CASP Commons

Y. Ishida, N. Denissova, G. Liu,
G. V. T. Swapna, G. T. Montelione,

G. Hura, S. Tsutakawa, J. Tainer

J. Moult

Chin-Hsien Emily Tai

K. Fidelis, A. Kryshtafovych

CASP Commons Overview (Guy)   3 min

CASP Commons Target Collection (Emily) 3 min

Modeling of CASP Commons Targets (Krzysztof)   6 min

Data Collection and Data Guided Prediction (Guy)   5 min
- Protein sample production
- NMR - 3 proteins

Data Collection SAXS - 5 - 6 proteins (Greg)   5 min
- What is SAXS? 1 to 2 slides - what is SAX
- Status of CASP Commons targets

How to Provide Data / Reports to the Nominators and Plan for Future
Audience Discussion - lead by Guy, Emily, Greg, John Moult, John Tainer, and Krzysztof
8 min
- Is this a valuable activity for the CASP community?
- how to bring in new targets?
- how to obtain funding for experimental activities?

CASP Commons Overview

Guy 3 min

CASP Commons

Vision Statement

Protein structures drive biology.

CASP Commons will engage the protein modeling community with the broader biological community to address important problems in biology and medicine.

CASP Commons

Implementation of the Vision

To engage the CASP scientific community in both ‘regular’ and ‘data-assisted’ protein structure modeling on a large number of biomedically-important proteins and complexes for which high-resolution experimental structures are not available.

Example CASP Commons Activities

- Targets broadly nominated by the biology community
- Targets involved in a particular disease (e.g. cancer), organism, or biological process.

CASP Commons

Targets and data are being generated by CASP Organizers

Proposed by high-impact biomedical research labs.

Range from 50 to 200 residues. May be monomers or oligomers.

No good templates can be identified for modeling.

Shallow multiple sequence alignments (N_{eff} / L < 2).

Structures to eventually be determined by CASP Organizers – may not have 3D structures for assessment for some years.

Assessment will be an ongoing activity.

CASP Commons Goals

Three key objectives:

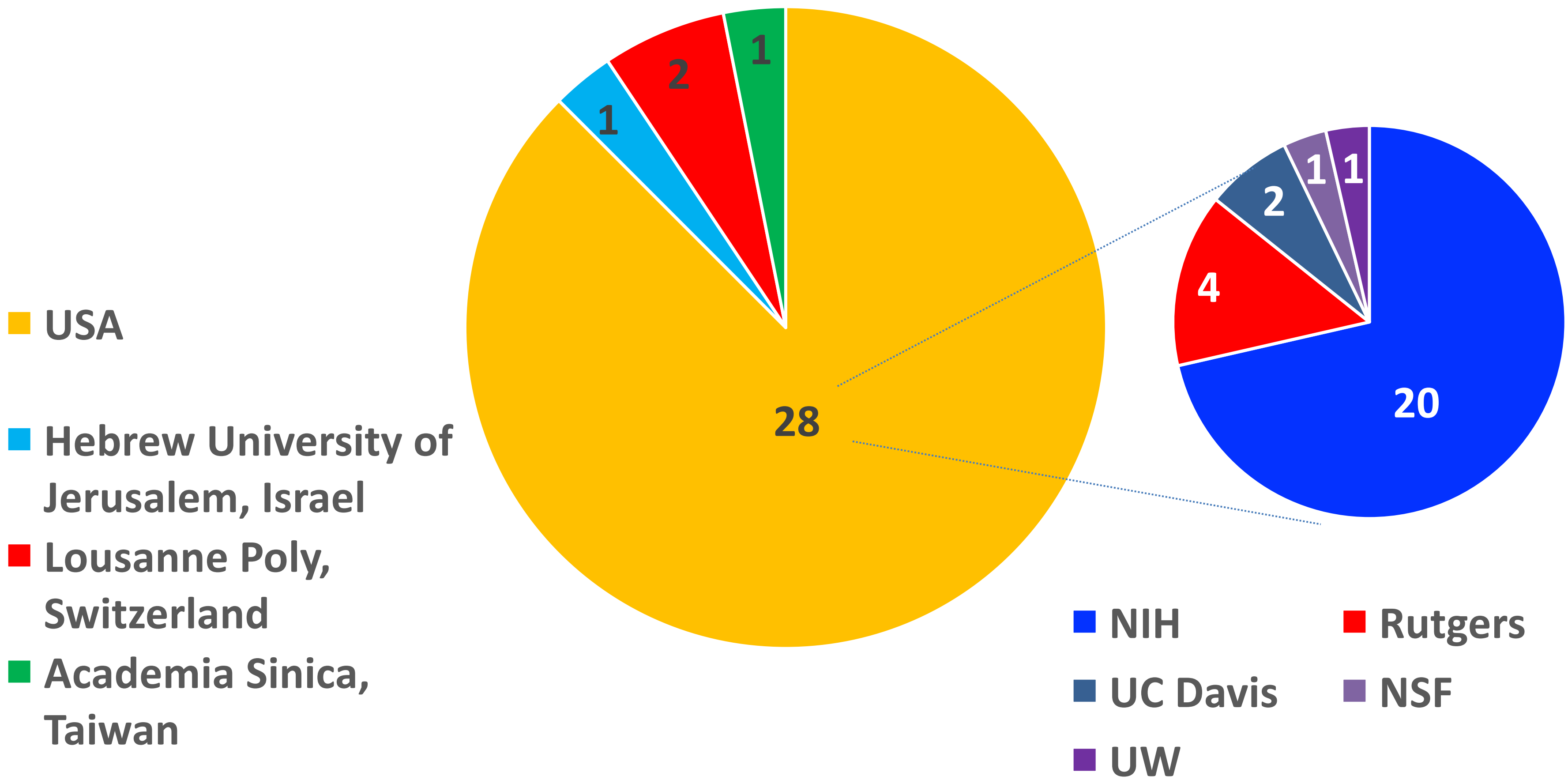
- Structural models for biology
- Provide a bridge between the modeling and biological communities
- Drive methods for data-assisted modeling

Goal today: To gain input from, and engage, the CASP modeling community.

