Five reasons to pay attention to Small Angle X-ray Scattering Assisted CASP

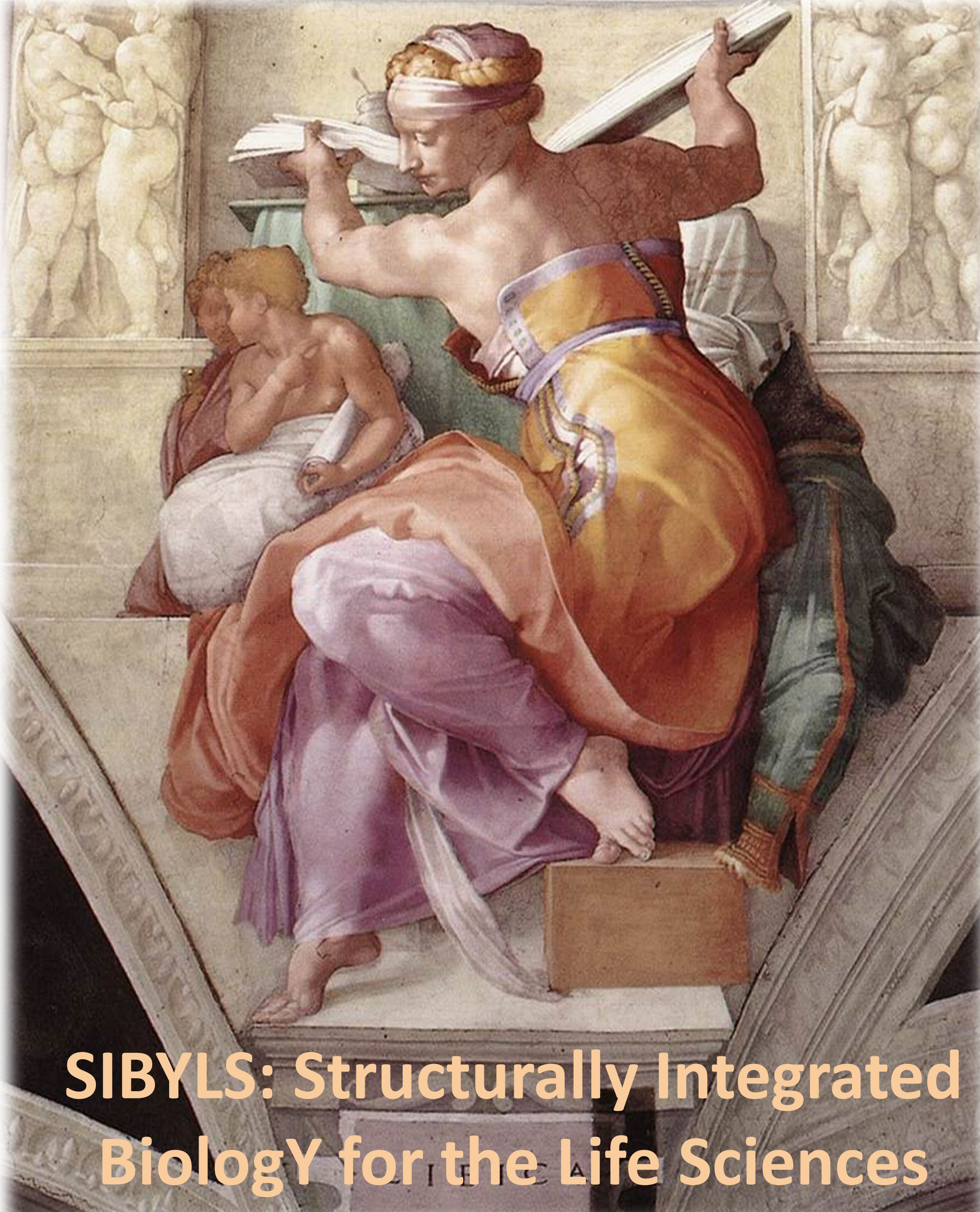
Susan Tsutakawa, Greg Hura, and John Tainer

Advanced Light Source Beamline 12.3.1 (SIBYLS)

BERKELEY LAB

sibyls.lbl.gov

www.bioisis.net



Why SAXS?

Five reasons to pay attention to Small Angle X-ray Scattering Assisted CASP

1. SAXS data can be collected at a fraction of the cost, time, and labor of other structural techniques.

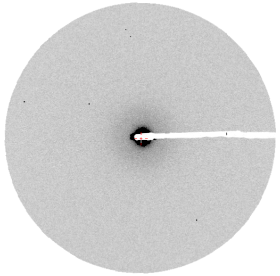
2. SAXS data provides distance information as experimental restraints for prediction algorithms.

3. SAXS data can filter starting models, provide shape, provide fold info, and orient domains and subunits.

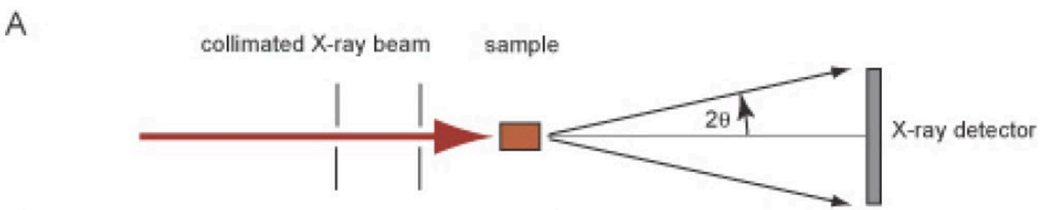
4. SAXS data can experimentally validate prediction models.

5. IT WORKS! (with need for improvement)

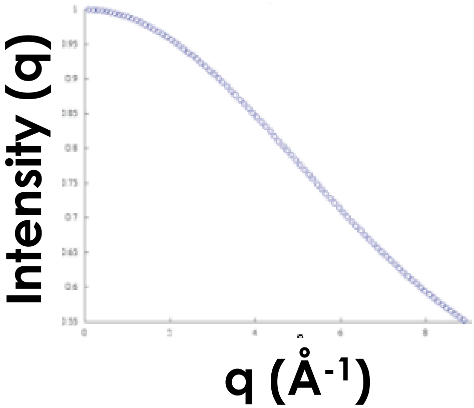
SAXS is a distance method, measuring shape and all electron pair distances (including flexible and cross-subunit).

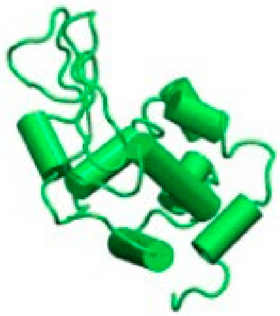


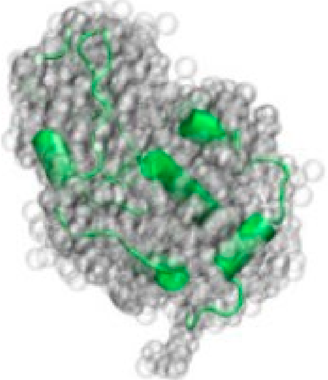
SAXS sample 30 ul
Protein 1-3 mg/ml
Exact Buffer

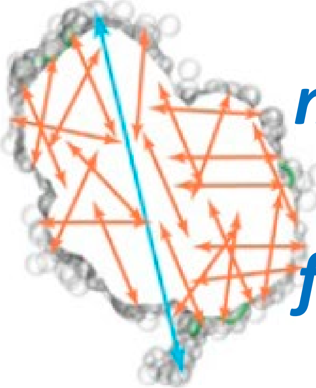


Scattering Curve









The p(r) is the sum of all "distograms" within the protein plus hydration layer.

SHAPE

Electron Pair Distances

SANS is a complementary method for complexes, separating the signal from one subunit from the others.

How good was the SAXS data provided for CASP13?

Note: ALL SAXS data was experimental. We have the throughput to do this.

The two SAXS data methods used are good for proteins that are flexible and/or multimerizing.

Technical challenges intrinsic to proteins in solution

1. Protein sample is low conc.
2. Protein is flexible
3. Protein is multimerizing
4. Stoichiometry is heterogeneous or protein is aggregated.

High throughput (HT-SAXS)	Size-exclusion Chromatography (SEC-SAXS)
Good signal to noise	Homogenous stoichiometry

H0980 and T0999 collected by outside groups.

We provided metrics, SAXS curves, and our quality assessment, including flexibility metric

Sample Quality

H0957 – Analysis 4/20/18

Small heterogeneity observed and corrected for in SEC-SAXS

Sample: CASP Chicago209Sample			
Variable	Value	Error +/-	Units
Rg	21.8	1	Angstroms
Porod Exponent	4.0		Scale (2-4)
Mass SAXS	32		kDa
Max Dimension	71	3	Angstroms
Radius of Cross Section	18.4	1	Angstroms
Volume	54545	5,000	Cubic Angstroms
Real Space Rg	21.58	3	Angstroms

Flexibility

Mass

Challenge

1:1 Complex

4 gold
6 silver
1 bronze

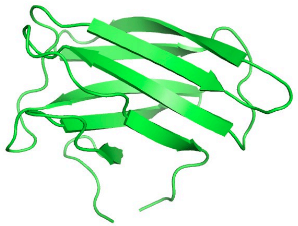
ChainA:
SNSFEVSSLPDANGKNHITAVKGDAKIPVDKIELYMRGKASGDLDSLQAEYNSLKDARISSQKEFAKDPNNAKRMEVLEKQIHNIERSQDMA
RVLEQAGIVNTASNNSMIMDKLLDSAQGATSANRKTSVVVSGPNGNVRIYATWTILPDGTKRLSTVTGTFK
ChainB:
SNAMINVNSTAKDIEGLESYLANGYVEANSFNDPEDDALECLSNLLVKDSRGGLSFCKKILNSNNIDGVFIKGSALNFLLLSEQWSYAFEYLTS
NADNITLAE LEKALFYFYCAKNETDPYPV PEGLFKKLMKRYEELKNDPD AKFYHLHETYDDFSKAYPLNN
Total 36.4kDa

Mass

What information about the targets is provided by SAXS (solution)? How does it compare to crystal (lattice/assessment)?

