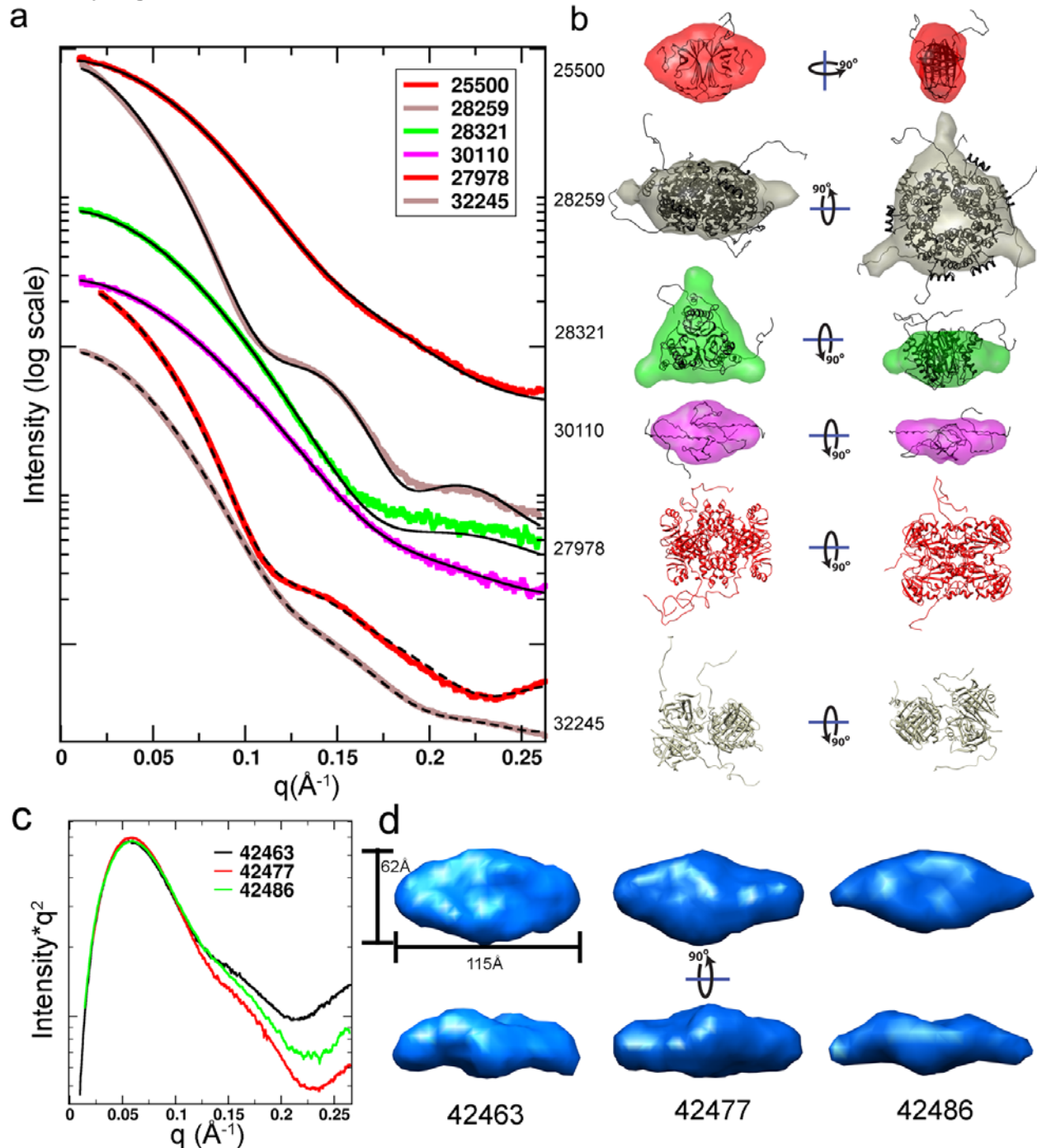
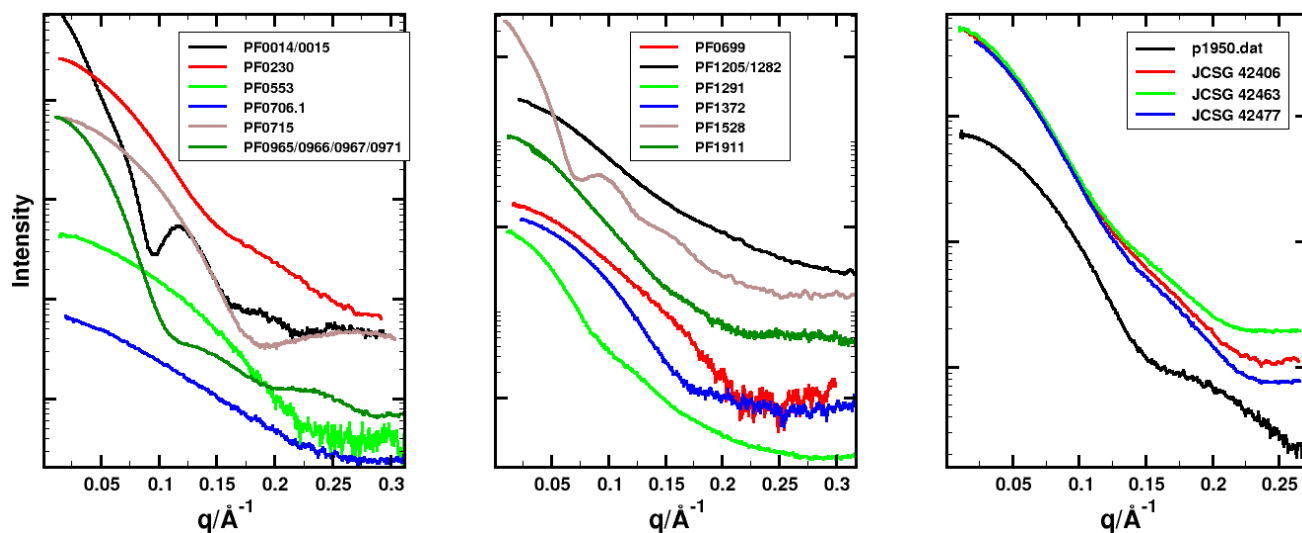