CASP Commons Target Collection

Emily 3 min

32 targets nominated from 26 labs worldwide



CASP Commons Targets

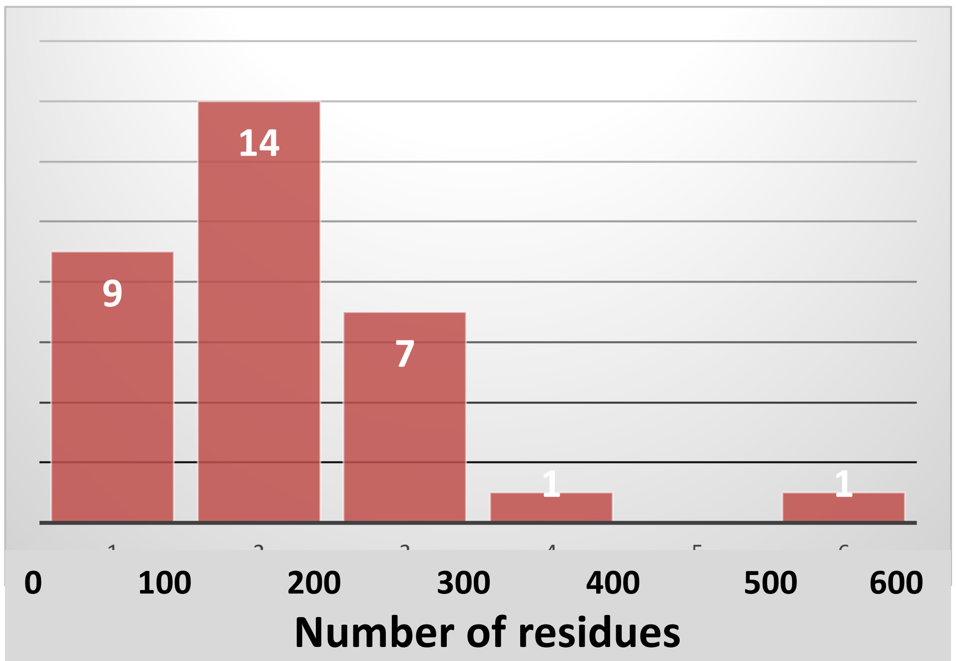
Modeling targets

lastname	firstname	institution	domain	length	HHsearch top templ	HHsearch prob	HHsearch coverage
Abriata	Luciano	Lousanne Poly	DHHC6_SH3	81	2RQR_A	85.6	0.49
Abriata	Luciano	Lousanne Poly	COA6	125	5J4Z_BH	99.6	0.54
Best	Sonja	NIH/NIAID	LIM	206	1RUT_X	99.7	0.72
Folco	Hernan	NIH/NCI	SPAC17C9	94	5MRC_VV	26.1	0.55
Golden	Andy	NIH/NIDDK	EMB1	81	5A31_D	56.1	0.48
Harmer	Stacey	UC Davis	XAP5	187	5NSA_A	56.2	0.2
Jacobs	Dakota	NIH/NCI	NELF_Cterm200	202	2MR5_A	64.6	0.29
Jacobs	Dakota	NIH/NCI	NELF	532	2BL0_A	83	0.07
Kimura	Shioko	NIH/NCI	SCGB_3A2	93	1CCD_A	97.7	0.69
Koepnick	Brian	UW, Baker Lab	UW_engnr	80	3HF3_D	71.6	0.85
Liang	Jake	NIH/NIDDK	HBx	154	6EU1_N	19	0.12
Masison	Cynthia	NIH/NCI	P12C39A	99	2KIX_D	16.1	0.15
Maurizi	Michael	NIH/NCI	PinA	161	2QL2_B	65.3	0.14
Michelmore	Richard	UC Davis	BIRXLR3	141	2LC2_A	95.8	0.55
Mushegian	Arcady	NSF/MCB, Virginia	NP_062900	189	3PUN_B	53.9	0.48
Myrum	Craig	NIH/NIA/IRP	ARC_Nterm1-140	140	1TQG_A	43.7	0.16
Myrum	Craig	NIH/NIA/IRP	ARC	396	4X3X_A	100	0.22
Prosser	Gareth	NIH/NIAID	EFD57440	107	3HFE_A	32.4	0.13
Rein	Alan	NIH/NCI	EIAV	68	4K02_B	11.9	0.31
Robert	Hufnagel	NIH/NEI	PNPLA6	167	5FYA_B	99.9	0.98
Schneider	Thomas	NIH	RepA	286	1REP_C	99.5	0.7
Schwartz	Daniella	NIH/NIAMS	NBCe1-B	160	5JHO_A	99.9	0.48
Ten-Hagen	Kelly	NIH/NIDCR	NP_476718	74	2HNW_B	37.6	0.14
Wang	Qinglan	NIH	PPE_Mtb	181	1KRK_.	16.4	0.04
West	Jennifer	NIH/NIDDK	MotB	162	4YTK_A	28.6	0.25
Yang-Yen	Hsin-Fang	IMB, Taiwan	PRAP1	149	1Y0G_C	29.7	0.11
Yarden	Oded	Hebrew U of Jerusalem	SSP1	149	5GNA_A	53.7	0.15
Young	Howard	NIH/NCI	BAE31734	78	2AAZ_B	38.6	0.47
Zhang	Zhiyong	Rutgers U	Etglo4	216	2LVF_A	98.2	0.3
Zhang	Zhiyong	Rutgers U	Etglo1	216	1S6D_A	98.2	0.31
Zhang	Zhiyong	Rutgers U	Etglo2	258	1S6D_A	98.1	0.25
Zhang	Zhiyong	Rutgers U	Etglo3	290	1S6D_A	97.9	0.22

Proteins from phage, human and plant virus, mycobacterium, fungus, plants, mouse, and human

Functions include pathogen infection, host immunity, cell cycle, long-term memory, detoxification, oncoprotein, chaperones, drug targets, splicing.

Target size



15 targets selected for NMR trial experiments

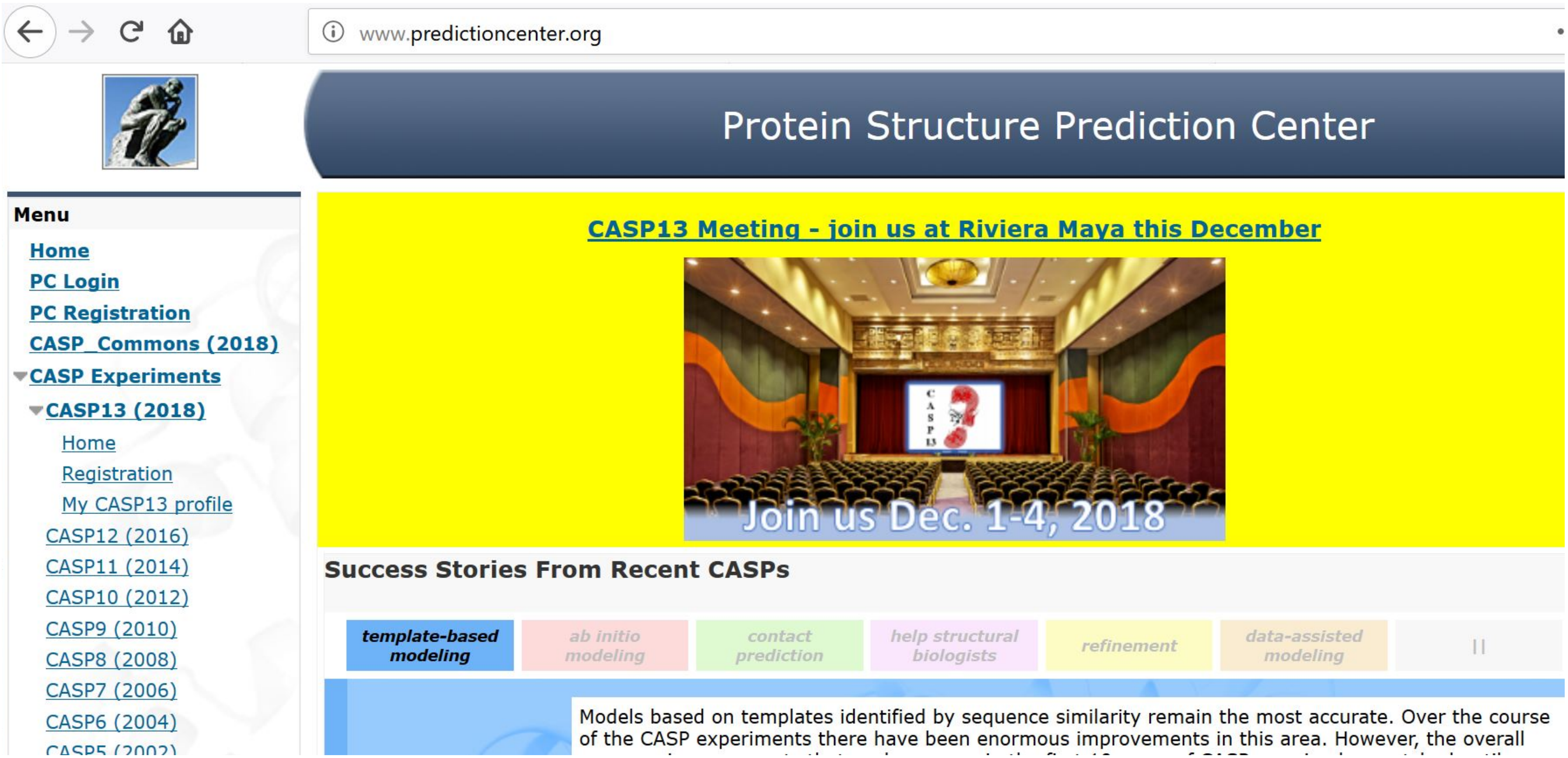
Selection criteria:

- Size (50 ~ 200 residues)
- Template availability
- Shallow multiple sequence alignments (N_{eff} / L < 2)
- Number of cysteine in the sequence no more than 10

	domain	status (DQF)	CYS	length (60-180)
1	EIAV	1	0	68
2	PRAP1	1	1	149
3	EMB1	1	1	81
4	UW_engnr	2	0	80
5	MotB	2	2	162
6	BAE31734	2	2	78
7	PinA	2	3	161
8	BIRXLR3	2	3	141
9	ARC_Nterm1-140	2	4	140
10	PPE_Mtb	3	0	181
11	XAP5	3	0	187
12	NP_062900	3	2	189
13	SSP1	3	7	149
14	Hbx	3	10	154
15	EFD57440	4	0	107
16	[illegible]	[illegible]	[illegible]	[illegible]

Modeling of CASP Commons Targets

Krzysztof 7 min


www.predictioncenter.org
Protein Structure Prediction Center
Menu
Home
PC Login
PC Registration
CASP_Commons (2018)
CASP Experiments
CASP13 (2018)
Home
Registration
My CASP13 profile
CASP12 (2016)
CASP11 (2014)
CASP10 (2012)
CASP9 (2010)
CASP8 (2008)
CASP7 (2006)
CASP6 (2004)
CASP13 Meeting - join us at Riviera Maya this December
Join us Dec. 1-4, 2018
Success Stories From Recent CASPs
template-based modeling
ab initio modeling
contact prediction
help structural biologists
refinement
data-assisted modeling
Models based on templates identified by sequence similarity remain the most accurate. Over the course
of the CASP experiments there have been enormous improvements in this area. However, the overall

CASP_Commons

[Register for CASP_Commons]

[Released Target List]

[Submit a prediction]

[Model Accuracy Estimates]

Modeling targets

lastname	firstname	institution	domain	length	HHsearch top templ	HHsearch prob	HHsearch coverage
Abriata	Luciano	Lousanne Poly	DHHC6_SH3	81	2RQR_A	85.6	0.49
Abriata	Luciano	Lousanne Poly	COA6	125	5J4Z_BH	99.6	0.54
Best	Sonja	NIH/NIAID	LIM	206	1RUT_X	99.7	0.72
Folco	Hernan	NIH/NCI	SPAC17C9	94	5MRC_VV	26.1	0.55
Golden	Andy	NIH/NIDDK	EMB1	81	5A31_D	56.1	0.48
Harmer	Stacey	UC Davis	XAP5	187	5NSA_A	56.2	0.2
Jacobs	Dakota	NIH/NCI	NELF_Cterm200	202	2MR5_A	64.6	0.29
Jacobs	Dakota	NIH/NCI	NELF	532	2BL0_A	83	0.07