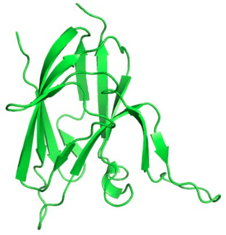
11 CASP13-SAXS targets include 4 monomers and 7 multimers, 14-340 kDa

T0992



14 kDa

T0949



17kD

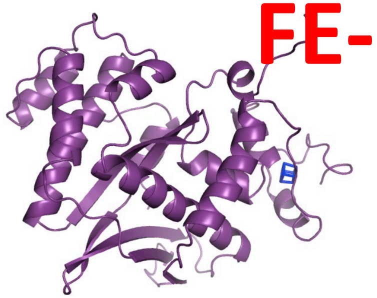
T0987

CASP:
Image redacted

46 kDa

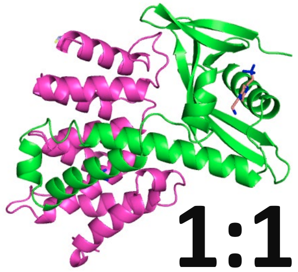
T0975

FE-S



38 kD

H0957

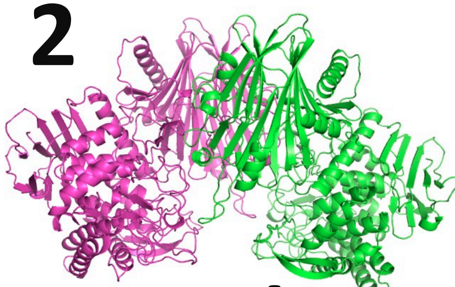


1:1

36 kD

T0985

2

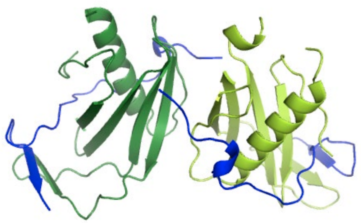


200 kDa

CASP:
Image redacted

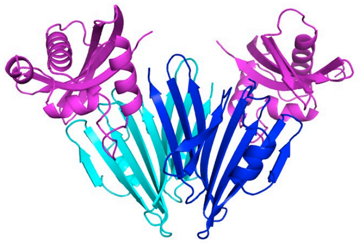
T0999

homodimer
340 kD



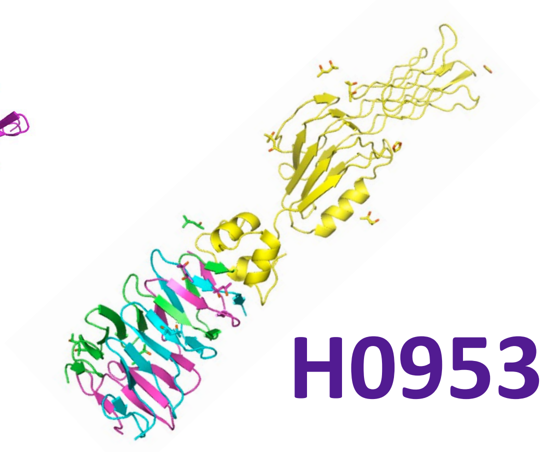
H0980

2:2 40 kD



H0968

54 kDa
2:2



H0953

3:1 48 kD

CASP:
Image redacted

T0981

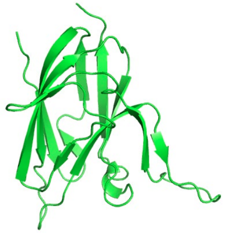
homotrimer
228 kD

Compared to 2016 CASP12-SAXS, the structures had much less disordered regions. Only 3 were missing more than 10% and those were maximum 18%.

T0992

15%

T0949



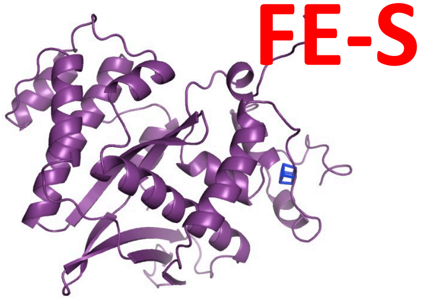
8%

T0987

CASP:
Image redacted

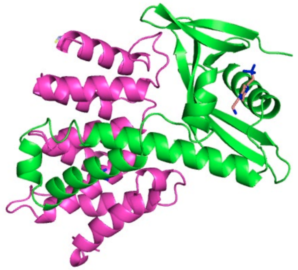
6%

T0975



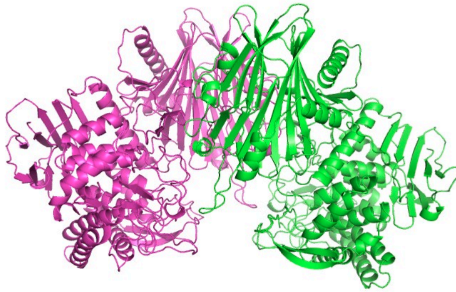
18%

H0957



3%

T0985

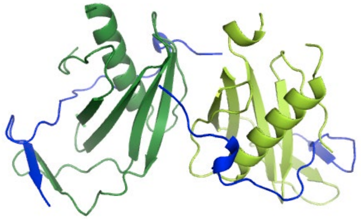


2%

CASP:
Image redacted

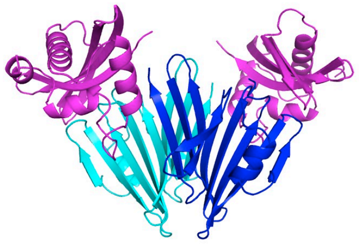
T0999

3%



H0980

14%



H0968

4%

H0953

2%

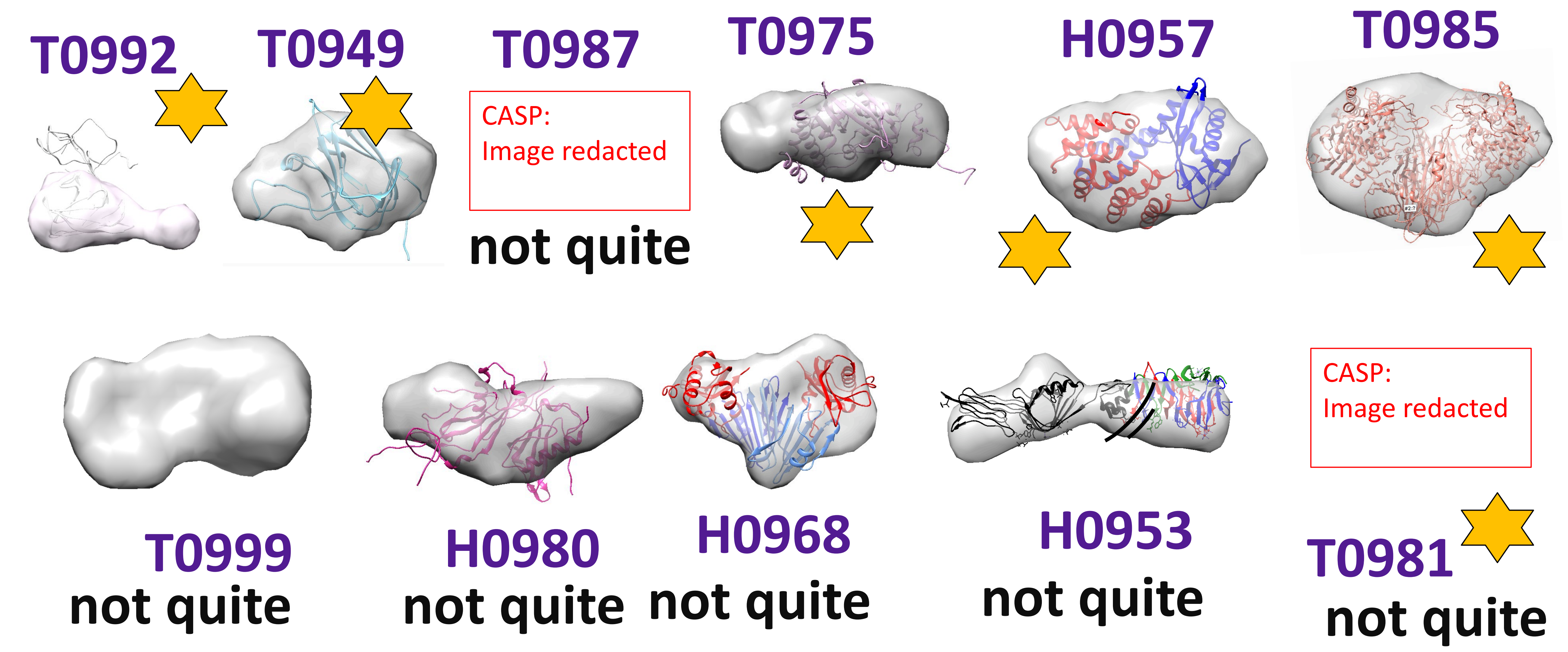
CASP:
Image redacted

T0981

10%

6/11 Crystal Structures fit within the SAXS-derived shape.

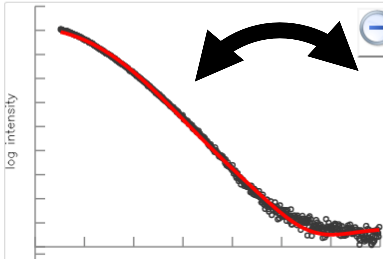
5/11 stick out a bit.



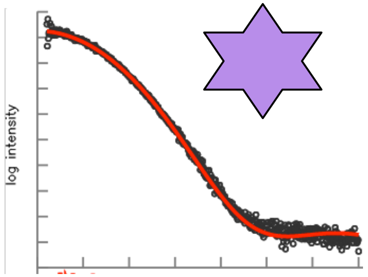
4 Crystal Structures+disordered tails agreed with solution SAXS data based on comparison of Scattering Curve.

6-7 Will likely Require Domain movements.