Supplementary Figure 1



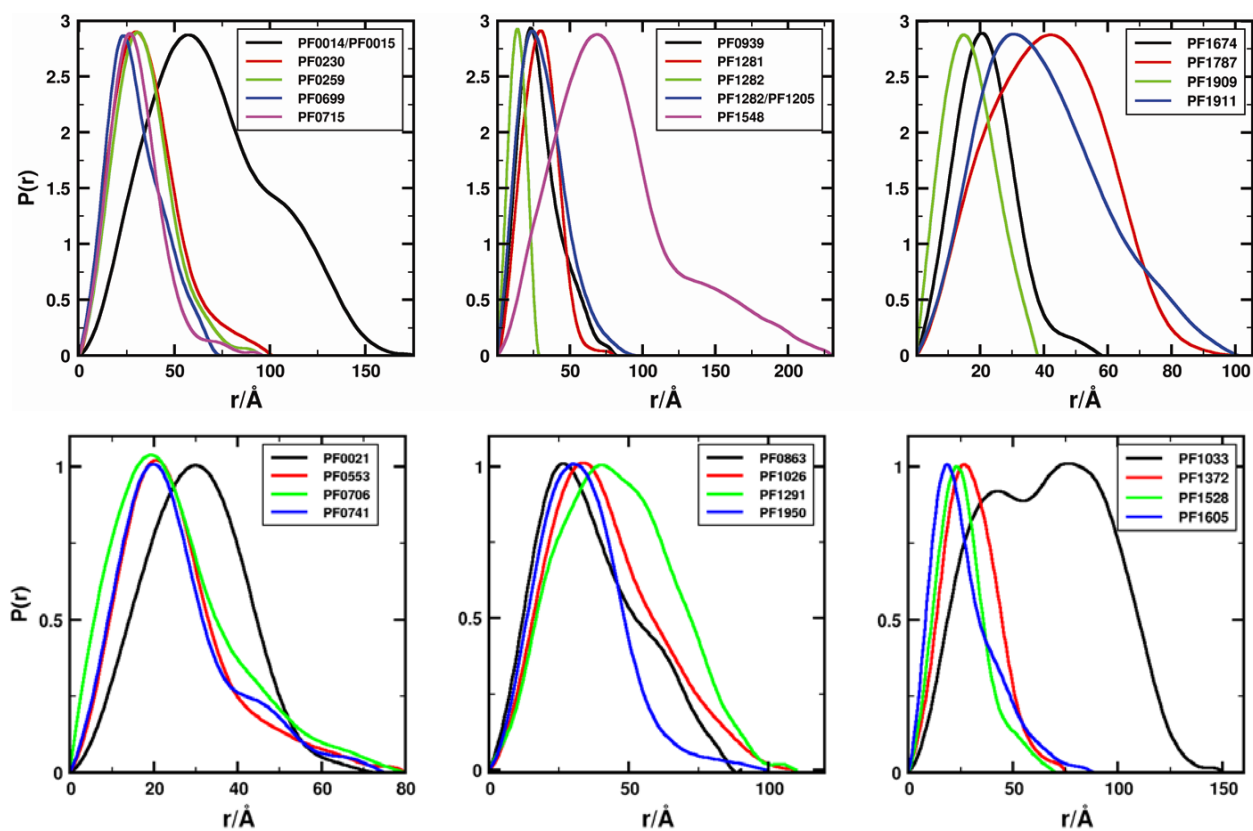
Supplementary Figure 1 SAXS shape and assembly for monodispersed targets from JCSG. (a) For the six proteins with existing structures, the experimental scattering data (colors) were compared with the scattering curve calculated from models based on existing structure (black). The structures were modified to include the His-tags. Conformation space for these His-tags were sampled as described in Pelikan et al.¹ and the single best fit conformation is shown. (b) The envelope determinations (colored as in a) were overlaid with the molded structures (ribbons). For two cases (27978 and 32245) envelopes do not agree in detail with the suggested atomic models. In both cases the flexible regions make up ~15% of the structure. (c) For the 3 proteins with no pre-existing structural information, the Kratky plot is shown. The similarities in the smallest q region suggest the proteins have structural similarities while the wider q decay identifies the most compact form as 42477. (d) GASBOR envelopes are shown for each structure modeled with 2-fold symmetry.