3)

Mes

CAS rele

In a deci prec rele part forn asse rour

CAS resu onli

Dea all c lear

lastname	firstname	institution	domain	length	HHsearch top templ	HHsearch prob	HHsearch coverage
Abriata	Luciano	Lousanne Poly	DHHC6_SH3	81	2RQR_A	85.6	0.49
Abriata	Luciano	Lousanne Poly	COA6	125	5J4Z_BH	99.6	0.54
Best	Sonja	NIH/NIAID	LIM	206	1RUT_X	99.7	0.72
Folco	Hernan	NIH/NCI	SPAC17C9	94	5MRC_VV	26.1	0.55
Golden	Andy	NIH/NIDDK	EMB1	81	5A31_D	56.1	0.48
Harmer	Stacey	UC Davis	XAP5	187	5NSA_A	56.2	0.2
Jacobs	Dakota	NIH/NCI	NELF_Cterm200	202	2MR5_A	64.6	0.29
Jacobs	Dakota	NIH/NCI	NELF	532	2BL0_A	83	0.07
Kimura	Shioko	NIH/NCI	SCGB_3A2	93	1CCD_A	97.7	0.69
Koepnick	Brian	UW, Baker Lab	UW_engnr	80	3HF3_D	71.6	0.85
Liang	Jake	NIH/NIDDK	HBx	154	6EU1_N	19	0.12
Masison	Cynthia	NIH/NCI	P12C39A	99	2KIX_D	16.1	0.15
Maurizi	Michael	NIH/NCI	PinA	161	2QL2_B	65.3	0.14
Michelmore	Richard	UC Davis	BlRXLR3	141	2LC2_A	95.8	0.55
Mushegian	Arcady	NSF/MCB, Virginia	NP_062900	189	3PUN_B	53.9	0.48
Myrum	Craig	NIH/NIA/IRP	ARC_Nterm1-140	140	1TQG_A	43.7	0.16
Myrum	Craig	NIH/NIA/IRP	ARC	396	4X3X_A	100	0.22
Prosser	Gareth	NIH/NIAID	EFD57440	107	3HFE_A	32.4	0.13
Rein	Alan	NIH/NCI	EIAV	68	4K02_B	11.9	0.31
Robert	Hufnagel	NIH/NEI	PNPLA6	167	5FYA_B	99.9	0.98
Schneider	Thomas	NIH	RepA	286	1REP_C	99.5	0.7
Schwartz	Daniella	NIH/NIAMS	NBCe1-B	160	5JHO_A	99.9	0.48
Ten-Hagen	Kelly	NIH/NIDCR	NP_476718	74	2HNW_B	37.6	0.14
Wang	Qinglan	NIH	PPE_Mtb	181	1KRK_.	16.4	0.04
West	Jennifer	NIH/NIDDK	MotB	162	4YTK_A	28.6	0.25
Yang-Yen	Hsin-Fang	IMB, Taiwan	PRAP1	149	1Y0G_C	29.7	0.11
Yarden	Oded	Hebrew U of Jerusalem	SSP1	149	5GNA_A	53.7	0.15
Young	Howard	NIH/NCI	BAE31734	78	2AAZ_B	38.6	0.47
Zhang	Zhiyong	Rutgers U	Etglo4	216	2LVF_A	98.2	0.3
Zhang	Zhiyong	Rutgers U	Etglo1	216	1S6D_A	98.2	0.31
Zhang	Zhiyong	Rutgers U	Etglo2	258	1S6D_A	98.1	0.25
Zhang	Zhiyong	Rutgers U	Etglo3	290	1S6D_A	97.9	0.22

CASP_Commons

[Register for CASP_Commons](#)

[Released Target List](#)

[Submit a prediction](#)

[Model Accuracy Estimates](#)

3)

Modeling targets

lastname	firstname	institution	domain	length	HHsearch top templ	HHsearch prob	HHsearch coverage
Abriata	Luciano	Lousanne Poly	DHHC6_SH3	81	2RQR_A	85.6	0.49
Abriata	Luciano	Lousanne Poly	COA6	125	5J4Z_BH	99.6	0.54
Best	Sonja	NIH/NIAID	LIM	206	1RUT_X	99.7	0.72
Folco	Hernan	NIH/NCI	SPAC17C9	94	5MRC_VV	26.1	0.55
Golden	Andy	NIH/NIDDK	EMB1	81	5A31_D	56.1	0.48
Harmer	Stacey	UC Davis	XAP5	187	5NSA_A	56.2	0.2
Jacobs	Dakota	NIH/NCI	NELF_Cterm200	202	2MR5_A	64.6	0.29
Jacobs	Dakota	NIH/NCI	NELF	532	2BL0_A	83	0.07

Mes
CAS
rele
In a
deci
pred
rele
part
form
asse
rour
CAS
resu
onli
Dea
all c
lear

CASP_Commons

Accuracy Estimates

C0001 (EIAV)

local score scale: (0.8; 1.0) (0.6; 0.8) (0.4; 0.6) (0.2; 0.4) (0.0; 0.2)

#	Model	ProQ3	QAcons.
1	C0001TS305_2	0.434	0.373
2	C0001TS305_4	0.418	0.330
3	C0001TS258_2	0.403	0.365
4	C0001TS001_3	0.395	0.414
5	C0001TS258_3	0.369	0.375
6	C0001TS308_3	0.351	0.327
7	C0001TS373_3	0.330	0.397
8	C0001TS001_2	0.329	0.396
9	C0001TS373_4	0.326	0.402
10	C0001TS305_5	0.317	0.341
11	C0001TS258_5	0.313	0.358
12	C0001TS258_4	0.313	0.381
13	C0001TS305_3	0.312	0.330
14	C0001TS308_1	0.311	0.331
15	C0001TS258_1	0.311	0.370
16	C0001TS305_1	0.304	0.359
17	C0001TS005_1	0.304	0.390
18	C0001TS308_2	0.284	0.337
19	C0001TS230_4	0.284	0.323
20	C0001TS005_5	0.276	0.389
21	C0001TS373_5	0.274	0.404
22	C0001TS230_2	0.267	0.332

Model Accuracy Estimates

Target: C0009 (ARC_Nterm1-140)

local score scale: (0.8; 1.0) (0.6; 0.8) (0.4; 0.6) (0.2; 0.4) (0.0; 0.2)

#	Model	ProQ3	QAcons.
1	C0009TS258_2	0.591	0.274
2	C0009TS305_2	0.568	0.216
3	C0009TS258_5	0.560	0.262
4	C0009TS258_3	0.557	0.278
5	C0009TS258_4	0.540	0.281
6	C0009TS373_1	0.526	0.302
7	C0009TS373_5	0.522	0.290
8	C0009TS305_4	0.518	0.192
9	C0009TS373_2	0.511	0.233
10	C0009TS258_1	0.510	0.281
11	C0009TS102_1	0.481	0.244
12	C0009TS005_3	0.480	0.291
13	C0009TS001_1	0.480	0.289
14	C0009TS001_5	0.458	0.306
15	C0009TS305_3	0.453	0.186
16	C0009TS005_4	0.447	0.285
17	C0009TS373_3	0.440	0.224
18	C0009TS305_5	0.424	0.192
19	C0009TS005_5	0.423	0.285
20	C0009TS001_4	0.418	0.283
21	C0009TS305_1	0.416	0.219

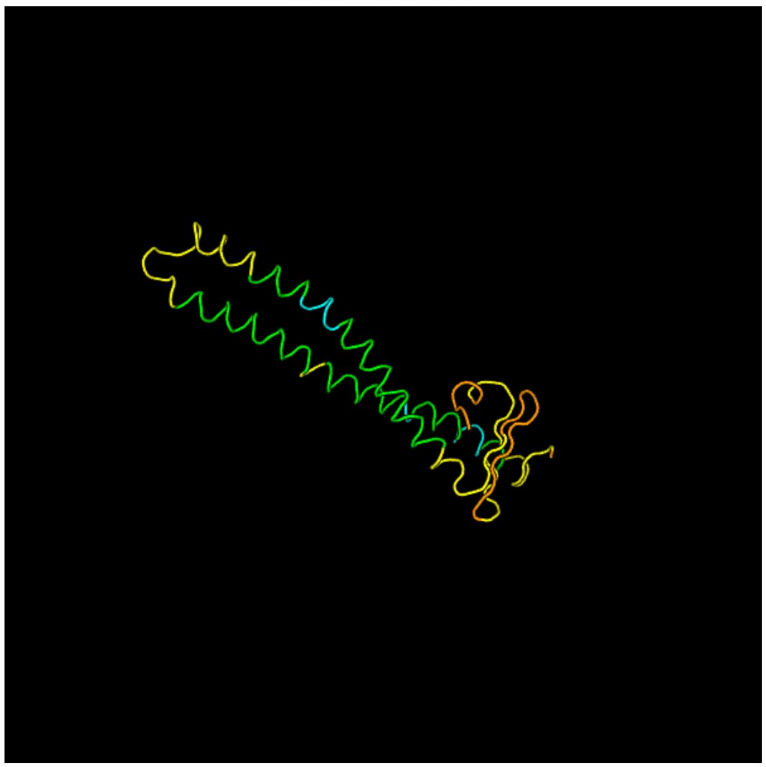
Model Accuracy Estimates

Target: C0009 (ARC_Nterm1-140)

local score scale: (0.8; 1.0) (0.6; 0.8) (0.4; 0.6) (0.2; 0.4) (0.0; 0.2)

