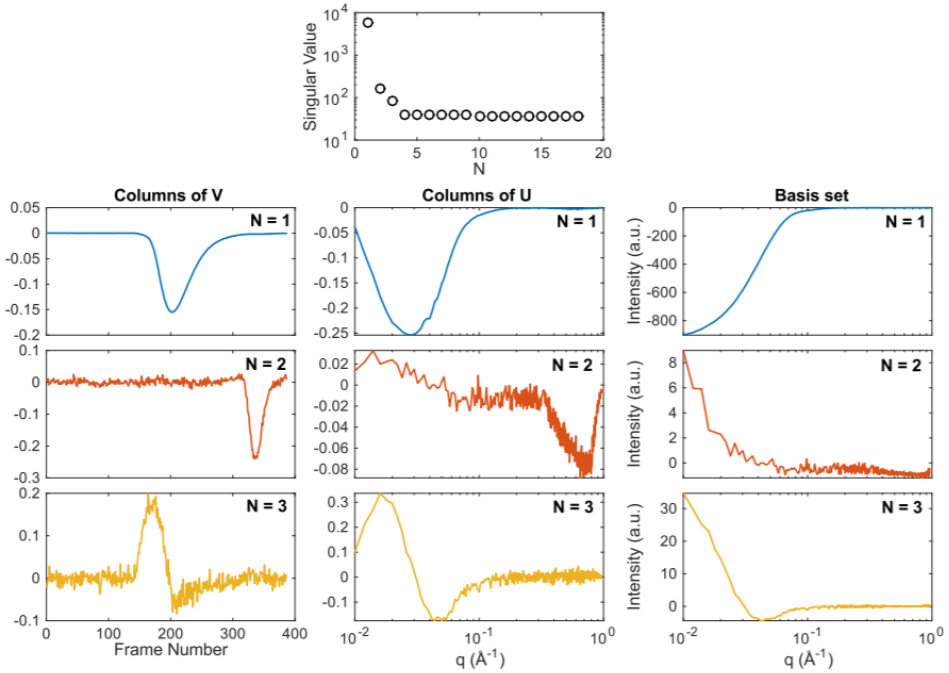
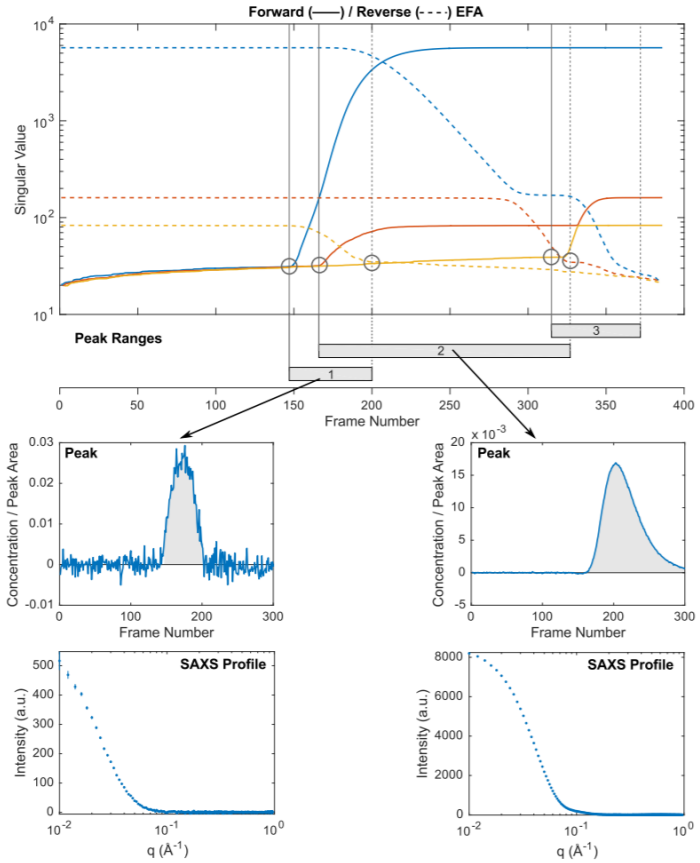


Figure S10: Conventional SVD of SEC-SAXS data from wt-PheH in 0 mM L-Phe. Three significant singular values are observed (top panel). Although the corresponding right singular vectors (columns of V) have shapes that are reminiscent of elution peaks, we find that there are sign changes within the curves. Moreover, when the scattering basis set is recovered by multiplying the columns of U with the experimental error, we find that there are non-physical sign changes within the curves. Because SVD produces orthogonal singular vectors, they cannot represent physical states, such as elution peaks and scattering intensities, which must be positive numbers.

Meisburger SP, Taylor AB, Khan CA, Zhang S, Fitzpatrick PF, Ando N. Domain Movements upon Activation of Phenylalanine Hydroxylase Characterized by Crystallography and Chromatography-Coupled Small-Angle X-ray Scattering. J Am Chem Soc. 2016 May 25;138(20):6506-16. doi: 10.1021/jacs.6b01563. Epub 2016 May 12. PMID: 27145334

SEC-SAXS: SVD-EFA



Evolving Factor Analysis
(Maeder M. Anal Chem. 1987;59:527–530)

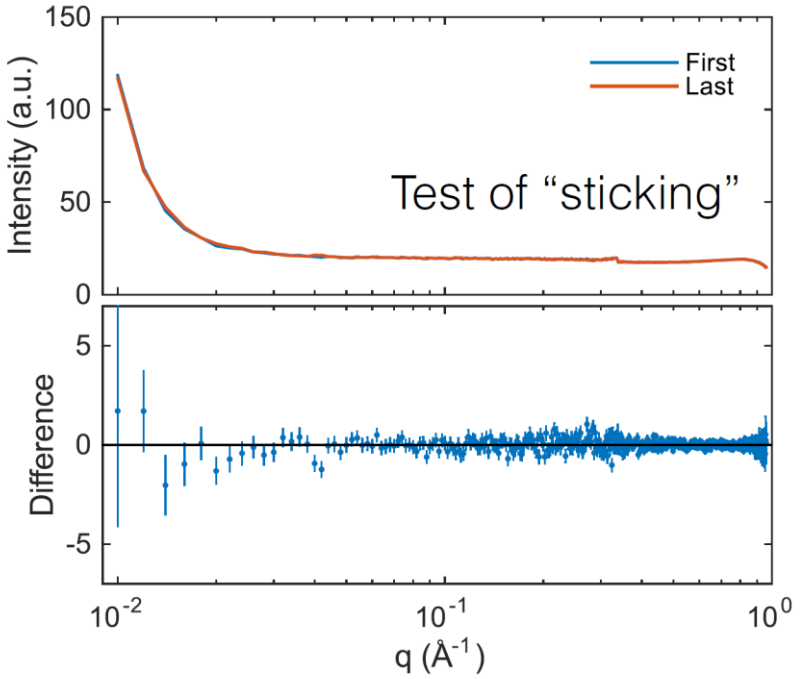
Variant of SVD that allows for the identification of ranges within each elution peak where separate species elute from the abrupt changes in the number of significant singular values as scattering profiles are added or removed from the matrix.