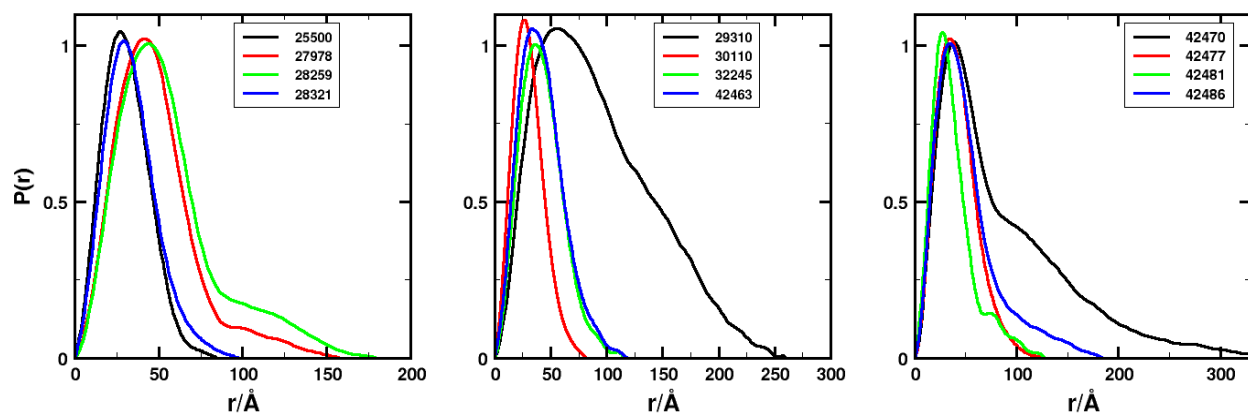
Supplementary Figure 2



Supplementary Figure 2 Scattering profiles of *Pfu* and JCSG samples not shown in main text or above. All SAXS data on each protein will be made available at www.bioisis.net

Supplementary Figure 3





Supplementary Figure 3 Pair distribution $P(r)$ functions for *Pfu* and JCSG samples. Pair distribution functions are determined using a Perl script for GNOM which automatically runs GNOM over many guesses of maximum dimensions (D_{max}). The most appropriate D_{max} has the following properties: the smallest value required to fit the data, smoothly approaches $P(r)=0$, and where a plot of R_g vs guessed D_{max} starts to plateau. The pair distribution function is accessible within 30 seconds of data collection and contains significant structural information. For example the ring like structure of PF1033 is readily apparent from its $P(r)$ function.