#	Model	ProQ3	QAcons.
1	C0009TS258_2	0.591	0.274
2	C0009TS305_2	0.568	0.216
3	C0009TS258_5	0.560	0.262
4	C0009TS258_3	0.557	0.278
5	C0009TS258_4	0.540	0.281
6	C0009TS373_1	0.526	0.302
7	C0009TS373_5	0.522	0.290
8	C0009TS305_4	0.518	0.192
9	C0009TS373_2	0.511	0.233
10	C0009TS25	0.51	0.281
11	C0009TS1	0.48	0.244
12	C0009TS0	0.48	0.291
13	C0009TS0	0.48	0.289
14	C0009TS0	0.45	0.306
15	C0009TS3	0.45	0.186
16	C0009TS0	0.44	0.285
17	C0009TS37	0.44	0.224
18	C0009TS3	0.42	0.192
19	C0009TS0	0.42	0.285
20	C0009TS0	0.41	0.283
21	C0009TS3	0.41	0.219

Model C0009TS258_2



Model C0009TS305_2

Model Accuracy Estimates

Target: C0017 (SCGB_3A2)

local score scale: (0.8; 1.0) (0.6; 0.8) (0.4; 0.6) (0.2; 0.4) (0.0; 0.2)

#	Model	10 20 30 40 50 60 70 80 90	ProQ3	10 20 30 40 50 60 70 80 90	QAcons.
1	C0017TS305_4		0.685		0.372
2	C0017TS308_4		0.685		0.372
3	C0017TS258_1		0.658		0.342
4	C0017TS258_3		0.657		0.303
5	C0017TS305_1		0.649		0.354
6	C0017TS308_1		0.644		0.354
7	C0017TS258_2		0.629		0.341
8	C0017TS305_3		0.615		0.318
9	C0017TS308_3		0.590		0.318
10	C0017TS305_5		0.577		0.363
11	C0017TS308_5		0.577		0.363
12	C0017TS258_4		0.554		0.319
13	C0017TS258_5		0.538		0.312
14	C0017TS308_2		0.481		0.360
15	C0017TS305_2		0.478		0.360
16	C0017TS005_1		0.471		0.388

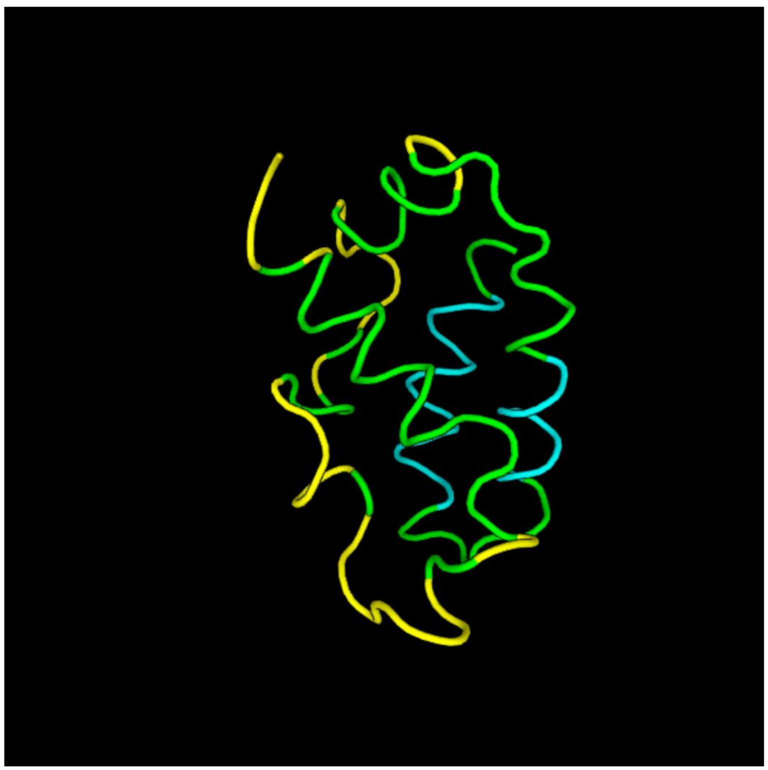
Model Accuracy Estimates

Target: C0017 (SCGB_3A2)

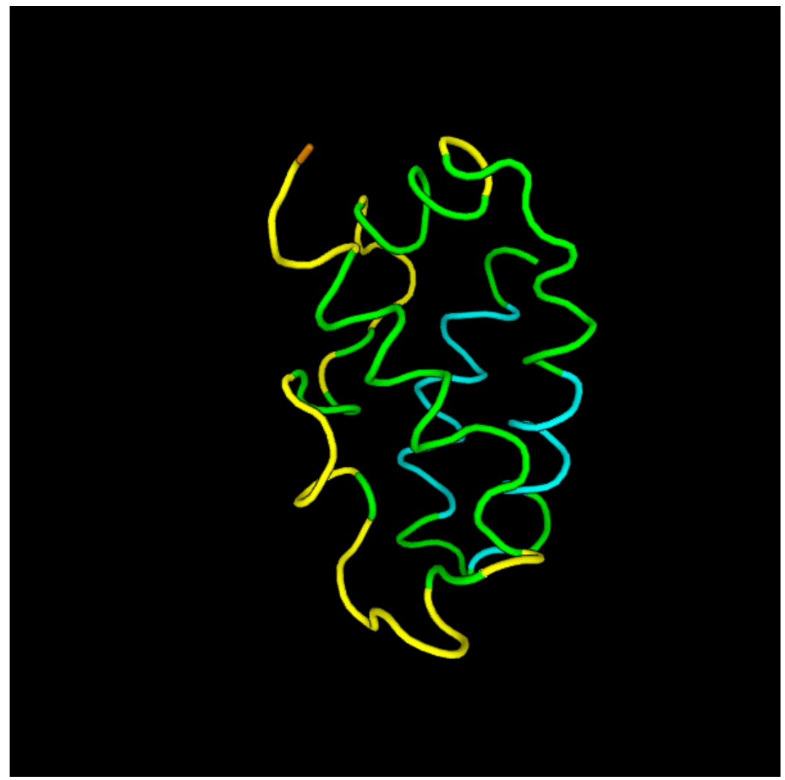
local score scale: (0.8; 1.0) (0.6; 0.8) (0.4; 0.6) (0.2; 0.4) (0.0; 0.2)

#	Model	ProQ3	QAcons.
1	C0017TS305_4	0.685	0.372
2	C0017TS308_4	0.685	0.372
3	C0017TS258_1	0.658	0.342
4	C0017TS258_3	0.657	0.303
5	C0017TS305_1	0.649	0.354
6	C0017TS308_1	0.644	0.354
7	C0017TS258_	0.629	0.341
8	C0017TS305_	0.615	0.318
9	C0017TS308_	0.590	0.318
10	C0017TS305_	0.577	0.363
11	C0017TS308_	0.577	0.363
12	C0017TS258_	0.554	0.319
13	C0017TS258_	0.538	0.312
14	C0017TS308_	0.481	0.360
15	C0017TS305_	0.478	0.360
16	C0017TS005_	0.471	0.388

Model C0017TS305_4



Model C0017TS308_4



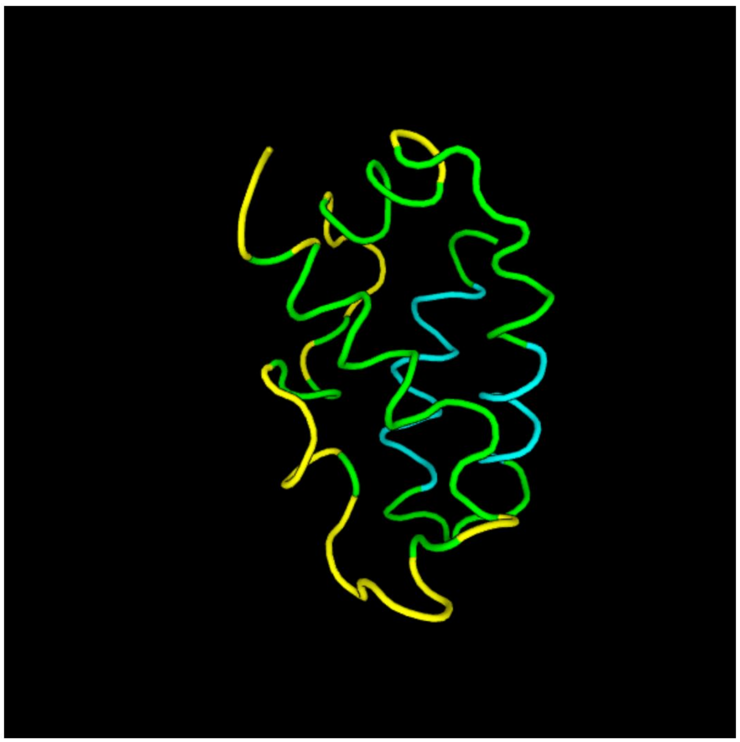
Model Accuracy Estimates

Target: C0017 (SCGB_3A2)