Meisburger SP, Taylor AB, Khan CA, Zhang S, Fitzpatrick PF, Ando N. Domain Movements upon Activation of Phenylalanine Hydroxylase Characterized by Crystallography and Chromatography-Coupled Small-Angle X-ray Scattering. J Am Chem Soc. 2016 May 25;138(20):6506-16. doi: 10.1021/jacs.6b01563. Epub 2016 May 12. PMID: 27145334

SEC-SAXS: Troubleshooting

1. Radiation Damage: sticking to windows
    1. Flow Rate increase
    2. Add glycerol
    3. Attenuation

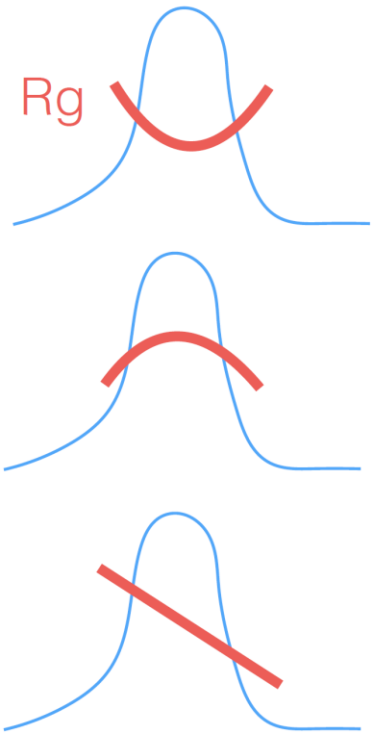
2. Column Resolution
    1. Overloading
    2. Injection volume too large
    3. Media interactions



Meisburger SP, Taylor AB, Khan CA, Zhang S, Fitzpatrick PF, Ando N. Domain Movements upon Activation of Phenylalanine Hydroxylase Characterized by Crystallography and Chromatography-Coupled Small-Angle X-ray Scattering. J Am Chem Soc. 2016 May 25;138(20):6506-16. doi: 10.1021/jacs.6b01563. Epub 2016 May 12. PMID: 27145334

SEC-SAXS: Troubleshooting

3. Rg changes within Peak (also a problem with SEC-MALS)



- **Smiling:**
  - repulsive inter-particle interference.
  - Consider lowering sample concentration, adding salt to the buffer.
- **Frowning:**
  - aggregating post-column
  - attractive structure factor (stickiness)
  - weakly-associated oligomer
- **Sloping**: (especially if peak is also broad)
  - exchanging oligomers
  - overlapping peaks from multiple species

Steven Meisburger

SEC-SAXS-MALS

DNA Repair 65 (2018) 11–19

Contents lists available at ScienceDirect

DNA Repair

journal homepage: www.elsevier.com/locate/dnarepair

The C-terminal tail of the NEIL1 DNA glycosylase interacts with the human mitochondrial single-stranded DNA binding protein

Nidhi Sharma[a], Srinivas Chakravarthy[b], Matthew J. Longley[c], William C. Copeland[c], Aishwarya Prakash[a,*]

[a] University of South Alabama, Mitchell Cancer Institute, 1660 Springhill Avenue, Mobile, AL 36604, United States
[b] Illinois Institute of Technology, Advanced Photon Source, Bldg. 435B/Sector 18, 9700 S. Cass Avenue, Argonne, IL 60439-4860, United States
[c] Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, 111 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States

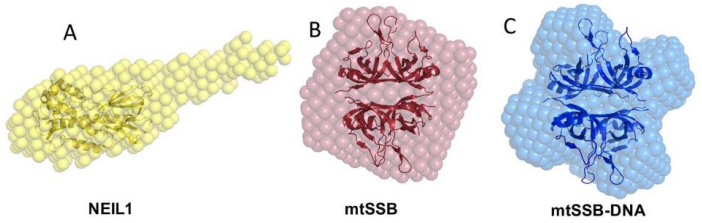


Table 2
SAXS data collection parameters.

	mtSSB (tetramer)	*NEIL1*	*mtSSB-DNA complex*
(A) Sample details			
Organism	*Homo sapiens*	*Homo sapiens*	*Homo sapiens*
Source	*E. coli*	*E. coli*	*E. coli*
UniProt sequence ID (residues in constructs)	Q04837 (17–148)	Q96FI4 (1–390)	
Extinction coefficients [A280, 0.1%(w/v)]	1.313	0.725	
SEC–SAXS column, 10/300 mm Superdex S200			
Loading concentration (mg ml^{-1})	7	11	8
Injection volume (μl)	300	300	300
Flow rate (ml/min)	0.75	0.75	0.75
Solvent (blanks taken from SEC flow through prior to elution of protein)	25 mM HEPES pH 7.4, 5% glycerol, 300 mM NaCl, and 1 mM		
(B) Data collection parameters			
Beamline	APS	APS	APS
Wavelength (Å)	1.03	1.03	1.03
Q Range ($Å^{-1}$)	0.0059–0.3892	0.0059–0.3898	0.0058–0.3587
Temperature (°C)	25	25	25
(C) Softwares used for data reduction, analysis and interpretation			
SAXS data reduction	PRIMUS (ATSAS 2.8.1)		
Extinction coefficient estimate	ProtParam (Expasy) [83]		
Basic analyses: Guinier, P(r), MW (V_p, V_c)	PRIMUS (ATSAS 2.8.1), Scatter [43], BioXTAS RAW [47]		
Shape/bead modeling	DAMMIF and DAMAVER (ATSAS 2.8.1)		
(C) Structural parameters			
I(0) from *P*(r)	44.88	51.49	22.74
R_g (Å) from *P*(r)	27.52	37.15	28.18
I(0) from Guinier	44.83 ± 0.10	49.20 ± 0.24	22.77 ± 0.03
R_g from Guinier	27.44 ± 1.21	33.04 ± 2.46	28.40 ± 0.86
Dmax (Å)	103.21	149.96	98.74
(D) Molecular weight determination (kDa)			
Expected Theoretical (Expasy)	60.78	44.72	78
MW(V_p)	61	46	79
MW(V_c)	59	36	71
MW (MALS)	59.8 ± 1.67%	48.5 ± 2.23%	69 ± 1.46%
(E) Modeling parameters			
Symmetry	P4	P1	P4
Particle anisometry	Oblate	Unknown	Oblate
# of modeling iterations	10	10	10
X^2 of the model	1.072	1.041	0.945

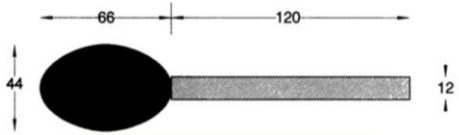
Advanced Data Analysis

- SEC and SEC-MALS
- SEC-SAXS
  - SVD-EFA
- ***Ab Initio* Reconstructions**
  - Calculating Profiles
  - DAMMIN/F
  - GASBOR
  - DENSS
  - MONSA