Supplementary Table 1 SAXS characterizations for sixteen JCSG samples including determined structural information.

Sample			New Structural Results			
Class	JCSG ID	Topsan ID	R_g	D_{max}	Assemblies	3-D-Envelope
Aggregates	25287	366924			Inseparable	No
	32541	375605			Separable	No
	42470	390709			Separable	No
	42482	390764			Separable	No
Mixtures	29310	372359			Large assembly with concentration dependent multimerization	No
	31314*	374450			Mostly 2-mer	No
	42481*	390756			Small multimers < 4mer	Yes
Matching Crystal Structure	25500	370637	24.0	85	Matching 2-mer	Yes
	27978	371321	35.0	158	Mostly 4-mer	Yes
	28259	371529	44.6	180	Matching 6-mer	Yes
	28321*	371708	26.3	98	Matching 3-mer	Yes
	30110*	371811	23.5	82	Matching 2-mer	Yes
	32245	375018	32.6	116	Matching 2-mer	Yes
Novel Structures	42463*	390716	32.8	120	2-mer	Yes
	42470*	390709	32.4	125	2-mer	Yes
	42486	390749	42.3	185	2-mer	Yes

*indicates samples which were initially aggregates and were aided by centrifuge filtration prior to SAXS experiments.

JCSG ID refers to an internal identification tag. The Topsan ID (The Open Protein Structure Identification Network) catalogs detailed target information at <http://www.topsan.org/>.

Supplementary Table 2 Comparison of radius of gyration (R_G) using two methods and mass based on Porod volume

ORF	R_G (Gunier) (Å)	R_G (from P(r) in (Å))	Mass from Porod Volume (kDa)	Mass from Porod Volume/monomeric mass
PF0014/PF0015	55.0	54.1	370	Unknown stoichiometry
PF0380	21	21.6	20.6	0.72
PF0699	23.7	23.3	25	2.1
PF0715	23.1	23.5	45	2.3
PF1282/PF1205	24.3	25.1	20	1.2
PF1281	22.1	22.1	44	3.1
PF1282	11.0	10.7	3	0.5
PF1674	16.7	16.88	15	0.8
PF1787	33.9	33.9	80	2.9
PF1911	30.9	31.3	60	1.8
PF2047.1	29.7	32	28	3.0

ORF	R_G (Gunier) (Å)	R_G (from P(r) in (Å))	Mass from Porod Volume (kDa)	Mass from I(0)/c	Mass from Porod Volume/monomeric mass
PF0094	28	34.6	40	42	1.6
PF0553	19.2	20.37	15	21	0.75
PF0706.1	18.6	20.74	6	6	1.1
PF0741	20	20.24	12	15	1.1
PF0863	27.4	28.7	35	39	1.95
PF0965/0966/ 0967/0971	36.9	37.3	170	234	
PF1026	31.8	32.32	72	70	1.45
PF1033	51.2	50.8	180	185	7.1
PF1061	17.7	16.3	8	9	1
PF1291	35.6	35.4	120	110	3.6
PF1372	23	23.2	33	37	3.7
PF1528	19.9	20.2	22	18	0.8
PF1605	21	22.0	10	13	1.3
PF1950	25.3	25.9	45	62	2.3

Supplementary Table 2 Global parameters for monodispersed samples. R_G (Gunier) is the radius of gyration extracted from Gunier analysis, while R_G (P(r)) is that extracted from the pair distribution function P(r). The mass from the Porod volume is extracted using a rule of thumb of 0.5 times the Porod Volume. The Porod volume was extracted using Porod button in the PRIMUS program.

1. Pelikan, M., Hura, G.L., & Hammel, M. Structure and flexibility with proteins as identified through small angle x-ray scattering. *General Physiology and Biophysics* (in press June 2009).