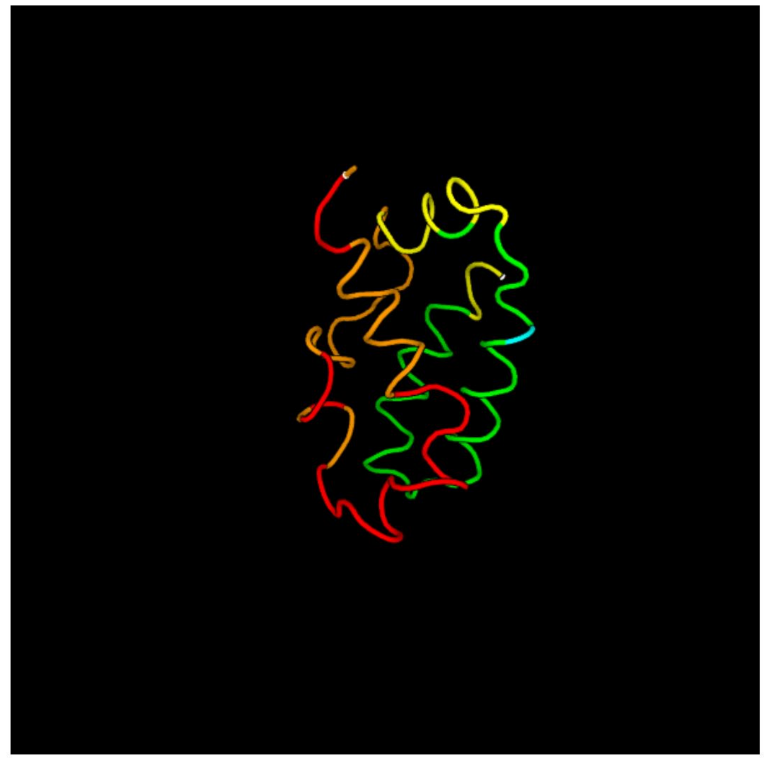
local score scale: (0.8; 1.0) (0.6; 0.8) (0.4; 0.6) (0.2; 0.4) (0.0; 0.2)

#	Model	10 20 30 40 50 60 70 80 90	ProQ3	10 20 30 40 50 60 70 80 90	QAcons.
1	C0017TS305_4		0.685		0.372
2	C0017TS308_4		0.685		0.372
3	C0017TS258_1		0.658		0.342
4	C0017TS258_3		0.657		0.303
5	C0017TS305_1		0.649		0.354
6	C0017TS308_1		0.644		0.354
7	C0017TS		0.629		.341
8	C0017TS		0.615		.318
9	C0017TS		0.590		.318
10	C0017TS		0.577		.363
11	C0017TS		0.577		.363
12	C0017TS		0.554		.319
13	C0017TS		0.538		.312
14	C0017TS		0.481		.360
15	C0017TS		0.478		.360
16	C0017TS		0.471		.388

Model C0017TS305_4



Model C0017TS308_4



Data Collection and Data Guided Prediction

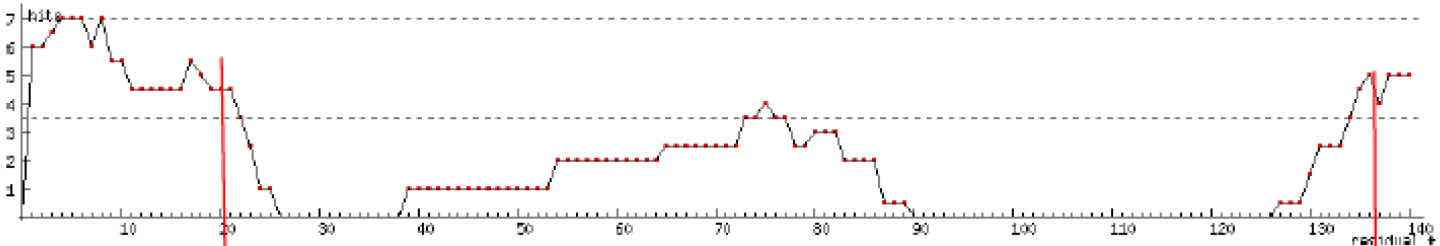
Guy 5 min

CASP Commons Data-Assisted Prediction

- CASP Commons Protein Targets for
  - 80 – 180 residues
  - solicited from biomedical research community
  - no good templates; shallow sequence alignments
- 15 targets selected for sample production
  - Gene synthesis, expression, purification
  - Oligomer state by Analytical Gel Filtration with Static Light Scattering
  - ^{15}N-HSQC spectrum
  - Proteins provided to LBL for SAXS
- Samples are then prepared with ^{13}C,^{15}N isotope-enrichment for NMR studies
- Data to be provided to predictors
  - Backbone resonance assignments
  - Backbone dihedral angles from Talos
  - Ambiguous contacts from ^{15}N-edited NOESY (no RDC data).
- Reference structures will be completed by NMR and/or X-ray crystallography

ANCHOR protein binding site (disordered)

Disorder Consensus (7 predictors used)



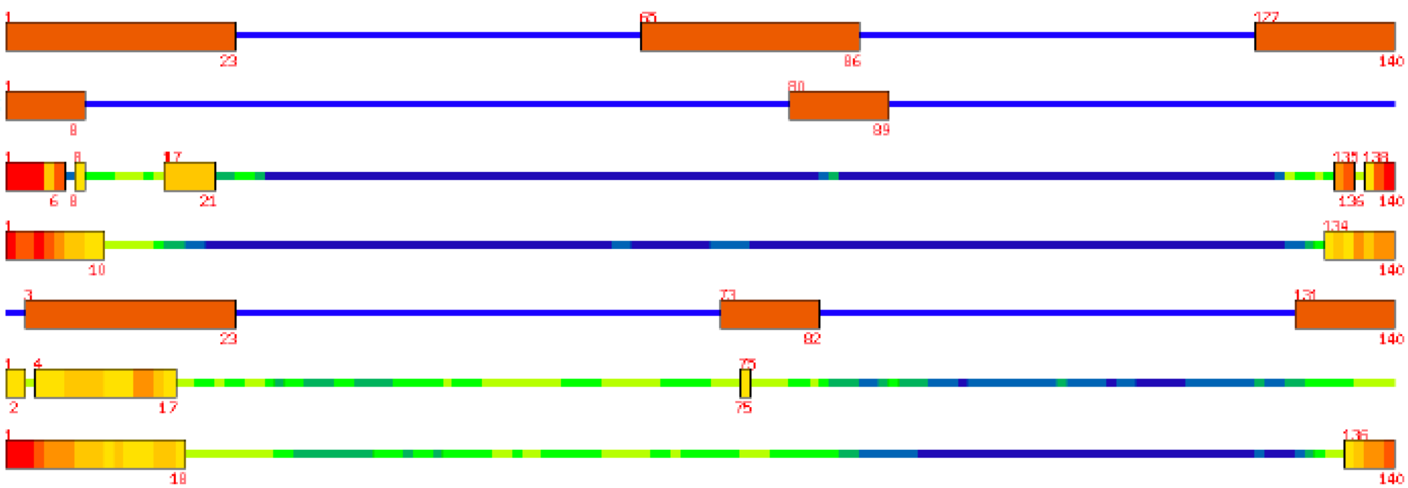
Detailed Disordered Region Prediction Results:

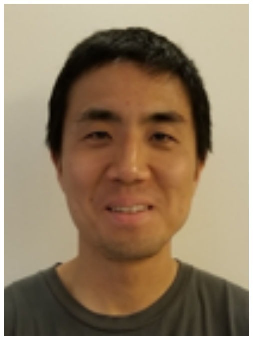
Failed to get the results of the following predictors:**FoldIndex,FoldUnfold,DRIPPRED**

with score		**w/o score**	
color	score	color	structure
	<10%		ordered
	10%-20%		disordered
	20%-30%		
	30%-40%		
	40%-50%		
	50%-60%		
	60%-70%		
	70%-80%		
	80%-90%		
	90%-100%		

```
        10         20         30         40         50         60
MELDHRTSGG LHAYPGPRGG QVAKPNVILQ IGKCRAEMLE HVRRTHRHLL AEVSKQVERE
        70         80         90        100        110        120
LKGLHRSVGK LESNLDGYVP TSDSQRWKKS IKACLCRCQE TIANLERWVK REMHVWREVF
       130        140
YRLERWADRL ESTGGKYPVG
a single construct 20 to 136 is recommended
```

