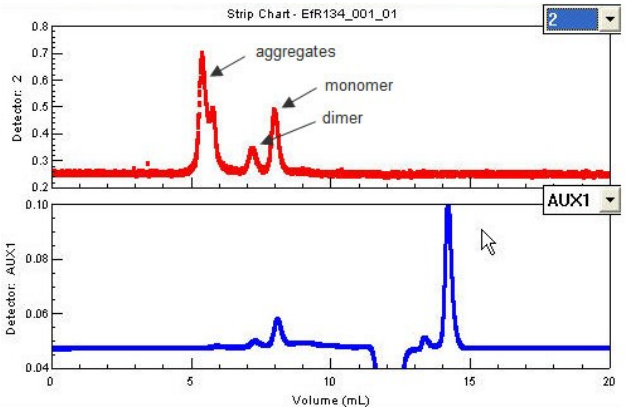
Northeast Structural Genomics Consortium
Protein Aggregation Screening using Analytical Gel Filtration with Static Light Scattering

In protein crystallography, homogeneous monodisperse protein samples are generally thought to be important for obtaining diffraction-quality crystals suitable for structure determination. Analytical gel filtration followed by multi-angle static light scattering (MALS) is an ideal means by which to measure the molar mass distribution of a protein solution. This is an extremely sensitive method for detecting the distribution of monomer, dimer, and/or higher-order oligomers and/or aggregates, as the static light scattering response is directly proportional to the weight-averaged molar mass (MW) of the sample being measured.

In the Northeast Structural Genomics Consortium (NESG; http://www.nesg.org/), a MALS protein aggregation screening system using analytical size-exclusion chromatography (SEC-HPLC) and Wyatt miniDAWN static light scattering system has been part of the standard quality control for protein samples since 2002. Data has been generated and archived in the NESG SPINE database (Goh, 2003) for more than two thousand protein samples (http://spine.nesg.org/index.cgi). Fig. 1 shows one of the miniDAWN's detector outputs, as well as the refractive index signal for NESG target EfR134. SEC-HPLC is done using a Protein 803 KW column from Shodex, (low salt isocratic mobile phase, 0.5 ml/min flow rate, 40 min run time, and 30 microliter sample). For EfR134, several different oligomeric states of the protein were separated into multiple peaks, and the molecular weights for these peaks were determined in the eluate by MALS. The molecular weights detected by MALS correlated with those expected for monomer, dimer, trimer, higher oligomers and aggregates of the protein.



The results are categorized as follows: (i) Monodisperse (> 90% homogenous wrt molecular weight); (ii) Predominantly monodisperse (80 - 90% of total protein in one major non-void peal; (iii) Mostly polydisperse (>50% total protein in major non-void peaks, no more than 3 peaks total); (iv) Polydisperse -- < 50% total protein in major non-void peaks, at least 3 peaks total; (v) Indeterminate -- protein not in void, but molecular weight of major peak obscured by ring-down from void. (vi) Aggregated -- protein in void.

Table 1 summarized data for 2,177 protein samples, along with the rates of crystal structure determination (PDB deposition) for these 6 categories. Crystallization and structure determination success rates are more than 10-fold for monodisperse (21%) or predominantly monodisperse (18%) protein samples compared to samples that are less than 80% monodisperse (1 – 4%). The results demonstrate the tremendous value of the MALS analysis for predicting crystallization success of proteins samples. Additional details of this systematic study are presented in Price (2009).

Table 1. PDB Deposition Rate 2001-2009 (PDB#/Protein Sample#)

Year	Monodisperse	Predominantly monodisperse	Mostly polydisperse	Polydisperse	Indeterminate	Aggregated
2001	1/2 (50.0%)					
2002	24/82 (29.3%)	0/4 (0.0%)				
2003	14/52 (26.9%)	1/3 (33.3%)				0/2 (0.0%)
2004	42/112 (37.5%)	6/19 31.6%)	0/11 (0.0%)	0/110 (0.0%)		0/35 (0.0%)
2005	37/148 (25.0%)	2/31 (6.5%)	1/9 (11.1%)	0/24 (0.0%)	0/20 (0.0%)	0/27 (0.0%)
2006	37/223 (16.6%)	14/41 (34.1%)	2/32 (6.3%)	2/30 (6.7%)	1/46 (2.2%)	1/25 (4.0%)
2007	47/277 (17.0%)	7/57 (12.3%)	0/26 0.0%)	0/19 (0.0%)	1/69 (1.4%)	1/70 1.4%)
2008	60/273 (22.0%)	4/27 14.8%)	1/6 (19.1%)	0/2 (1.0%)	1/3 (33.3%)	1/12 (8.3%)
2009	29/188 (15.4%)	1/13 (7.7%)	0/11 (0.0%)	0/7 (0.0%)	0/0 (0.0%)	0/29 (0.0%)
Total	291/1357 (21.5%)	35/195 (17.9%)	4/96 (4.2%)	2/192 (1.1%)	2/138 (2.2%)	3/199 (1.5%)

Goh, C.-S.; Lan, N.; Echols, N.; … Acton, T. ; Montelione, G.T.; Gerstein, M. Nucleic Acids Res. 2003, 31: 2833 - 2838. SPINE 2: A system for collaborative structural proteomics within a federated database framework.

Price, W.N.; Chen, Y.; Handelman, … DeTitta, G.T.; Rost, B.; Montelione, G.T.; Hunt, J.F. Nature Biotechnology 2009, 27: 51 - 57. Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data.