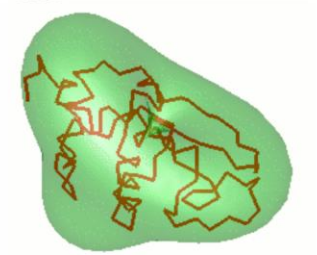
Ab Initio Reconstructions

Simple shapes

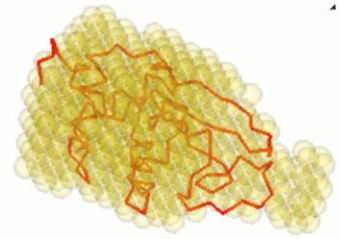
Collections of simple shapes



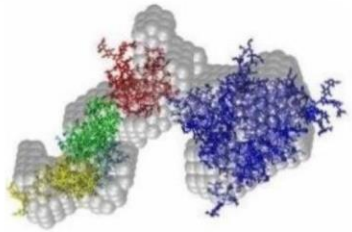
Spherical harmonics envelopes

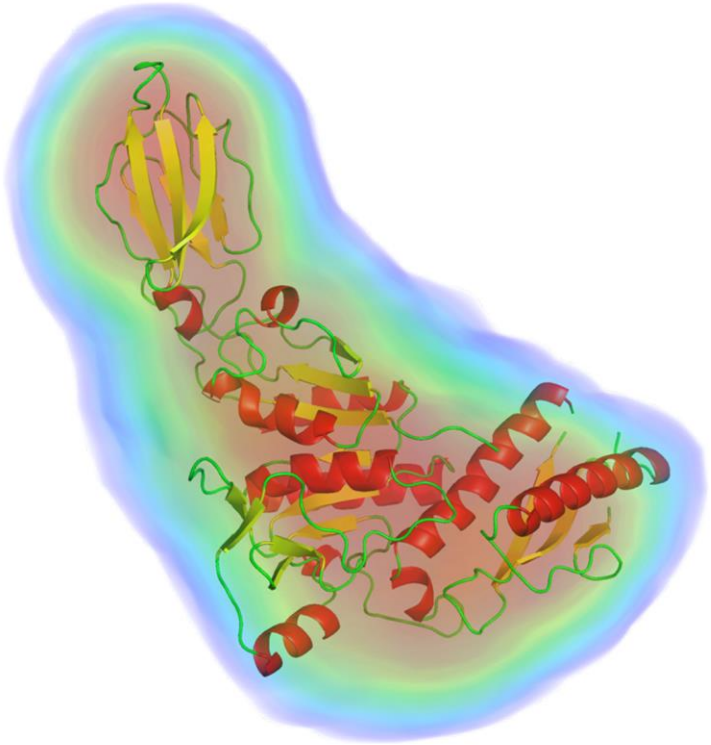


Bead modeling



Hybrid modeling





Direct Electron Density Calculation

Thomas Grant

Calculating SAXS profiles from Models

- Debye equation:

$$I(q) = \sum_i \sum_j f_i(q) f_j(q) \frac{\sin\left(q \cdot r_{ij}\right)}{q \cdot r_{ij}}$$

- ❑ Calculation of reciprocal space scattering from model
- ❑ Accomplished vis Debye Equation or Spherical Harmonics Approximation
- ❑ Modeling solvent shell and excluded volume is the technical challenge

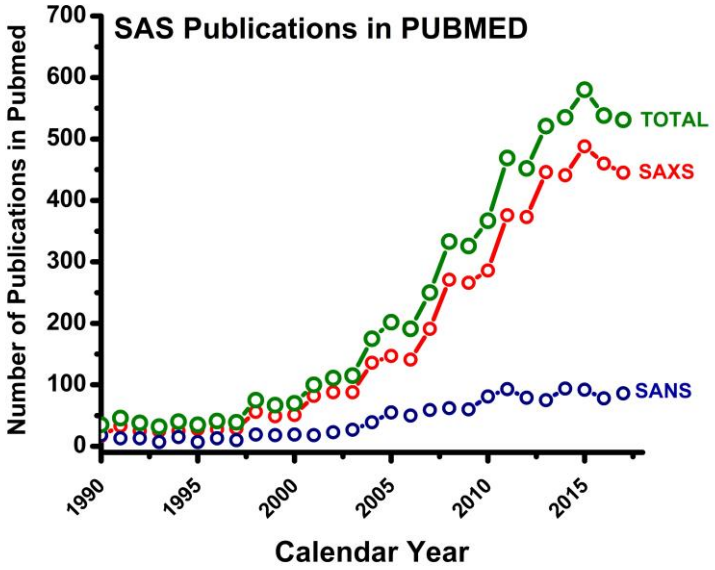
Calculating SAXS profiles from Models

Table 1 Methods for theoretical profile calculation

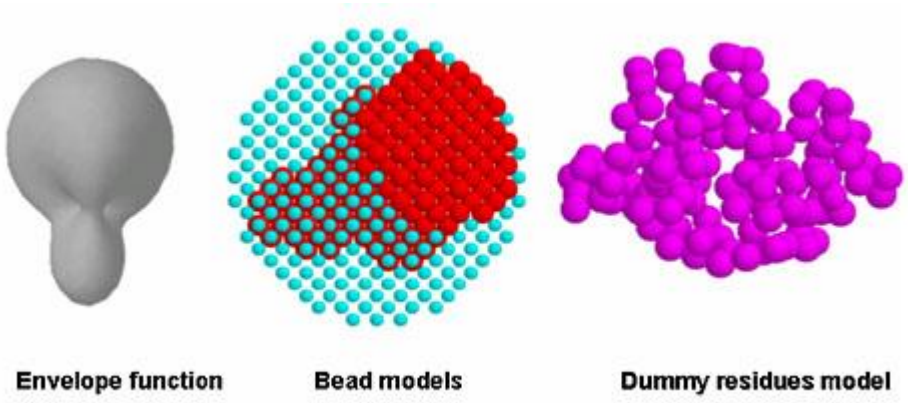
Method	Spherical Averaging	Hydration Layer	Representation	Availability
CRYSOL [34]	Multipole expansion	Implicit water layer based on envelope function	Atomic	Server, download http://www.embl-hamburg.de/biosaxs/crysol.html
solX [35]	Debye formula	-	Atomic	
ORNL_SAS [36]	Monte-Carlo sampling	Implicit water layer	Grid representation	Download http://www.ornl.gov/sci/csd/Research_areas/MS_csmb_comp_methods.htm
SoftWAXS [37]	Numerical quadrature	Implicit water layer	Atomic	
Fast-SAXS [38]	Debye formula	Explicit placement of water molecules	Coarse-grained residue level	http://yanglab.case.edu/software.html
Park et al. [39]	Spherical quadrature	Explicit placement of water molecules	Atomic	
Stovgaard et al. [40]	Debye formula	-	Coarse-grained, 1 or 2 points per-residue	
AXES [41]	Numerical quadrature	Explicit placement of water molecules	Atomic	Server http://spin.niddk.nih.gov/bax/nmrserver/saxs1/
FoXS [42]	Debye formula	Implicit water layer based on surface accessibility	Atomic or coarse-grained residue level	Source code, server, download, Chimera http://salilab.org/foxs/
AquaSAXS [43]	Cubature formula	AquaSol solvent density map	Atomic	Server http://lorentz.dynstr.pasteur.fr/aquasaxs/aquasaxs_submission.php
Virtanen et al. [44]	Debye formula or Cube model	HyPred based on MD simulations	Atomic, MD simulation	
Zernike Polynomials [45]	Zernike polynomial expansions	Hydration layer from voxelized representation	Atomic	Source code, server, download http://sastbx.als.lbl.gov/cgi-bin/intensity.html