Predictor	**Parameter**	**Disorder_prediction**
DISEMBL	Hot-loops	
DISEMBL	Remark-465	
DISOPRED	5_fp	
DISpro	default	
GlobPlot2	default	
IUPred	long	
IUPred	short	





Yojiro
Ishida

Small Scale Expression and Solubility Screening

Genes synthesized and expressed with
N-terminal His6-TEV-tag

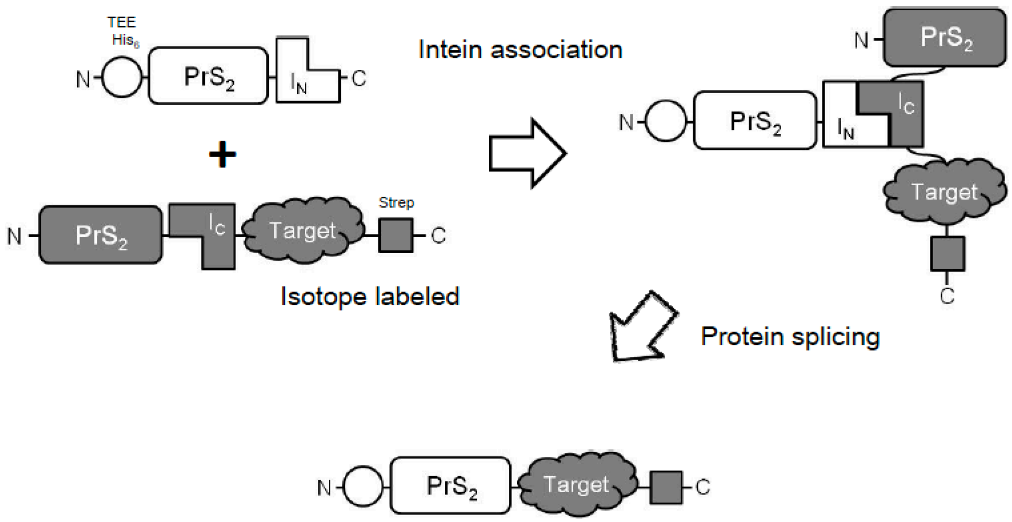
			LB/17°C/o/n		MJ9/17°C/o/n	
			T	S	T	S
1	EIAV	10.46 / 4564.42	3	2	3	1
2	EIAVfl (precipitates)	10.24 / 7882.21	3	0	3	0
3	PRAP1	4.55 / 14,700	3	5	5	5
4	EMB1	5.76 / 9310.49	1	3	2	1
5	UW_engnr	5.12 / 9539.65	3	5	3	5
6	MotB (no expression	9.23 / 18217.19	2	0	4	0
7	BAE31734	9.00 / 4256.82	2	0	1	0
8	PinA (precipitate)	4.48 / 18816.14	2	0	5	5
9	BIRXLR3	6.67 / 13949.97	1	0	3	3
10	ARC_Nterm1-140	9.78 / 13778.94	2	0	4	0
11	PPE_Mtb some prec	4.35 / 17502.62	1	0	3	1
12	XAP5	8.18 / 19222.72	3	1	3	1
13	NP_062900	9.46 / 20356.63	1	0	3	0
14	Hbxfl some precipitat	8.58 / 16565.26	3	0	3	0
15	Hbx	8.63 / 11807.75	1	0	1	0
16	EFD57440 (Carlie)	5.82 / 11519.86	2	0	2	0

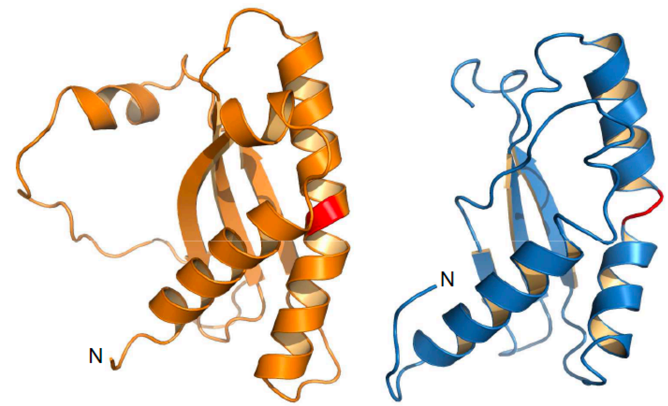
J Biomol NMR (2012) 52:303–313
DOI 10.1007/s10858-012-9610-0

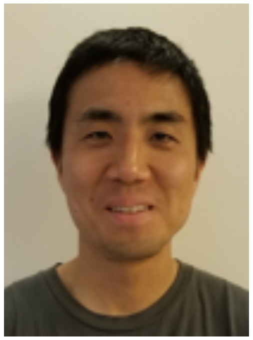
ARTICLE

Segmental isotope labeling of proteins for NMR structural study using a protein S tag for higher expression and solubility

Hiroshi Kobayashi · G. V. T. Swapna · Kuen-Phon Wu · Yuliya Afinogenova · Kenith Conover · Binchen Mao · Gaetano T. Montelione · Masayori Inouye







Yojiro

Small Scale Expression and Solubility Screening

			LB/17°C/o/n		MJ9/17°C/o/n		17C/PST	solubility with PST	NMR samples	NMR quality	SAXS
			T	S	T	S	T	S	Yes or No	Good or bac	comments
1	EIAV	10.46 / 4564.42	3	2	3	1	5	5			sent
2	EIAVfl (precipitates)	10.24 / 7882.21	3	0	3	0	5	5			sent
3	PRAP1	4.55 / 14,700	3	5	5	5	n/a	n/a	Yes	done	yes
4	EMB1	5.76 / 9310.49	1	3	2	1	5	5			yes
5	UW_engnr	5.12 / 9539.65	3	5	3	5	n/a	n/a	Yes	done	yes
6	MotB (no expression	9.23 / 18217.19	2	0	4	0	2	5	Yes	bad	need more work
7	BAE31734	9.00 / 4256.82	2	0	1	0	5	5	Yes	good	sent
8	PinA (precipitate)	4.48 / 18816.14	2	0	5	5	5	5	Yes		sent
9	BIRXLR3	6.67 / 13949.97	1	0	3	3	5	5			
10	ARC_Nterm1-140	9.78 / 13778.94	2	0	4	0	0	0	N/A	N/A	N/A
11	PPE_Mtb some prec	4.35 / 17502.62	1	0	3	1	5	5			sent
12	XAP5	8.18 / 19222.72	3	1	3	1	5	5	Yes	good	sent
13	NP_062900	9.46 / 20356.63	1	0	3	0	5	0			sent
14	Hbxfl some precipitat	8.58 / 16565.26	3	0	3	0	5	5			
15	Hbx	8.63 / 11807.75	1	0	1	0	5	2			
16	EFD57440 (Carlie)	5.82 / 11519.86	2	0	2	0	5	5			sent

Current Status on Sample Production

Produced for ^{15}N-NMR

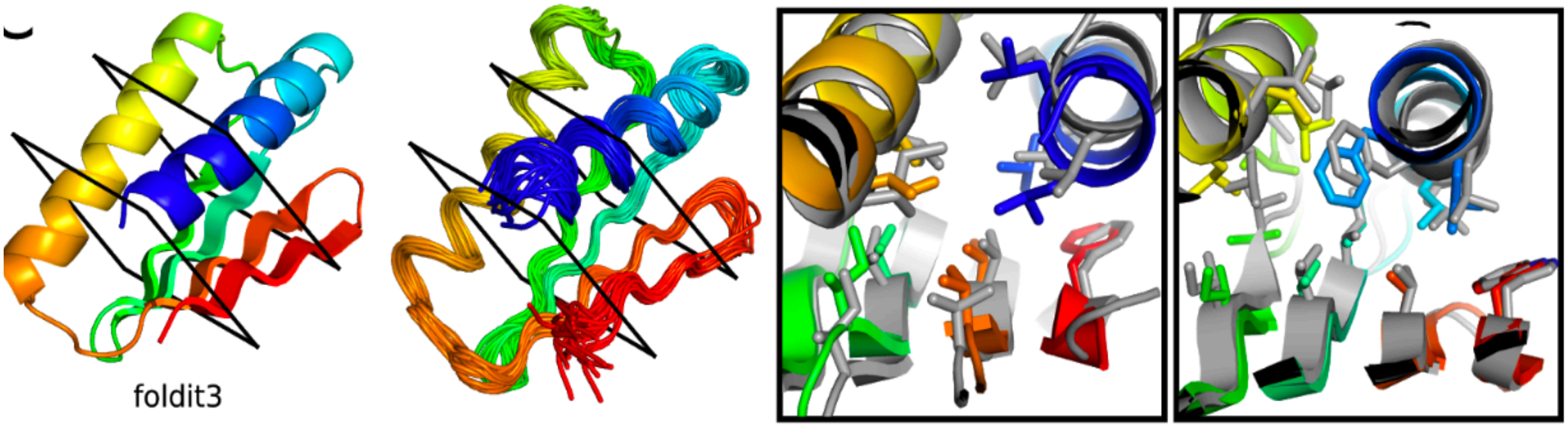
No Tag	PRAP1, UW_eng
His$_6$-tag	MotB, BAE31734, PinA, XAP5
S-tag	EIAV, EIAV_fl, EMB1, MotB, BAE31734, PinA, BlRXLR3, PPE_Mtb, XAP5, Hbl_fl, HBX, EFD57440