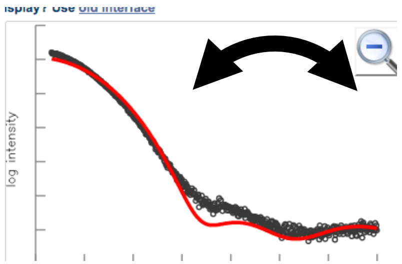
T0992



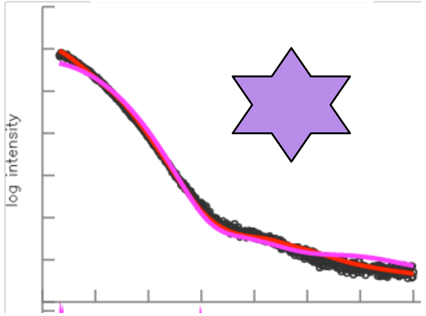
T0949



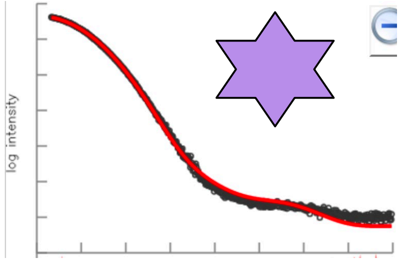
T0987



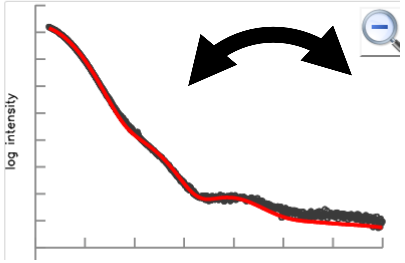
T0975



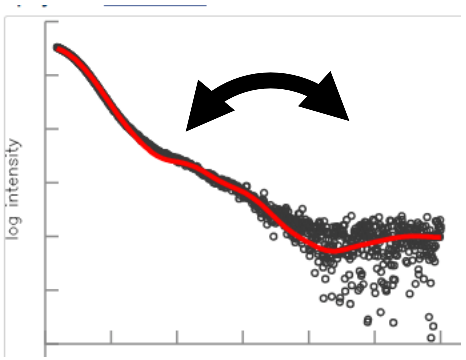
H0957



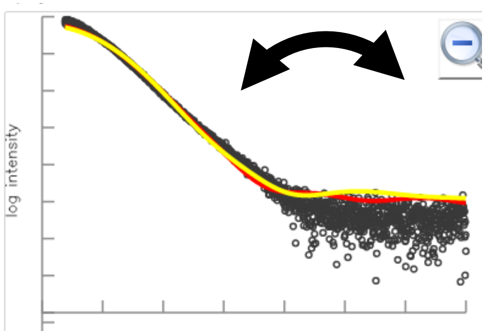
T0985



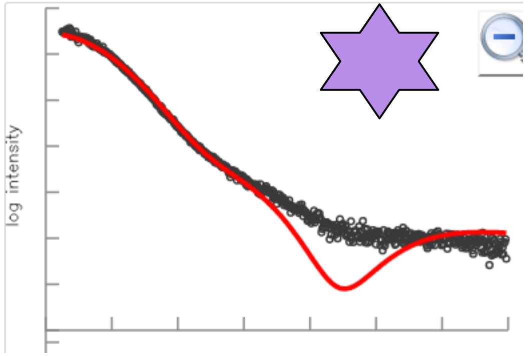
T0999



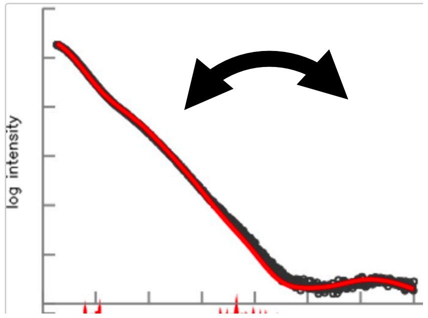
H0980



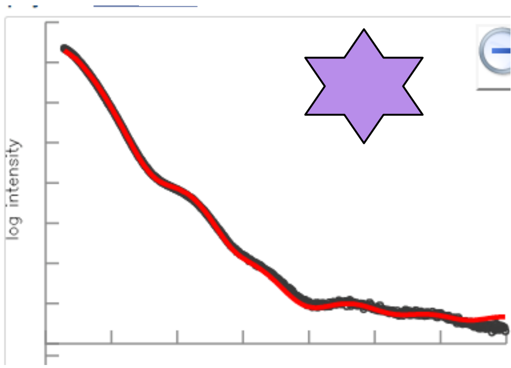
H0968



H0953

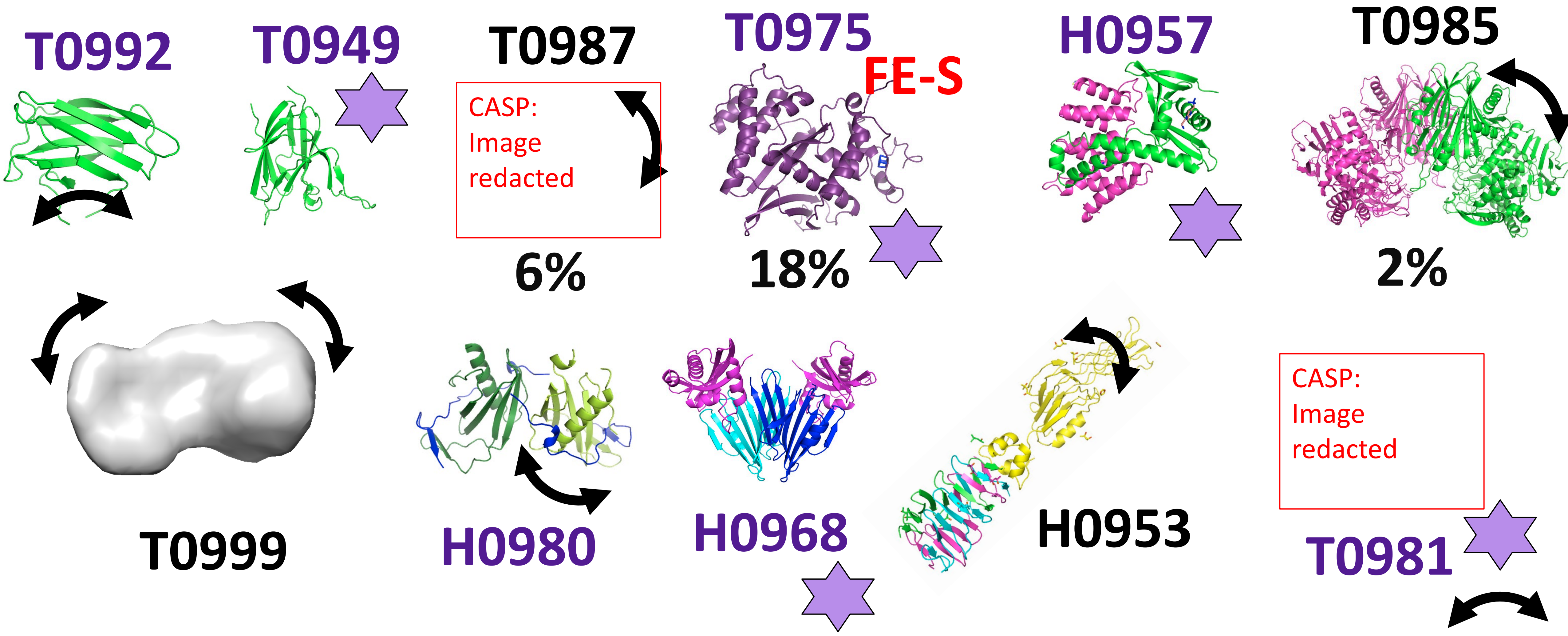


T0981

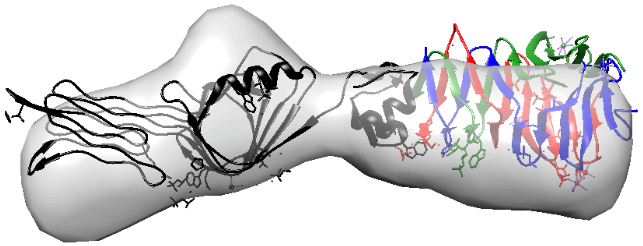


4 Crystal Structures+disordered tails agreed with solution SAXS data based on comparison of Scattering Curve.

6-7 Will likely Require Domain movements.



How did the predictors do? The Devil is in the details.



Two ways that SAXS data helps.

1. **Improve overall shape (density using gmfit tool)**
2. **Improve fold (GDT-TS, QCS score)**

Considerations

1. **Some groups only submitted SAXS predictions. Cannot assess SAXS-based improvement.**
2. **Should see improvement on whole protein. GDT-TS & QCS scoring by domains or polypeptides.**
3. **SAXS data in solution does not necessarily match crystal.**

Discrepancy between model sequence and SAXS sample suggests modelers might be mislead by fitting for something that is suppose to be there or not there.

6/11 models – AA sequence matched SAXS sample

Targets H0953, H0957, H0968, T0975, H0980 (-6 nt), T0999

2/11 models –prediction seq sometimes did not match SAXS sample

3/11 models –all model seq did not match SAXS sample.

Prediction models -102 to +32 AA.