Schneidman-Duhovny, et.al. (2012) *BMC Structural Biology*

Ab Initio Reconstructions



Shape Reconstruction



- Synchrotron/Reactor Sources
- Home Source Rotating Anodes
- Detectors
- Software and Computing Power
- Algorithms for Shape Reconstruction

Spherical Harmonics (Envelope Function)
Stuhrmann, 1970; Svergun & Stuhrmann, 1991; Svergun et al., 1996

Bead Models (DAMMIN, MONSA)
Chacon et al., 1998; Svergun, 1999; Walther, Cohen & Doniach, 2000
Svergun, Petoukhov & Koch, 2001

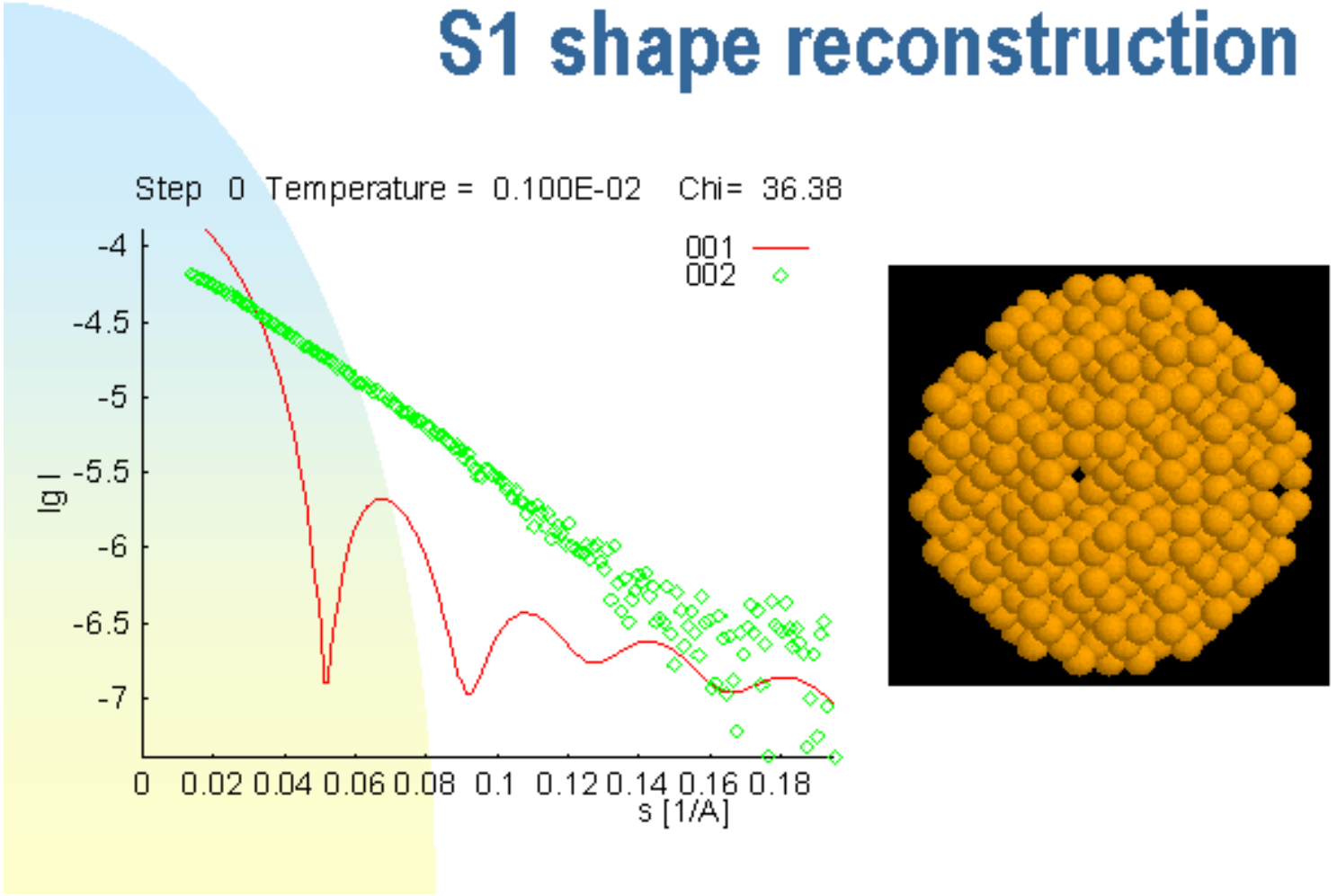
Dummy Residue Models (GASBOR)
Svergun, Petoukhov & Koch, 2001

http://www.embl-hamburg.de/workshops/2008/embo/d-sv_abinitio.html

Ab Initio Reconstructions:DAMMIF/N

- DAMMIN/F uses a dummy atom/bead modeling approach that applies spherical harmonics
- Relies on low resolution data where s*Rg <7-8 or q< 0.3, as contributions from solvent at high q can lead to errors
- 3D volume fits the data with physical constraints applied
- Penalties for envelopes that are loose, compact, or disconnected
- Simulated Annealing Method – each calculation is slightly different.

