New characterization and computational methods for understanding protein aggregation

Ewa Marszal, PhD

FDA/CBER

PEGS 2014

Disclaimer

My comments are an informal communication and represent my own best judgment. These comments do not bind or obligate FDA.

Disclaimer

Also, I do not endorse any of the methods and instruments that I may mention.

Outline

- Protein aggregation aspects
  - A common phenomenon
  - Assurance of product safety and efficacy
- Aggregation prone regions
  - Properties, roles and prediction algorithms
  - Optimization for better product solubility
- Quantification and characterization of aggregates
  - Continuous improvements of analytical tools
  - Examples
- Final thoughts

Aggregation caused by thermodynamic destabilization

- Contributing factors:
  - Mutations
  - Loss of translational fidelity
  - Post-translational modifications
  - Folding intermediates
  - Local fluctuations in the structure
  - Stress

Protein aggregation, a common phenomenon

- *In vivo*
  - In conformational diseases (more than 40 diseases identified)
  - During expression of recombinant proteins
  - Possible during administration of protein products
- *In vitro*
  - During manufacture, storage and handling of protein products

Goal of considering protein aggregates in biologics

To ensure safety, quality and efficacy of the products

Potential impact of protein aggregates on safety and efficacy

- Innate immune system stimulation
  - Inflammation
  - Anaphylactoid reactions
- Adaptive immune system stimulation
  - Anaphylaxis
  - Decreased or lost efficacy
    - Inhibitory antibodies
    - Reduced half-life
  - Autoimmune disease
    - Neutralization of patient's endogenous protein (natural counterpart)
- Cytotoxicity
- Capillary clogging

Aggregation prone regions
APRs

Properties of APRs

- 5 – 15 hydrophobic amino acids in sequence
- β-propensity
- Low net charge
- Packing density
- Generally in the hydrophobic core or at protein interaction sites

Role of APRs in protein aggregation

- Can be exposed under stress
- Considered to be a driving force in protein aggregation
- Form β-structured aggregates
- Their aggregation is selective and specific
- Macromolecular morphology of aggregates: from amorphous to fibrous

APRs prediction algorithms

- Tango
- STVIIE
- Zyggregator
- PAGE
- 3D Profile
- Aggrescan
- PASTA
- SALSA
- Betascan
- Waltz
- FoldAmyloid
- AMYLPRED2
- GAP

Beerten *et al*. (2012) Aggregation prone regions and Gatekeeping residues in protein sequences. *Curr Top Med Chem* 12, 2470-8
Tsolis *et al*. (2013) A consensus method for the prediction of 'aggregation-prone' peptides in globular proteins. *PLoS One* 8(1)
Thanqakani *et al* (2014) GAP: Towards almost hundred percent prediction for β-strand mediated aggregating peptides with distinct morphologies. *Bioinformatics* March 28 [Epub ahead of print]

Frequency of APRs occurrence

- Globular domains – 18% of residues on average in APRs
- Intrinsically disordered proteins – 8% residues
- 5% of globular domains have no APRs
- 40% of unstructured proteins have no APR
- Proteins generally have more than 1 APR

Beerten *et al*. (2012) Aggregation prone regions and gatekeeping residues in protein sequences. *Curr Topics Med Chem*, 12

Properties of APRs predicted in different databases by Tango

Dataset	Sequences with ≥ 1 APR (%)	Average Sequence length	Total number of APRs	Average APR length	Average proportion of APR residues (%)
R10000	68.9	100	10474	8.83 ± 3.15	13.42 ± 7.22
N10000	62.6	100	8965	8.49 ± 2.88	12.16 ± 6.33
SF49500	71.4	149 ± 35	59878	8.48 ± 2.85	10.27 ± 5.40
F495	60.4	152 ± 34	452	7.47 ± 1.99	7.85 ± 4.08
All-α	63.6	139 ± 37	62	7.79 ± 2.73	8.91 ± 4.89
All-β	64.2	150 ± 38	47	7.60 ± 2.06	7.63 ± 3.17
A+β	63.6	153 ± 31	82	7.23 ± 1.82	7.20 ± 2.81
α/β	79.1	167 ± 22	66	7.11 ± 1.42	8.35 ± 3.53
Others	57.1	151 ± 34	203	7.55 ± 1.91	7.65 ± 4.44
IDP536	75.2	763 ± 562	1316	9.13 ± 3.9	6.95 ± 5.37

Adapted from Buck *et al*. (2013) On the Role of Aggregation Prone Regions in Protein Evolution, Stability, and Enzymatic Catalysis: Insights from Diverse Analyses. *PLoS Comput Biol* 9(10)