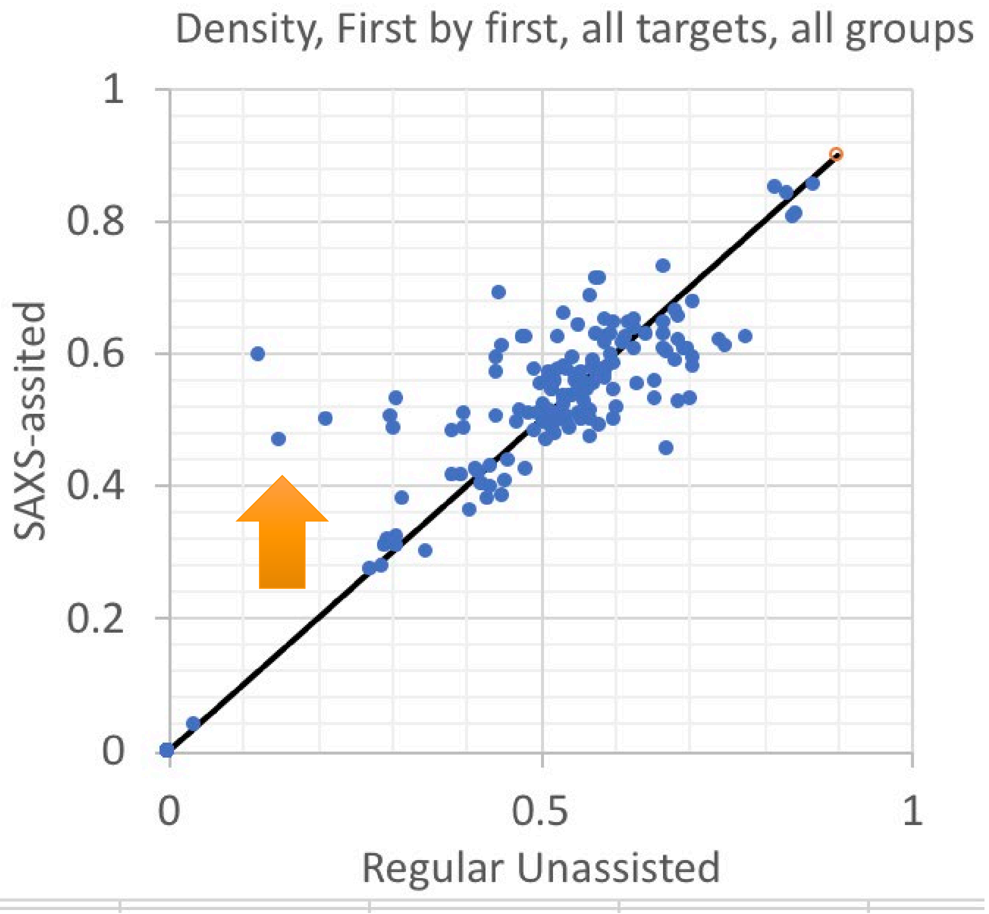
Crystal sequence might not match SAXS sample

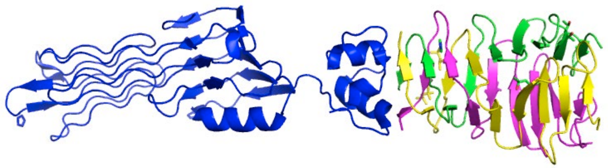
SAXS measures ALL electron pairs of protein & its hydration layer.

Please double check that model sequence matches the SAXS sample

Predictions showed improvement in overall shape.

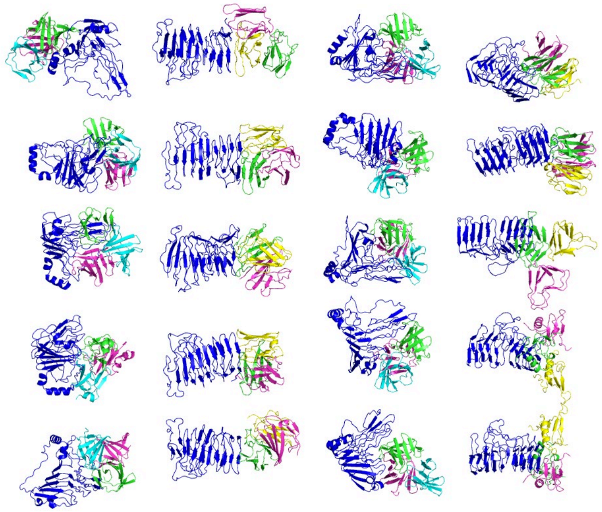
Shape – Predictors improved their overall shape with SAXS data.



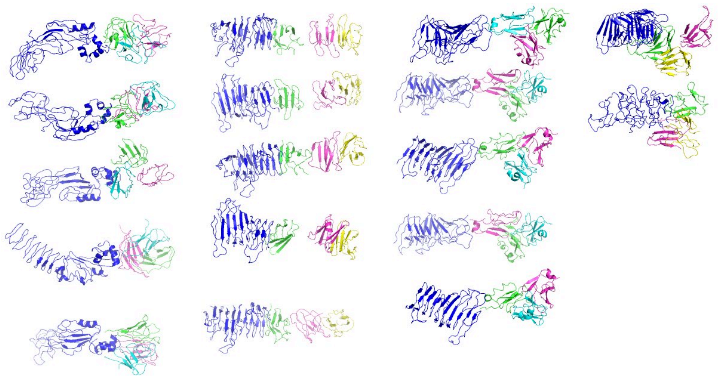


H0953 elongated

Regular (unassisted)