Ab Initio Reconstructions:DAMMIF/N



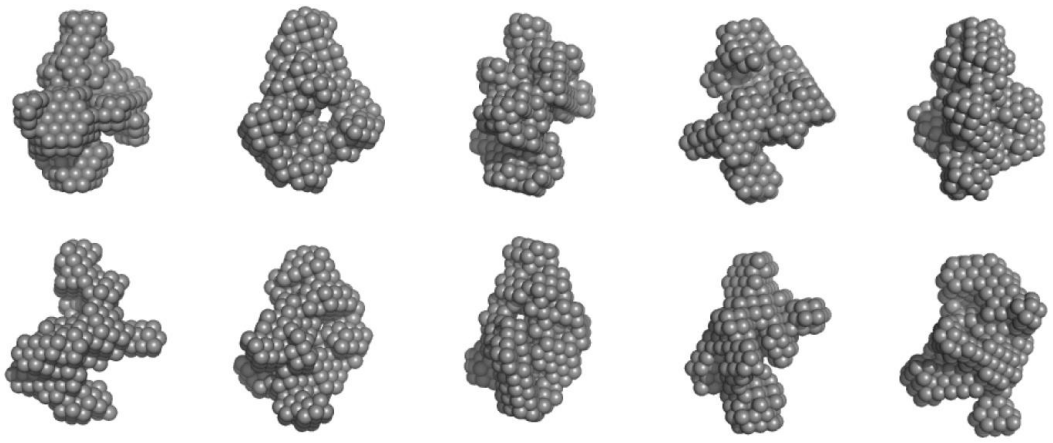
http://www.embl-hamburg.de/workshops/2008/embo/d-sv_abinitio.html

Ab Initio Reconstructions:DAMMIF/N

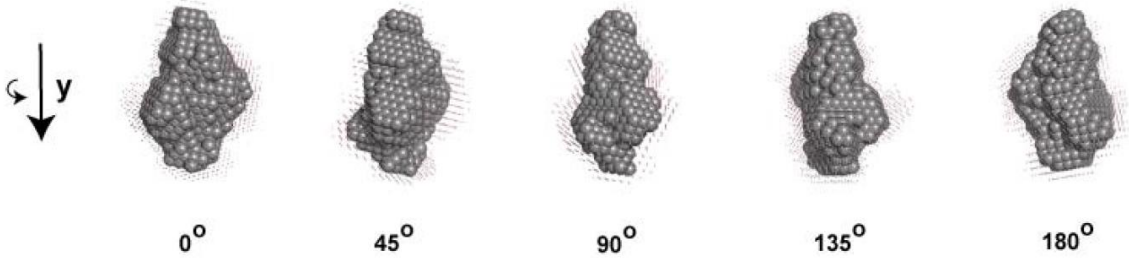
Process:

- DAMMIN/F: Calculate 5-20 ab initio reconstructions
- DAMSUP/DAMAVER/DAMFILT:
- Models are aligned
- Normalized Spacial Discrepancy (NSD) calculated:
  - ~Avg of 0.5 implies good stability
  - ~0.7-0.9 implies fair stability, but more common with anisotropic particles
  - ~>1.0 implies poor stability
- Ensemble then averaged and filtered to yield final result

Ab Initio Reconstructions:DAMMIF/N

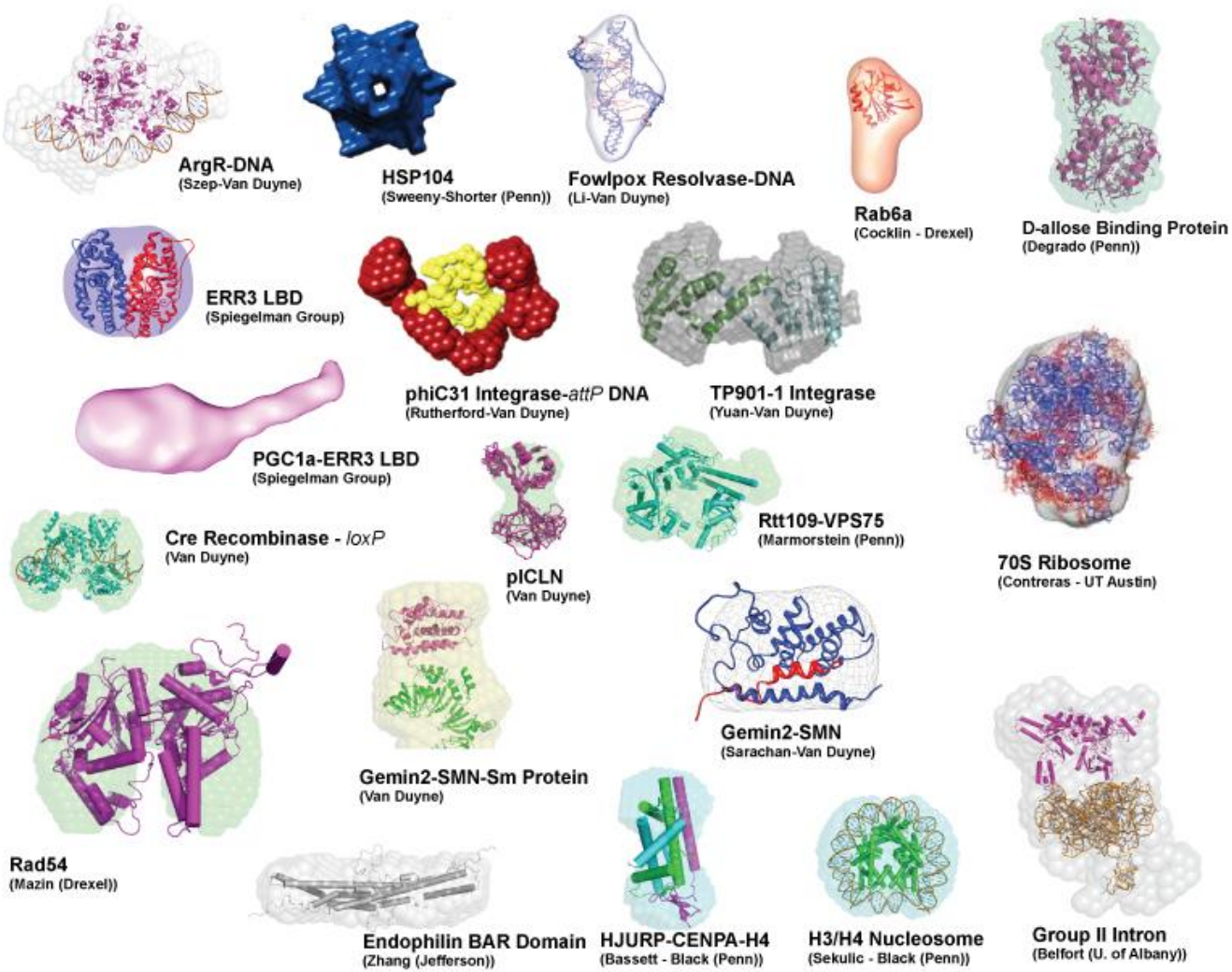


Gallery of ten DAMMIN shape reconstructions for [redacted] bound to 60-mer Holliday Junction, rendered as bead models with a sphere radius of 3.25 Å. The overall normalized spatial discrepancy (NSD) for these ten reconstructions range range from 0.70 to 0.76, indicating a stable solution.