PLOS | COMPUTATIONAL BIOLOGY

Aggregation gate keepers

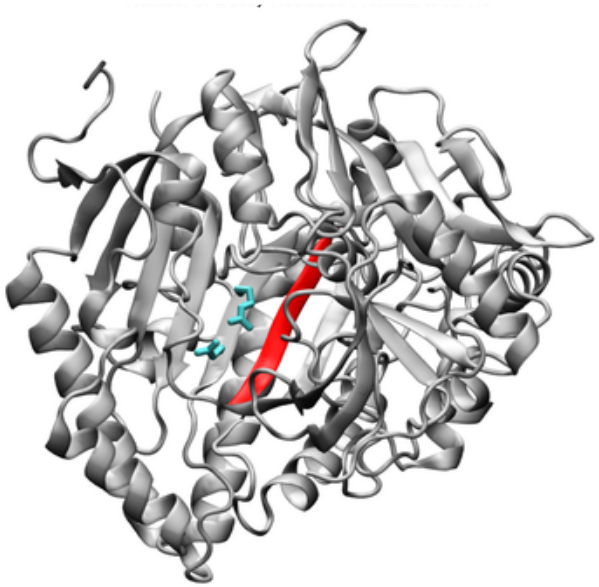
- Charged residues
  - Arg, Lys, Asp, Glu
  - Entropic factor
- Amino acids incompatible with β-structure
  - Pro, Gly

Role of APRs in protein folding, stability, and function

- Proposed to play a role in promoting structural order in native folds
  - Based on the tendency to be in ordered sequence regions
  - Based on prevalence in protein-protein interaction sites
- APRs conservation
  - 'close homologues' with the same fold (≥ 80% sequence identity) – conservation slightly lower than average sequence identity
  - 'distant homologues' (≥ 50% sequence identity) conservation higher than average sequence identity
- Stronger stabilizing effect than other equal length sequences
- Catalytic residues in close proximity

Buck *et al*. (2013) On the Role of Aggregation Prone Regions in Protein Evolution, Stability, and Enzymatic Catalysis: Insights from Diverse Analyses. *PLoS Comput Biol* 9(10)

Catalytic residues and APRs



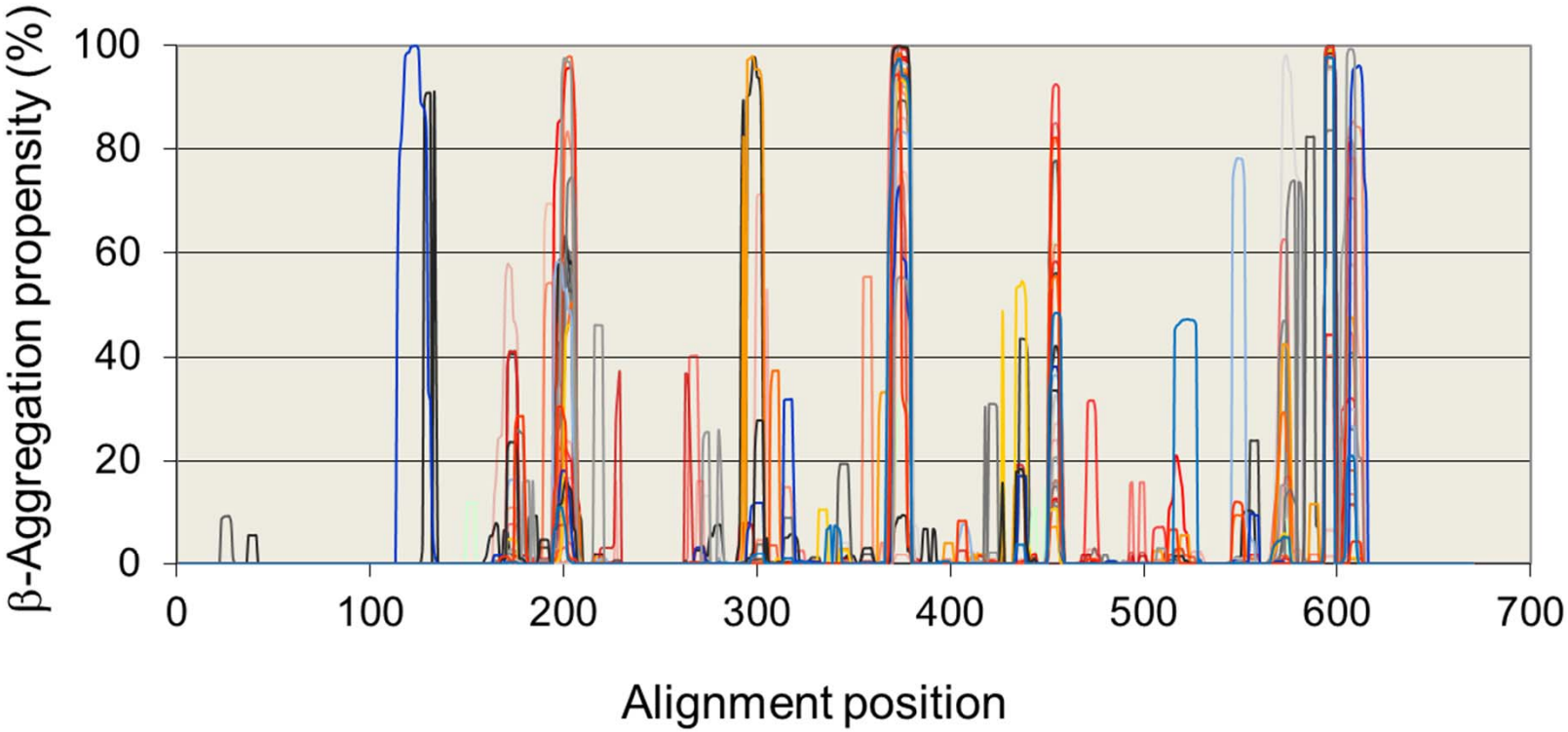
Catalytic residues are in contact with APRs more often than expected by random chance