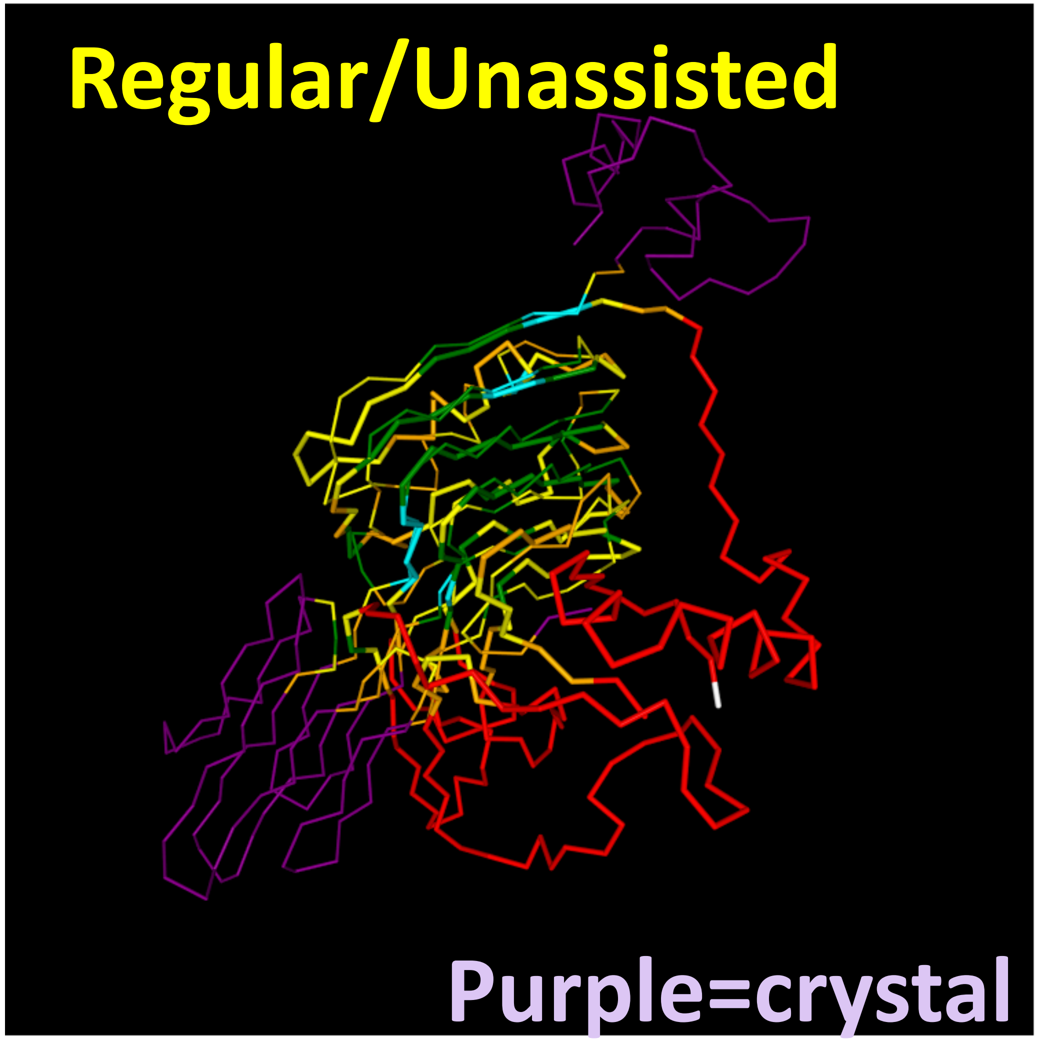
SAXS-assisted

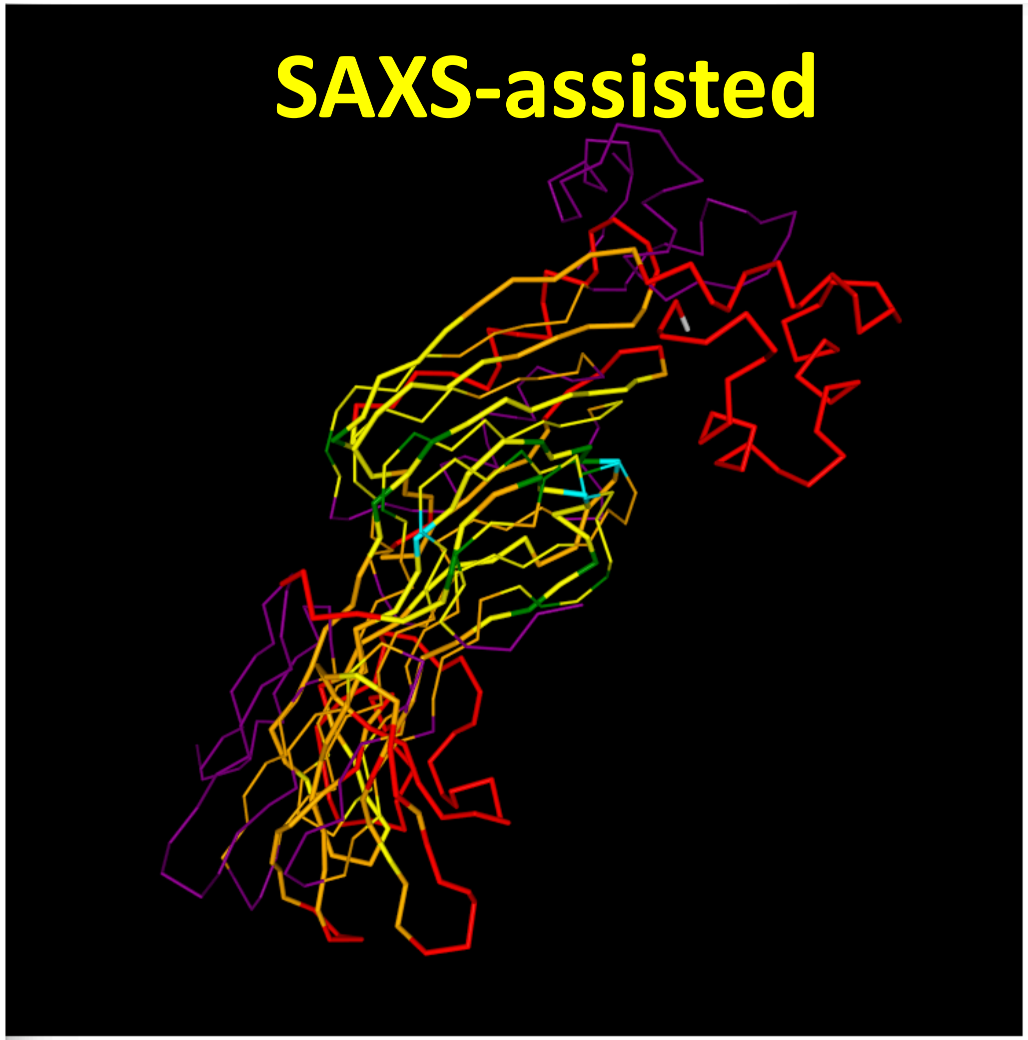


Gmfit – Dmytro Guzenko

For the top scoring GDT_TS model for H0953s2, can visibly see improved fold

196 Grudinin





SAXS improves Domain Definition Observed even though GDT_TS low.

D-Haven without SAXS
GDT_TS 16
Density 0.625
TM 0.28

Only 4 groups participated (D-Haven, kozakov-vajda, Grudinin, SBROD)– why?

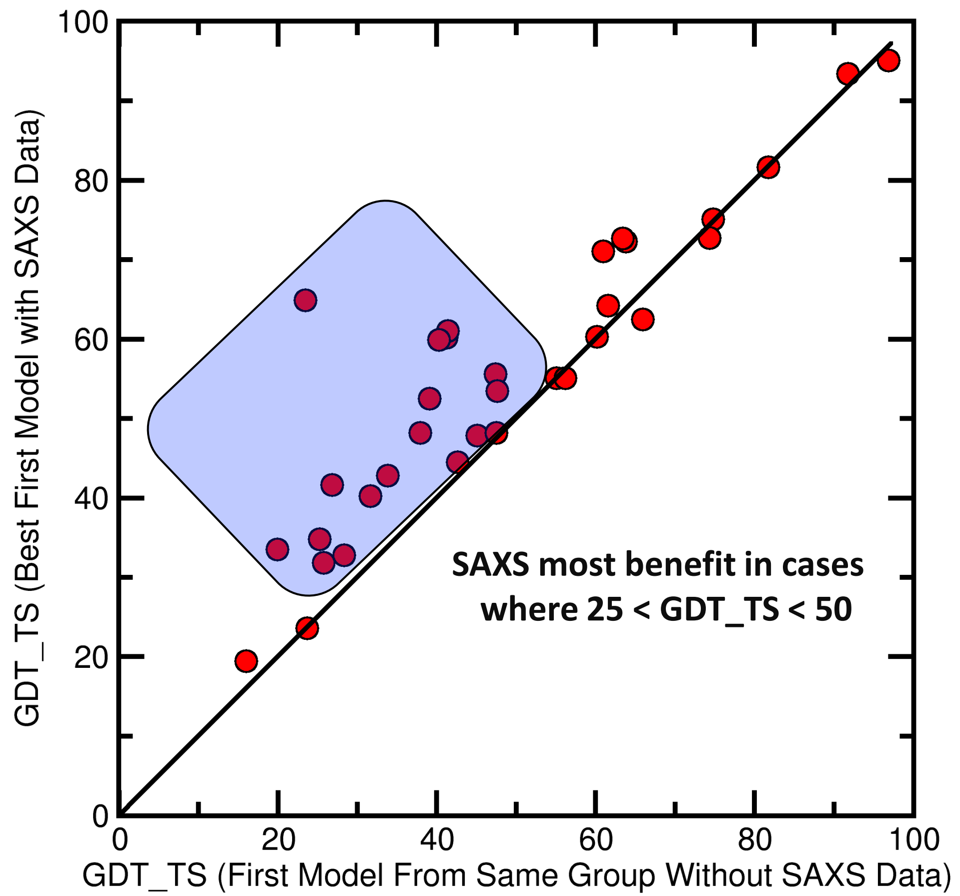
D-Haven with SAXS
GDT_TS 15
Density 0.625
TM 0.37

CASP:
Image redacted

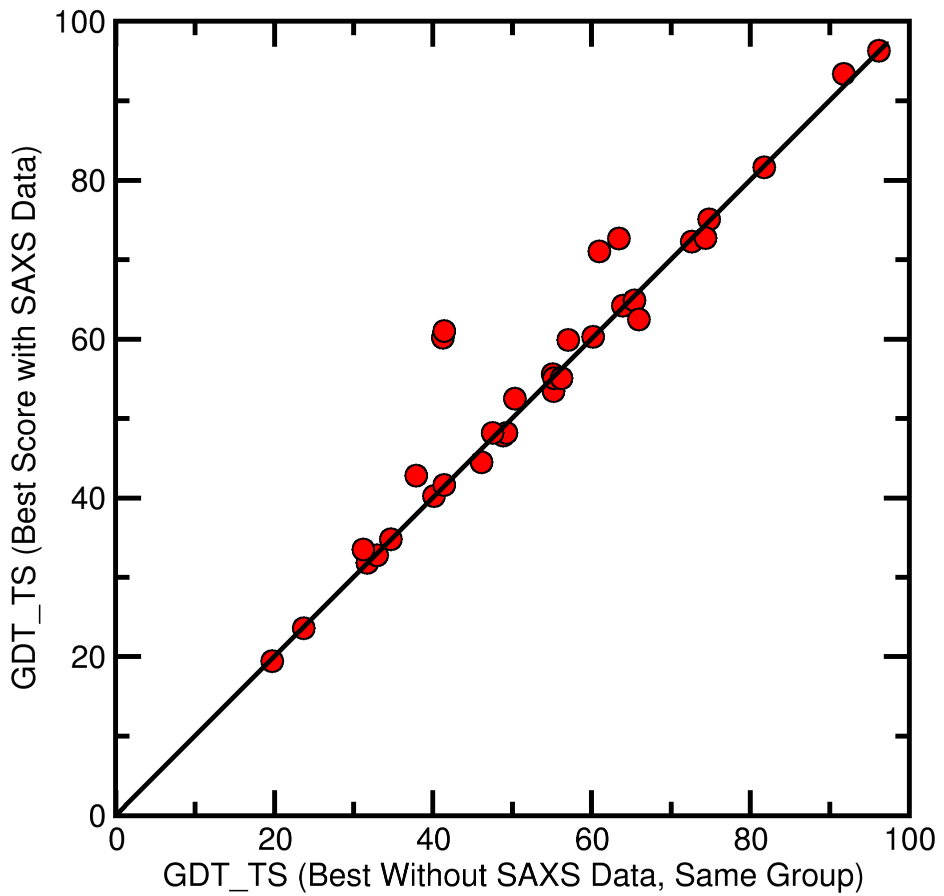
Some improvement on SAXS-assisted fold.

SAXS helps with fold and with ranking in top scoring SAXS-assisted GDT_TS models.