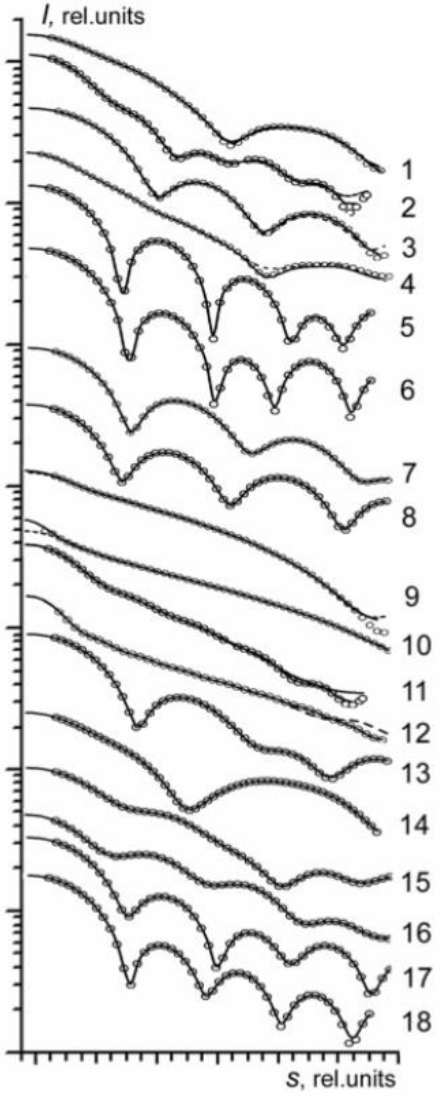


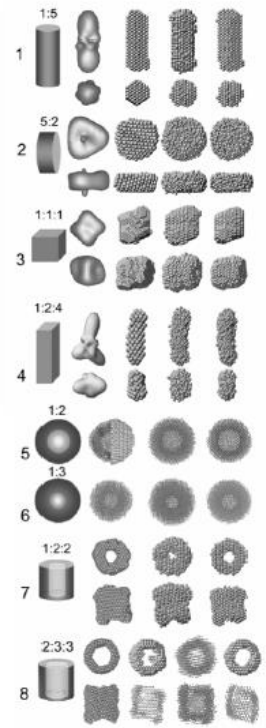
Final averaged and filtered reconstruction is rendered in grey spheres with a radius of 3.25 Å, rotated around the y axis. For reference, the averaged envelope without filtering is shown as small spheres. The shape has dimensions of 104.0 Å x 68.3 Å x 64.4 Å. According to HYDROPRO analysis, the shape has a predicted Stokes radius of 38 Å and a predicted S value of 5.3, largely consistent with DLS (R_s=35 Å) and Sedimentation velocity measurements ($S_{20,w}$ of 4.4).

Ab Initio Reconstructions: DAMMIF/N and GASBOR Examples

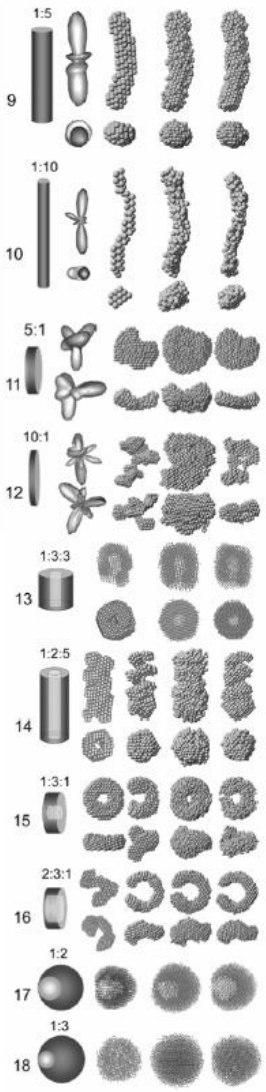


Ab Initio Reconstructions: Uniqueness



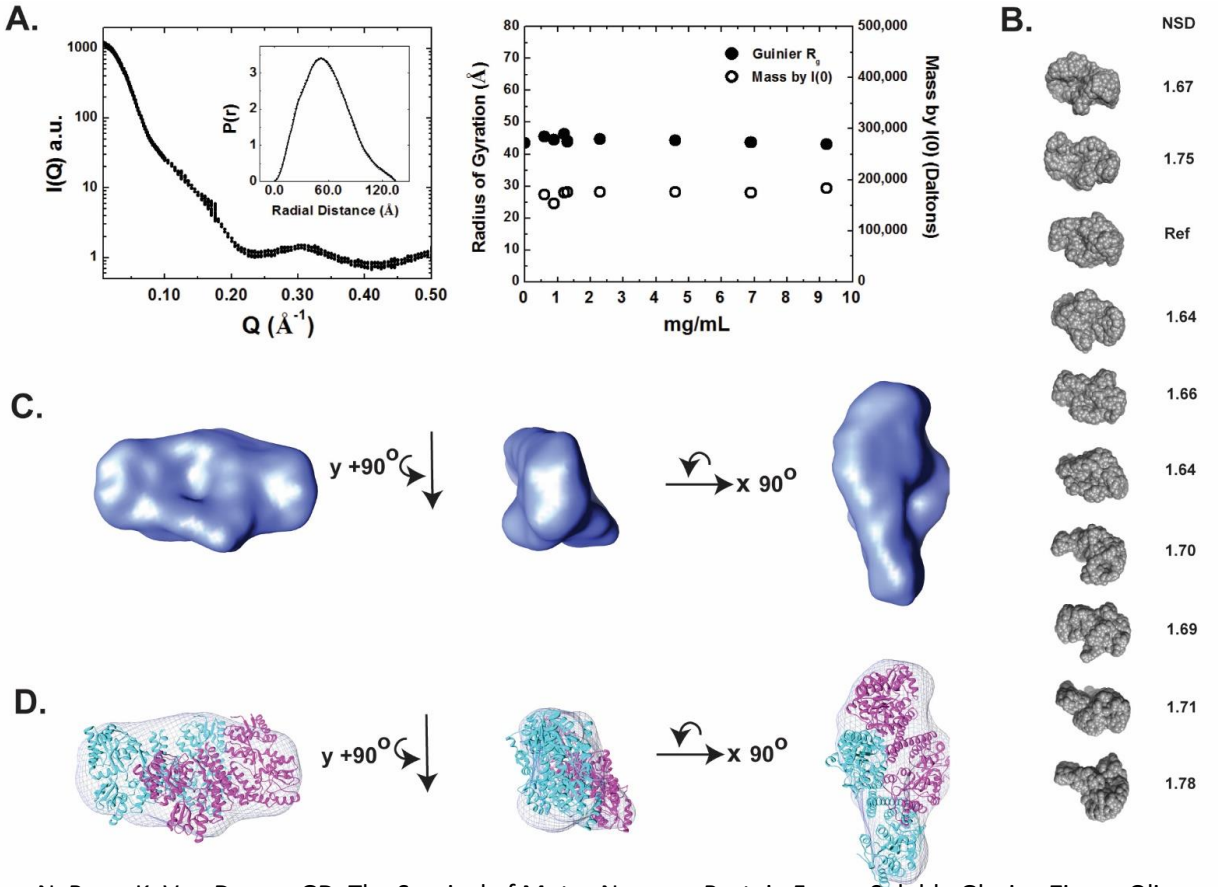


Volkov, V. V. and D. I. Svergun (2003). "Uniqueness of ab initio shape determination in small-angle scattering." Journal of Applied Crystallography **36: 860-864.**