Produced for SAX

No Tag	PRAP1, EMB1, UW_eng
With S- tag	EIAV, EIAV_fl, EMB1, BAE31734, PinA, PPE_Mtb, XAP5, EFD57440

First CASP COMMONS Structure
De novo Design by Citizen Scientists

Brian Koepnick, et al., submitted



Gaohua Liu

1.0 Å rmsd between medoid conformer and design structure

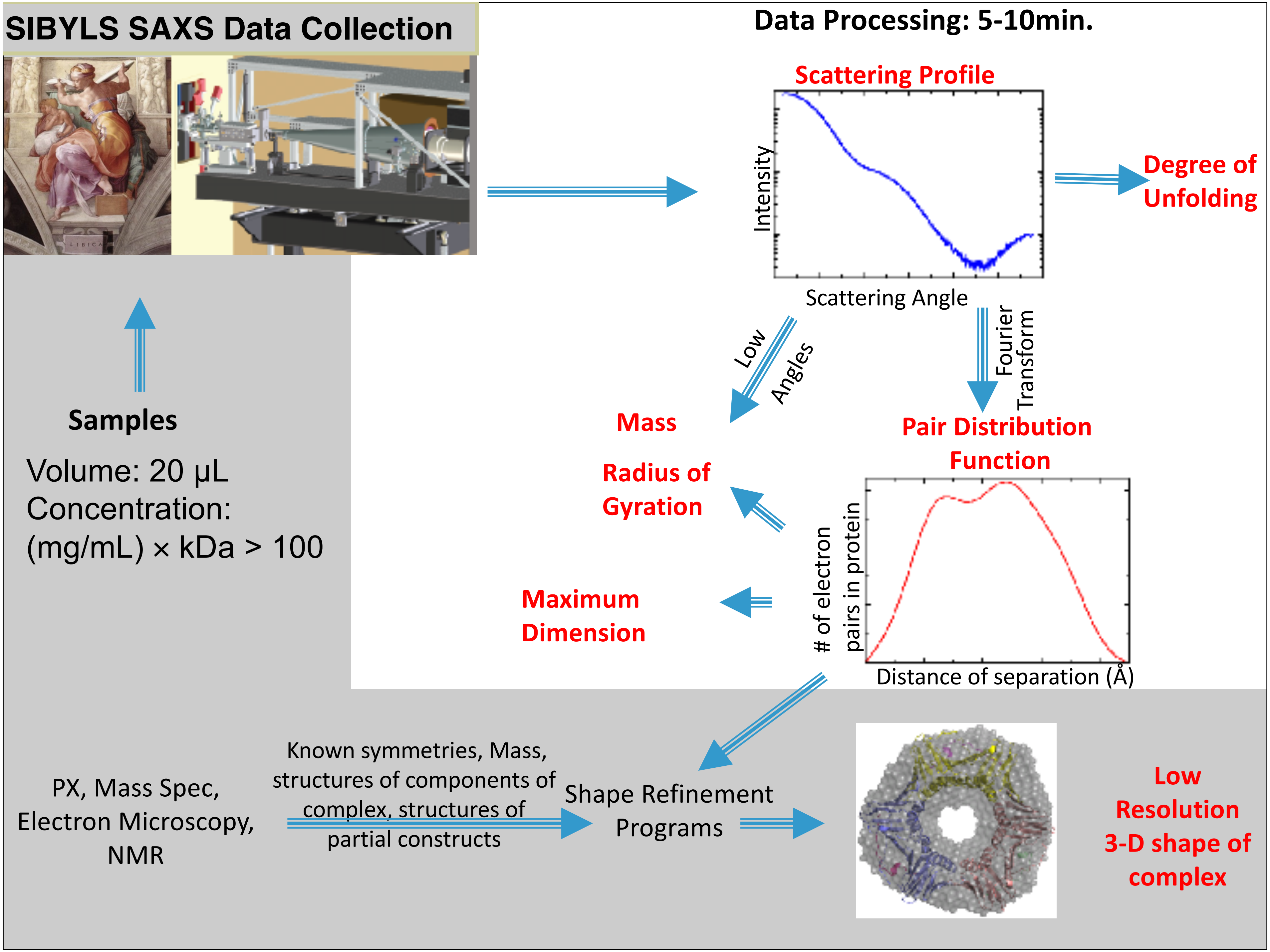
SAXS in CASP COMMONS

Greg Hura 5 min

Coupling Synchrotron X-rays with Robotics SAXS Data is Collected in High Throughput

100 wells in 2 hours with 4 fold improvement expected this year

SIBYLS collects data for 100 groups per year (all would benefit from good predictions)


SIBYLS SAXS Data Collection
Data Processing: 5-10min.
Scattering Profile
Intensity
Scattering Angle
Degree of Unfolding
Low Angles
Fourier Transform
Mass
Radius of Gyration
Maximum Dimension
Pair Distribution Function
# of electron pairs in protein
Distance of separation (Å)
Samples
Volume: 20 µL
Concentration:
(mg/mL) × kDa > 100
PX, Mass Spec, Electron Microscopy, NMR
Known symmetries, Mass, structures of components of complex, structures of partial constructs
Shape Refinement Programs
Low Resolution 3-D shape of complex