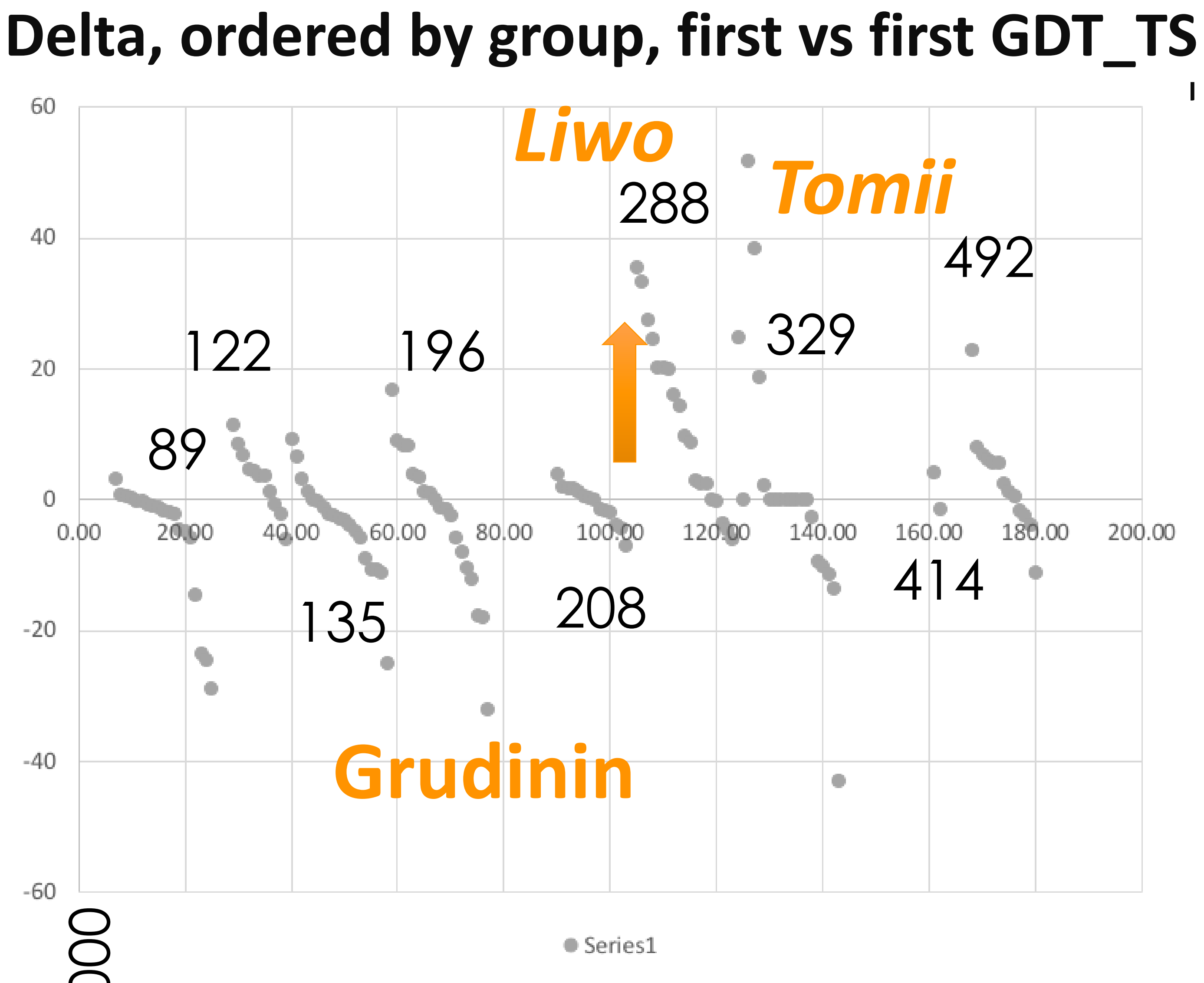
SAXS Model 1 vs Reg Model 1



SAXS Best vs Best



By group, Liwo consistently improves. Grudinin and Tomii see improvements with some targets, where they become one of the top scoring groups..



Liwo and Grudinin are discussion participants in Workshop 1 (after this session)

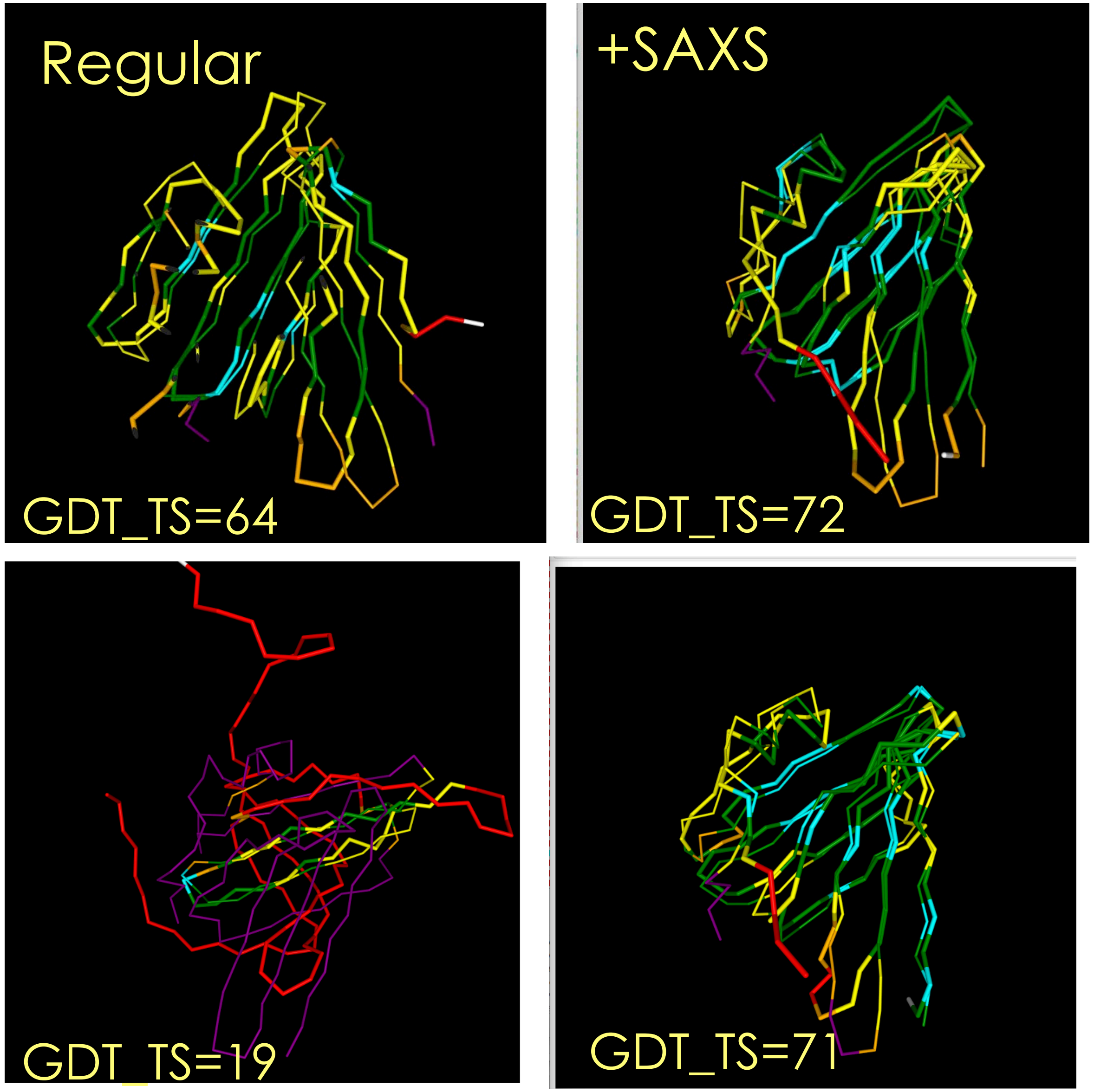
Tomii is a discussion participants in Workshop 2 (3 pm today)

Individual Example

For S0968 s2 which showed GDT_TS improvement, see that edges improves. For biologists, edges are important – it's where active sites and interfaces are.

196/Grudinin (#1 saxs)

329/D-Haven

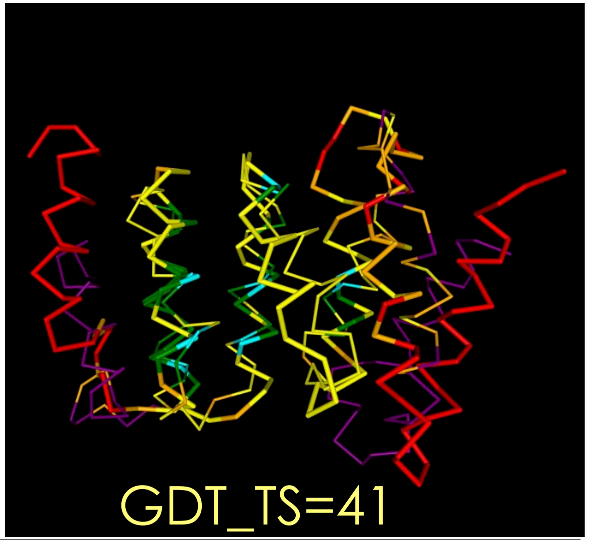


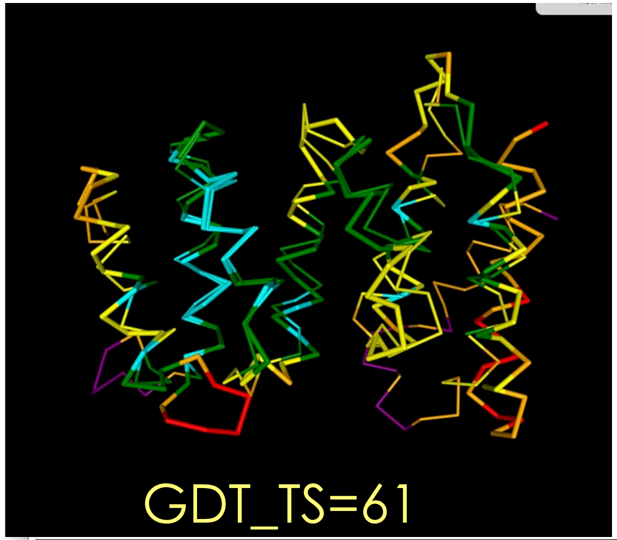
GDT_TS score improves to 72 (top)

For S0957 s2 which showed GDT_TS improvement, see that biologically-important edges improves.

329/D-Haven

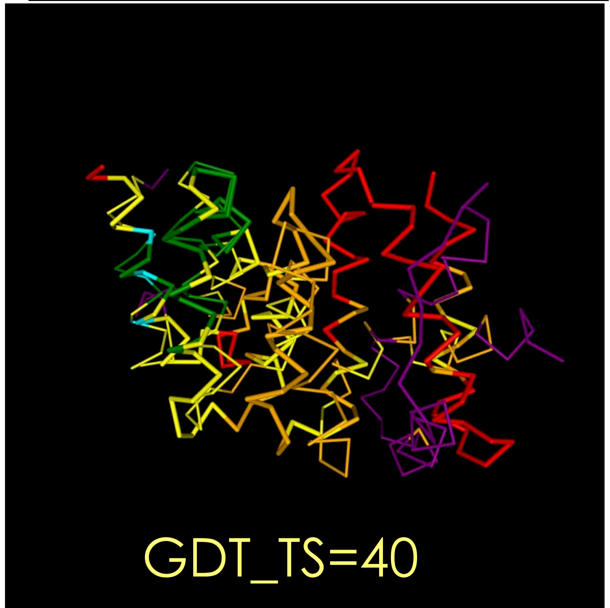
#	Model	(T0)	(S)	(S) Δ
1.	_0957s2TS329_1	41.452	60.968	19.516
2.	_0957s2TS196_1	40.000	56.935	16.935
3.	_0957s2TS135_1	40.000	49.355	9.355