Ab Initio Reconstructions: GASBOR example

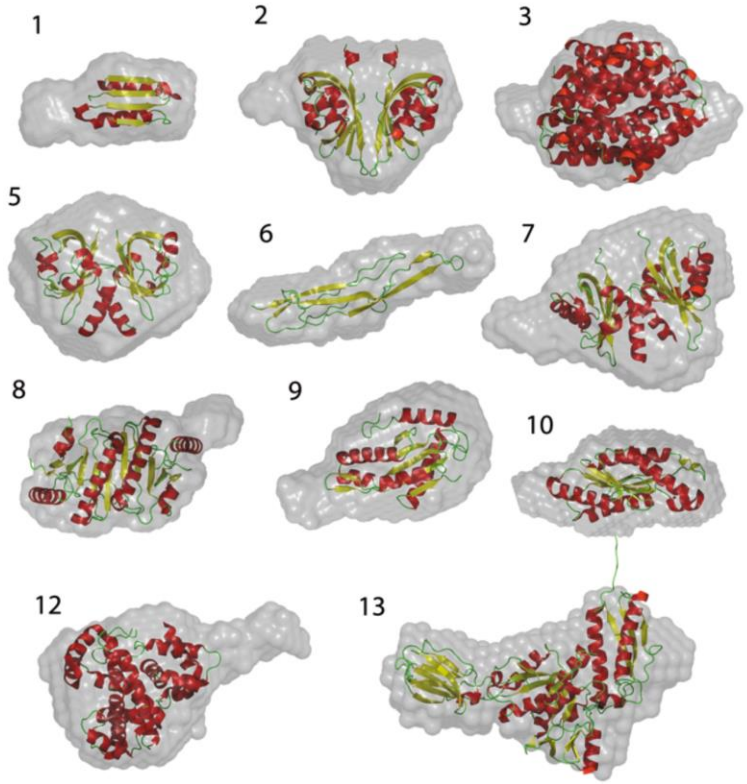
- GASBOR uses dummy residue approach with explicit solvent models, which then allows for use of higher resolution data.



Martin R, **Gupta K**, Ninan N, Perry K, Van Duyne, GD. The Survival of Motor Neurons Protein Forms Soluble Glycine Zipper Oligomers. Structure 20(11):1929-39 (2012).

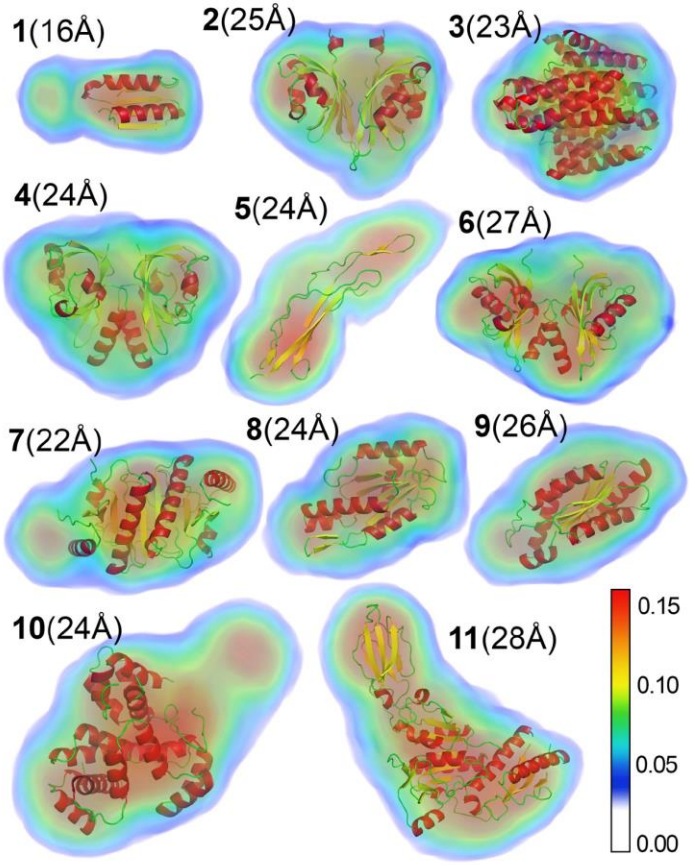
Ab Initio Reconstructions: DENSS

Bead modeling (uniform density)



Grant, et.al. (2011). *Biopolymers*, **95**, 517-530.

Direct Electron Density Calculation

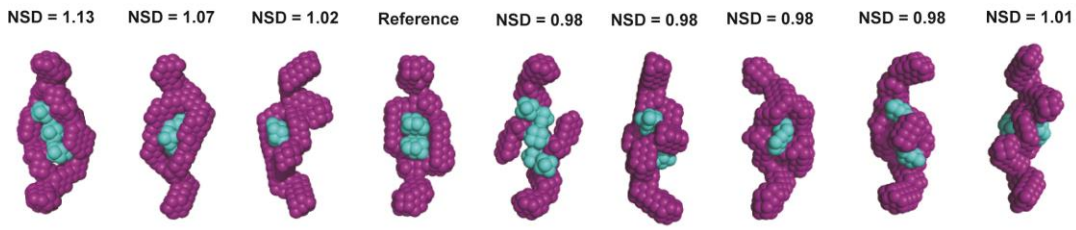


Grant, T.D. (2018) *Nature Methods.*

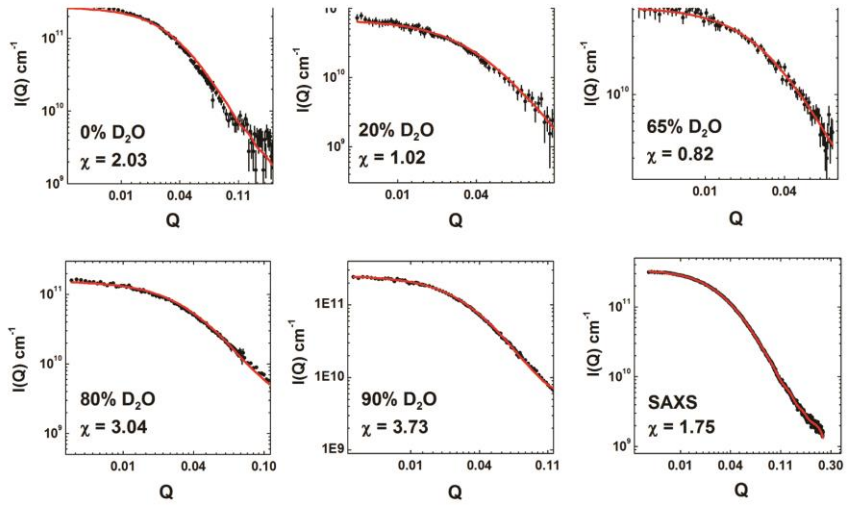
Ab Initio Reconstructions: MONSA

D.I. Svergun (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879-2886.