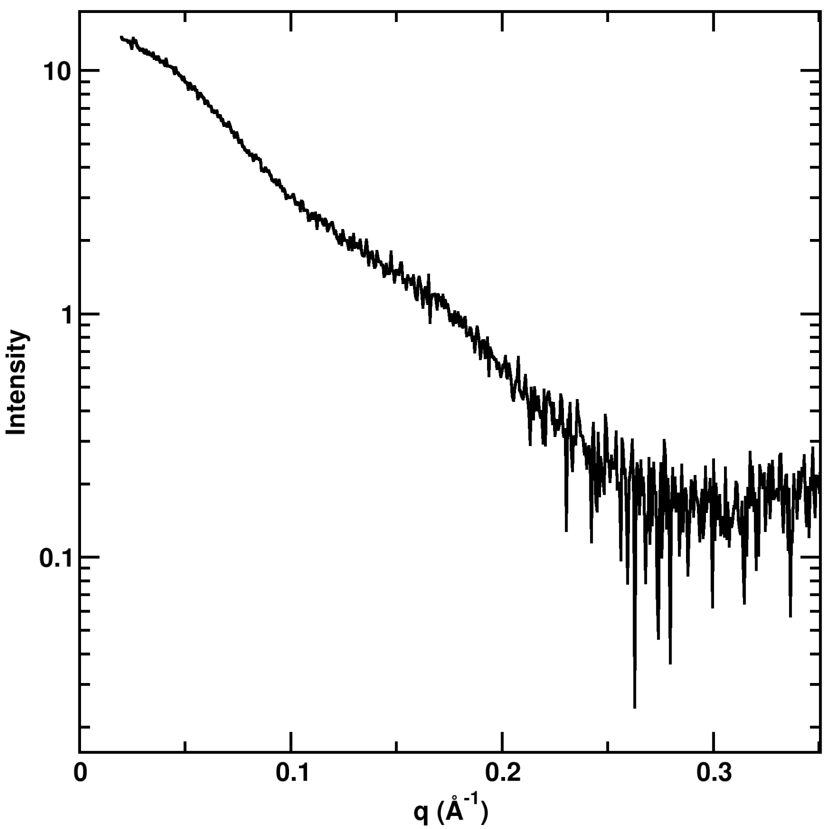
CASP Commons 4– Analysis 7/26/18

No HT-SAXS. Dmax hard to assign

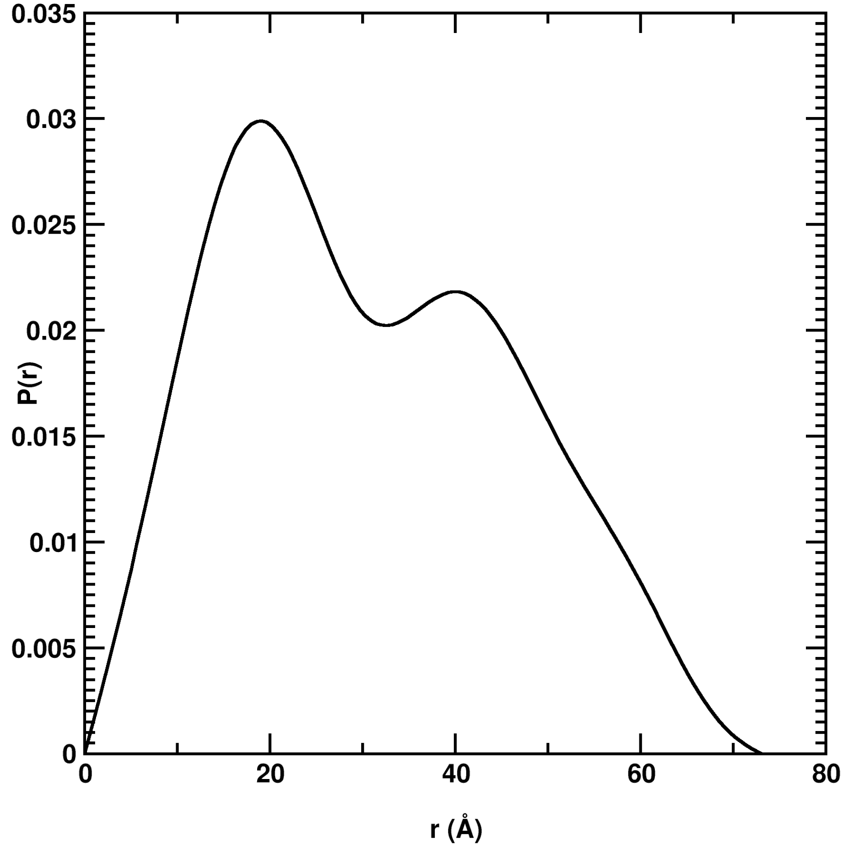
Flexibly linked dimer

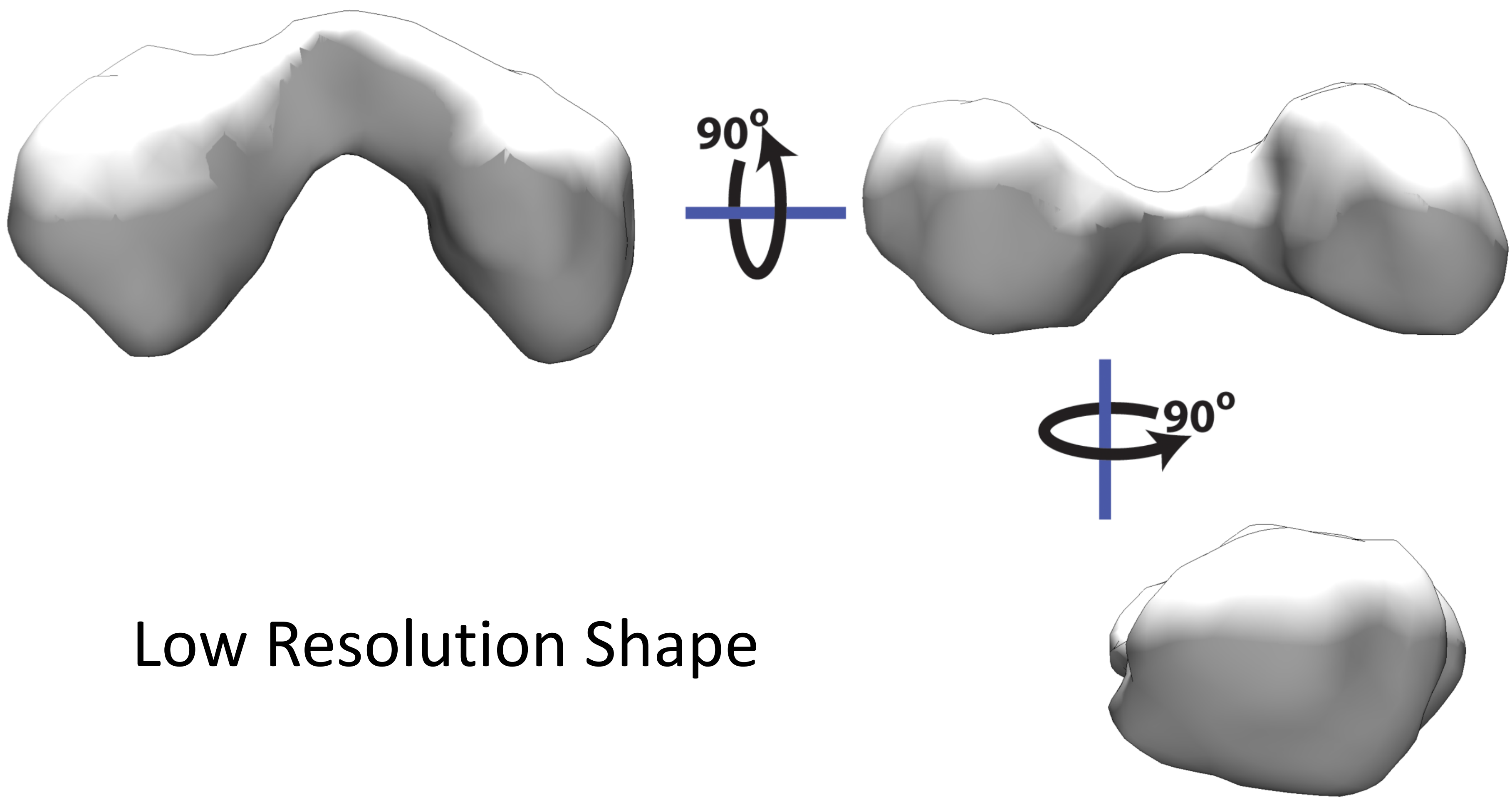
Sample: Commons 4			
Variable	Value	Error +/-	Units
Rg	23.6	1	Angstroms
Porod Exponent	2.0		Scale (2-4)
Mass SAXS	18	5	kDa
Max Dimension	73	3	Angstroms
Radius of Cross Section	11.78	1	Angstroms
Volume	100,000	5,000	Cubic Angstroms
Real Space Rg	24.03	3	Angstroms

MALMYPFHVAQPPLNWSEHLWVSEVSPAKESFITTICEHRQAQWDNQDLLRHLQDSVAILTREDQRH
VNAVHAAANMPANP
Mass 9.3kDa



Histogram of Scaler Distances
Between Atoms within Molecule





90°
90°
Low Resolution Shape

Audience Discussion

Emily Tai, Greg Hura, John Tainer,
John Moult, and Krzysztof Fidelis

How to Provide Data / Reports to the Nominators

- Predicted structures and reliability metrics
- Expression plasmids
- Atomic coordinates by NMR or X-ray Crystallography
- Consulting discussions

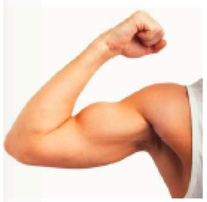
Plan for Future

- Is this a valuable activity for the CASP community?
- What is the incentive for a predictor? For a data provider?
- Can we require the nominating biologist to test the CASP models? How can we get feedback from the biologists?
- How to bring in new targets? Biological focus areas?
- Value of production and characterization of fusion proteins.
- How to obtain funding for these experimental activities?
- Consortium of commercial groups?

PRAP1

Nominator: Yang-Yen IMB Taiwan-China

No HT-SAXS.
Large Variation in max dimension



Largely flexible

CASP Commons 3– Analysis 7/26/18

Sample: Commons 3			
Variable	Value	Error +/-	Units
Rg	36.7	1	Angstroms
Porod Exponent	1.7		Scale (2-4)
Mass SAXS	23	5	kDa
Max Dimension	113-143	3	Angstroms
Radius of Cross Section	17.2	1	Angstroms
Volume	160,000	5,000	Cubic Angstroms
Real Space Rg	35.3	3	Angstroms

GAAPAHQVPVKTKGKHVFPEQETEKVWDTRALEPLEKDNQLGPLLPEPKQKPAAAEEKRPDAMTWVETEDILSHLRSPLQGPELDLDSIDH
PMSDDVQDEEVPQSRPILYRQVLQGPEEDLDHLAHSMEDS
Mass 15kDa

EMB1

Nominator: Andy Golden NIH / NIDDK

No HT-SAXS. Dmax hard to assign

Flexibly linked dimer

CASP Commons 4– Analysis 7/26/18

Sample: Commons 4			
Variable	Value	Error +/-	Units
Rg	23.6	1	Angstroms
Porod Exponent	2.0		Scale (2-4)
Mass SAXS	18	5	kDa
Max Dimension	73	3	Angstroms
Radius of Cross Section	11.78	1	Angstroms
Volume	100,000	5,000	Cubic Angstroms
Real Space Rg	24.03	3	Angstroms

MALMYPFHVAQPPLNWSEHLWVSEVSPAKESFITTICEHRQAQWDNQDLLRHLQDSVAILTREDQRHVNAVHAAANMPANP
Mass 9.3kDa

UW_engnr / foldit3 / CASP n1008

Nominator: Brian Koepnick, David Baker UW

No HT-SAXS.

CASP Commons 5– Analysis 7/26/18

Sample: Commons 5			
Variable	Value	Error +/-	Units
Rg	14.8	1	Angstroms
Porod Exponent	3.9		Scale (2-4)
Mass SAXS	10	5	kDa
Max Dimension	39-43	3	Angstroms
Radius of Cross Section	11.8	1	Angstroms
Volume	17,000	5,000	Cubic Angstroms
Real Space Rg	13.7	3	Angstroms

TDELLERLRQLFEELHERGTEIVVEVHINGERDEIRVRNISKEELKKLLERIREKIEREGSSEVEVNVHSGGQTWTFNEK

Mass 9.5kDa