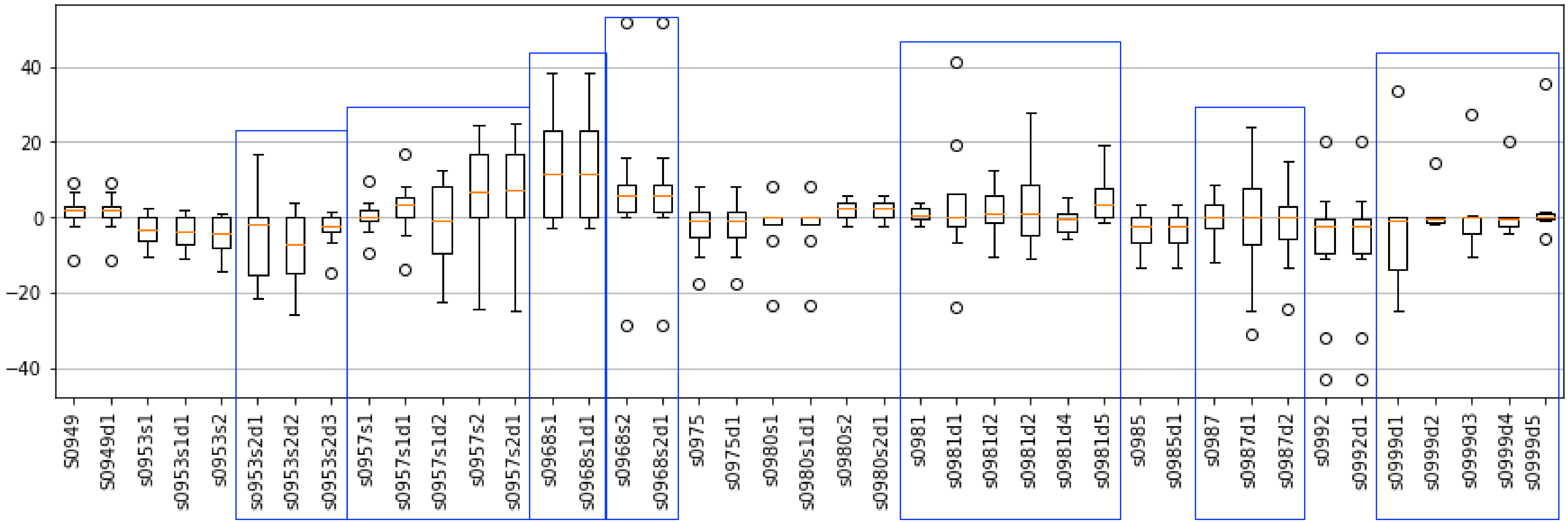


GDT_TS score improves to 60 (top).

196/Grudunin

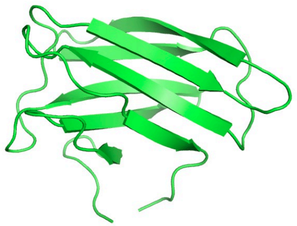


There was some improvement for all models, but most significant improvement in certain cases.



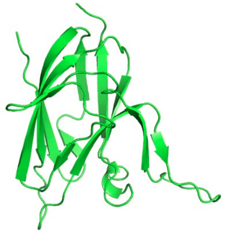
See most improvements in GDT_TS score in heteromeric complexes, medium size, AA correct.

T0992



14 kDa

T0949



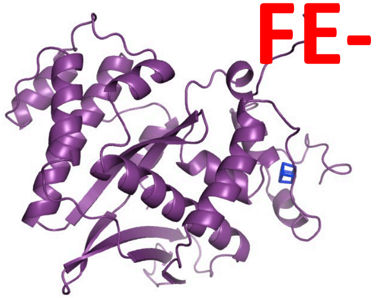
17kD

T0987

CASP:
Image redacted

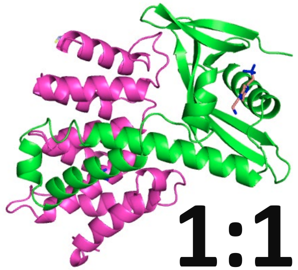
46 kDa

T0975



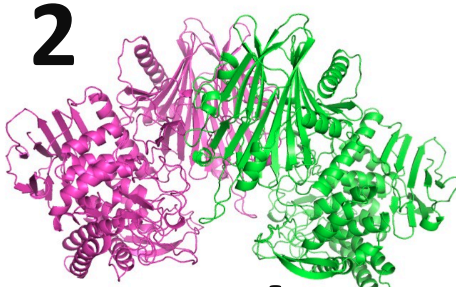
38 kD

H0957



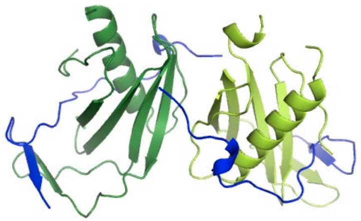
36 kD

T0985

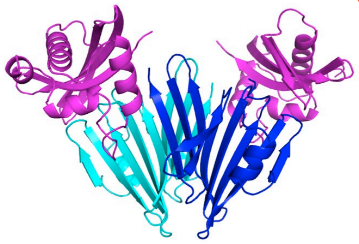


200 kDa

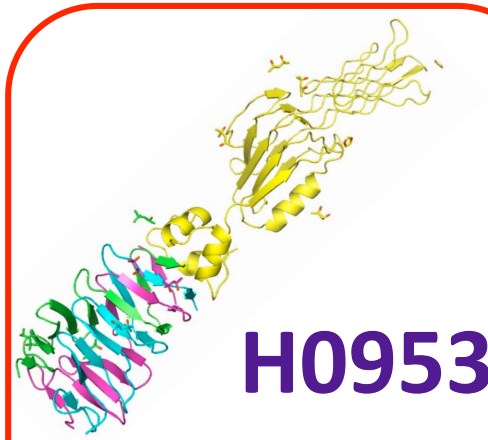
T0999
homodimer
340 kD



H0980
2:2 40 kD



H0968
54 kDa
2:2



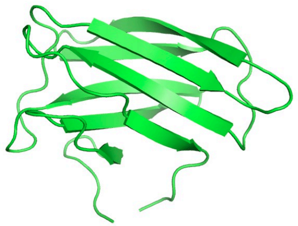
H0953
3:1 48 kD

CASP:
Image redacted

T0981
homotrimer
228 kD

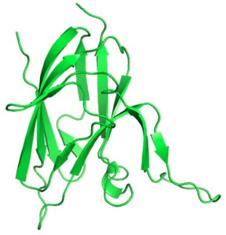
Improved models do not correlate with flexibility or fit in the SAXS envelope…

T0992



14 kDa

T0949



17kD

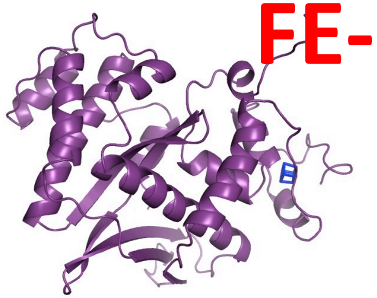
T0987

CASP:
Image redacted

46 kDa

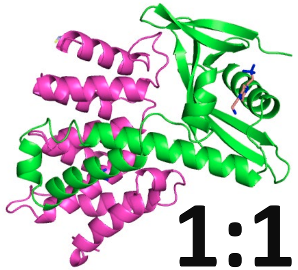
T0975

FE-S



38 kD

H0957

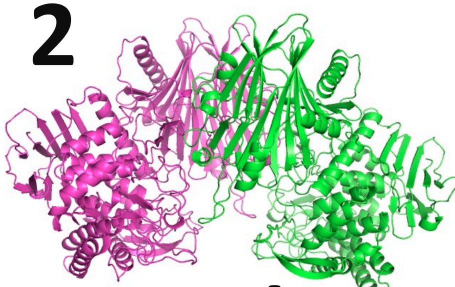


1:1

36 kD

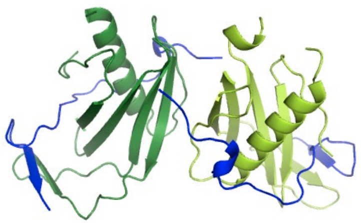
T0985

2



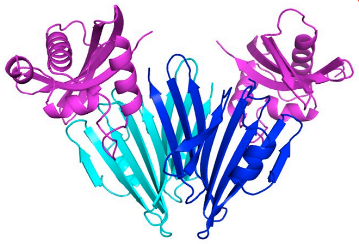
200 kDa

T0999
homodimer
340 kD



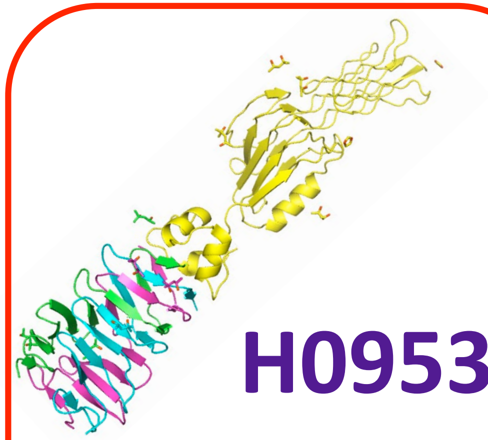
H0980

2:2 40 kD



H0968

54 kDa
2:2



H0953

3:1 48 kD

CASP:
Image redacted

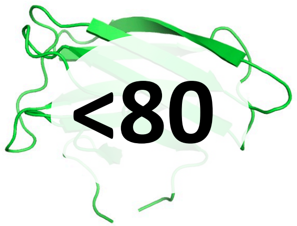
T0981
homotrimer
228 kD

Structures that predictors overall had low scores were in the not improved.