A.



B.

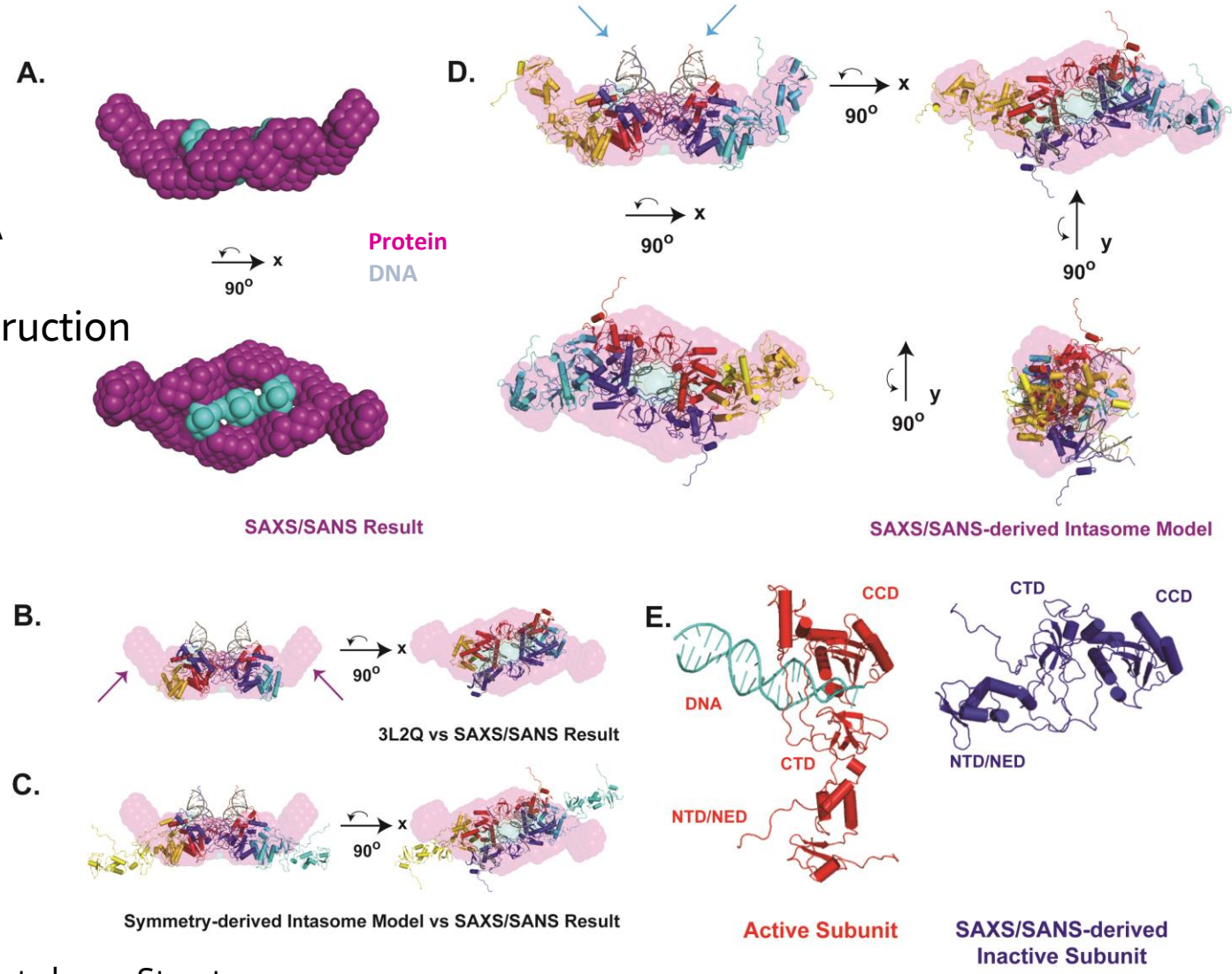


- Multiple Profiles
  → Increased Data-to-Parameters
- Experimental Contrast Information
- Simultaneous Solution

Gupta et al 2012 Structure

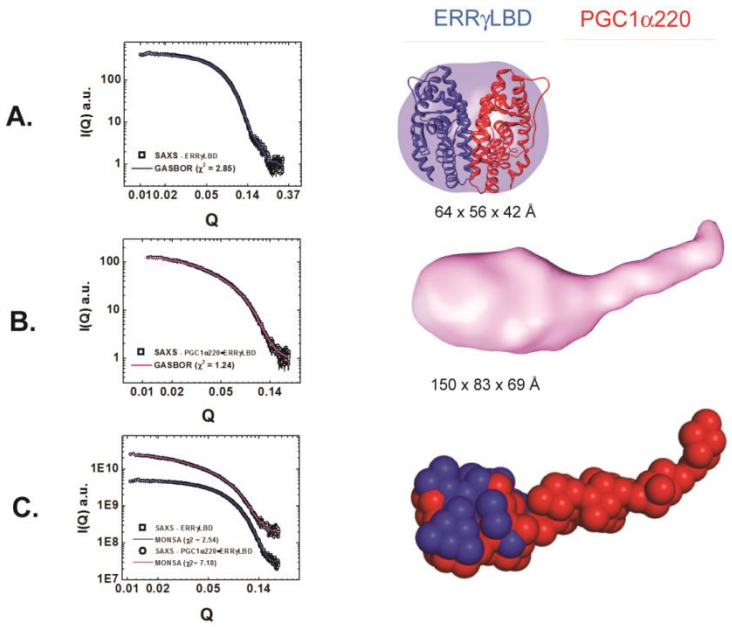
Ab Initio Reconstructions: MONSA

MONSA Shape Reconstruction



Gupta et al 2012 Structure

Ab Initio Reconstructions: MONSA



Supplemental Table S3 – Comparison of the hydrodynamic parameters from shape reconstructions and experimentally determined values

	ERRγLBD	Binary Complex
Sedimentation Coefficient of SAXS reconstructed shape[a]	3.5	3.4
Sedimentation Velocity Analysis	3.8	3.95
R_s of SAXS reconstructed shape[a]	35.3 Å	50.6 Å
R_s from SEC Analysis	33.2 Å	51.7 Å

[a] As calculated by HYDROPRO

Disorder-to-Order Structural Transition in the Assembly of the PGC-1a/ERRg Metabolic Hub Srikripa Devarakonda[a], Kushol Gupta, Michael J. Chalmers[c], John F. Hunt[d], Patrick R. Griffin[c], Gregory D. Van Duyne[b], Bruce M. Spiegelman PNAS 2011

Biophysical properties determined by orthogonal approaches should agree.