small

template

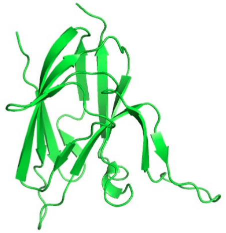
T0992



<80

14 kDa

T0949



17kD

T0987

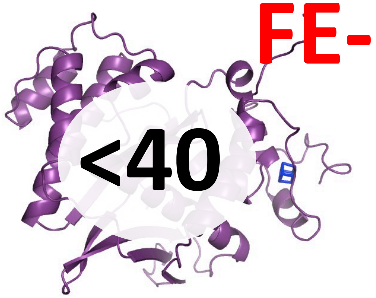
CASP:
Image redacted

<25

46 kDa

T0975

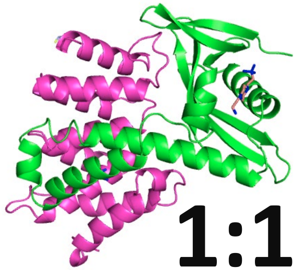
FE-S



<40

38 kD

H0957

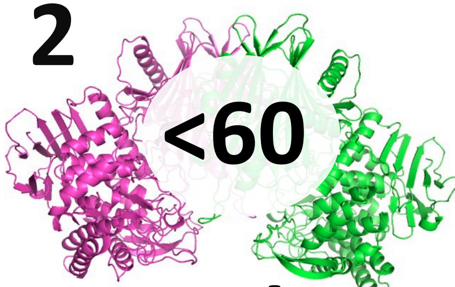


1:1

36 kD

T0985

2



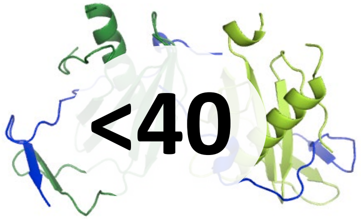
<60

200 kDa

T0999

homodimer

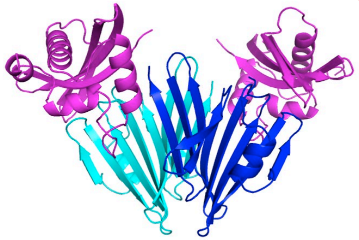
340 kD



<40

H0980

2:2 40 kD



H0968

54 kDa

2:2

H0953

3:1 48 kD

<20

CASP:
Image redacted

T0981

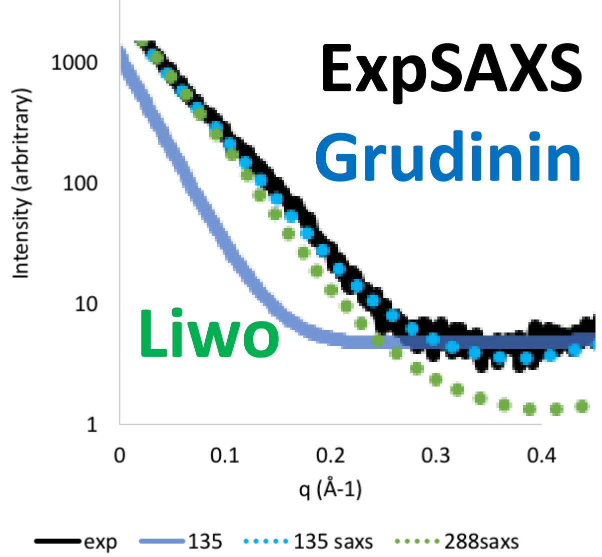
homotrimer

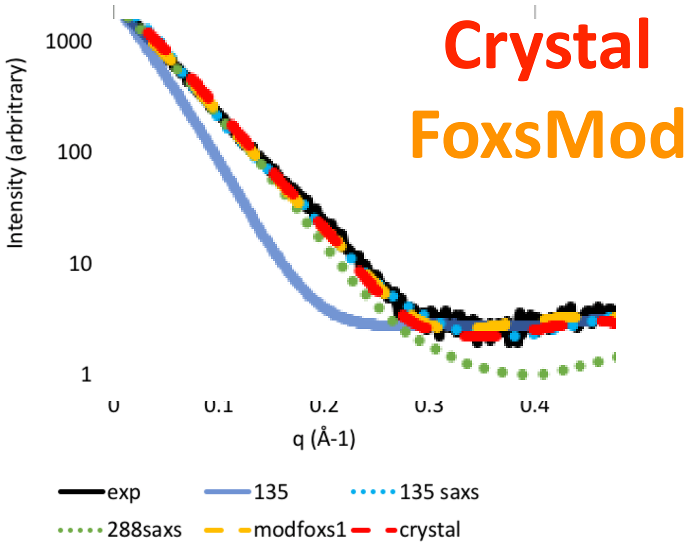
228 kD

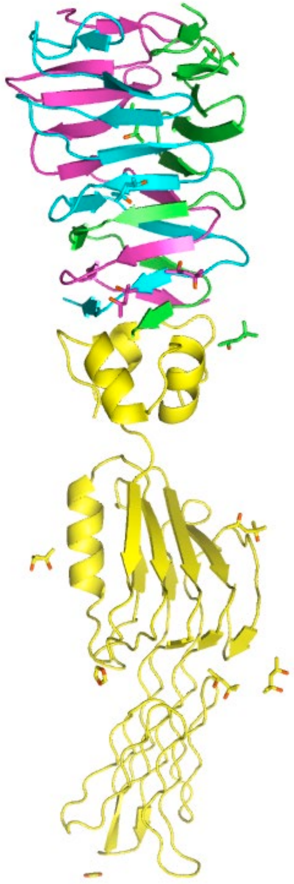
Could the SAXS have helped to reach the crystal structure?

Predictors have room for improvement – crystal structure fits the exp data best.

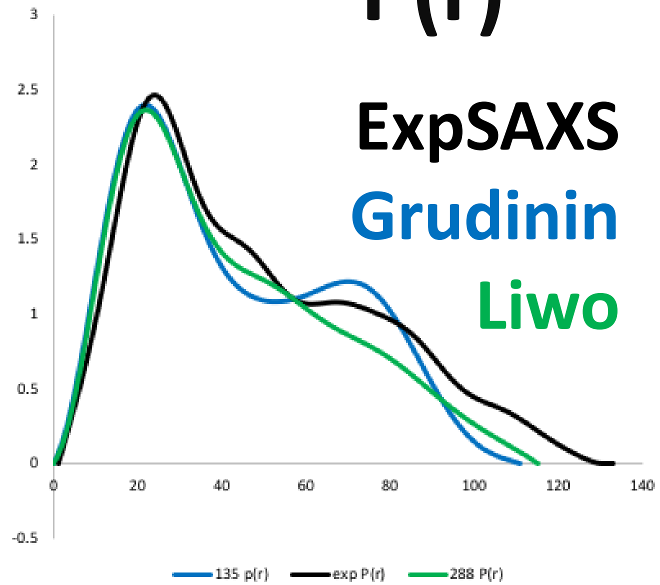
Log scale



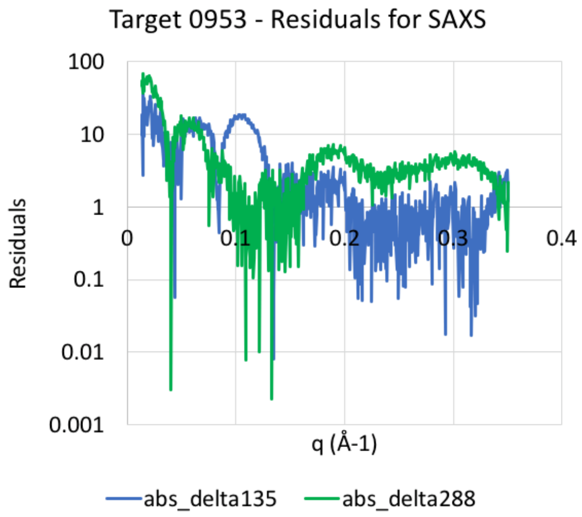


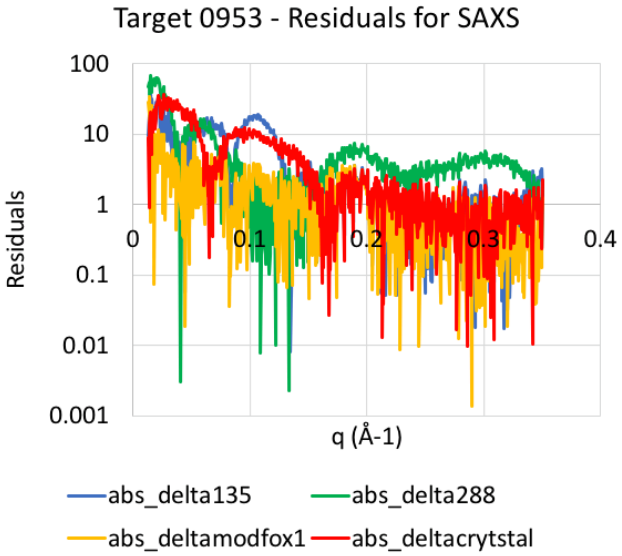


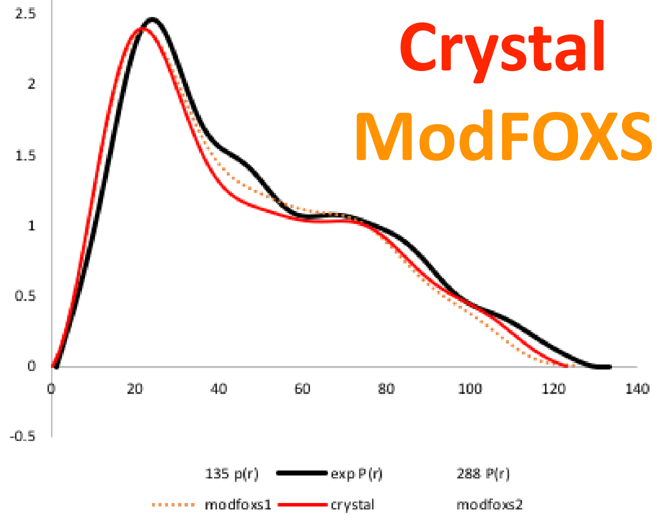
P(r)



Log scale







CASP13 Assessor Results for SAXS-assisted category. Target variability makes unifying summary difficult

Improvement of models for some targets

SAXS provided accurate guidance in most cases

Further improvement is required (please work with our beamline)

Workshops 1 and 2

Improved metrics??

Improved integration with algorithms??

Simulated SAXS data for crystal lattice structures??

Higher resolution SAXS data??

Novel methods to reduce signal from disorder regions??

Dynamic predictions?

SIBYLS Related Staff

John Tainer   Michal Hammel   Scott Classen

Jane Tanamachi

Curtis Hodge

David Shin

Kathryn Burnett

Susan Tsutakawa

Daniel Rosenberg

CASP Committee

+ Dmytro Guzenko