Cholesterol oxidase from *B. Sterolicum*

Buck *et al*. (2013) On the Role of Aggregation Prone Regions in Protein Evolution, Stability, and Enzymatic Catalysis: Insights from Diverse Analyses. *PLoS Comput Biol* 9(10)

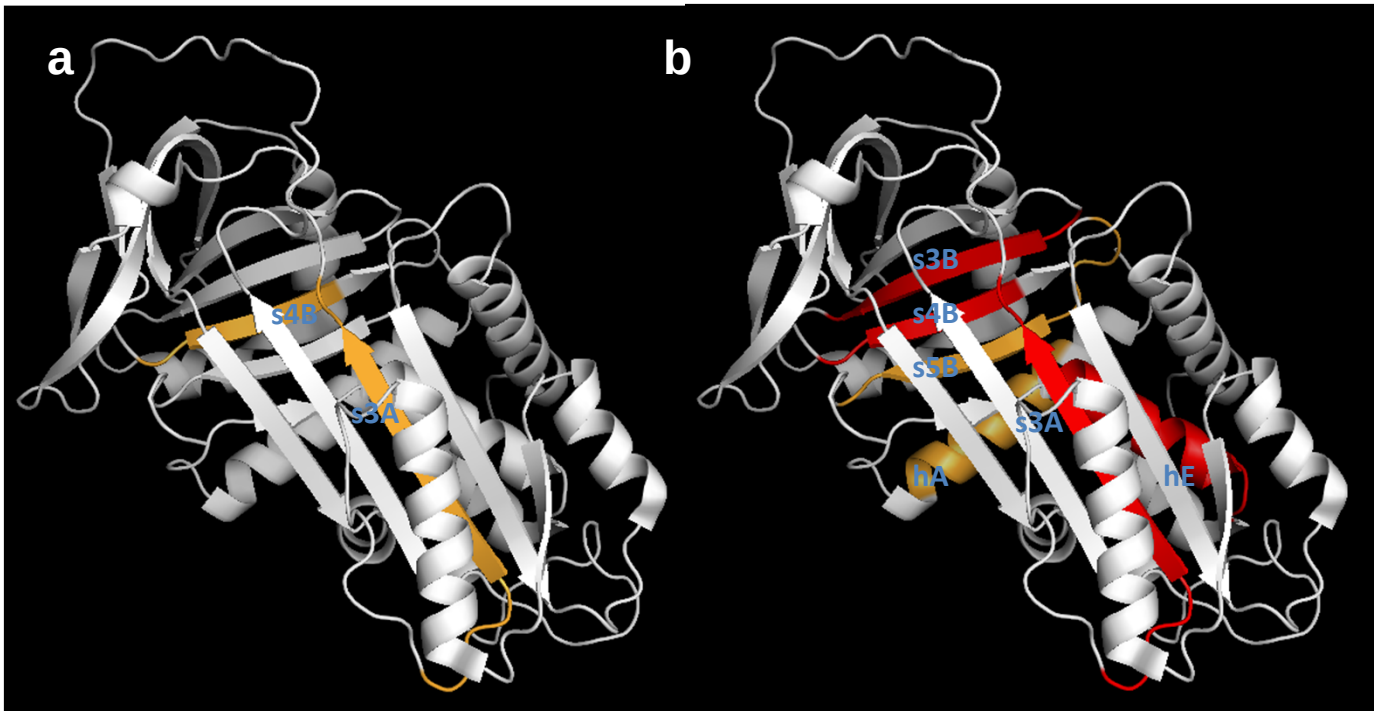
PLOS | COMPUTATIONAL BIOLOGY

β-Aggregation propensity of aligned serpins by Tango



Christine Stuart

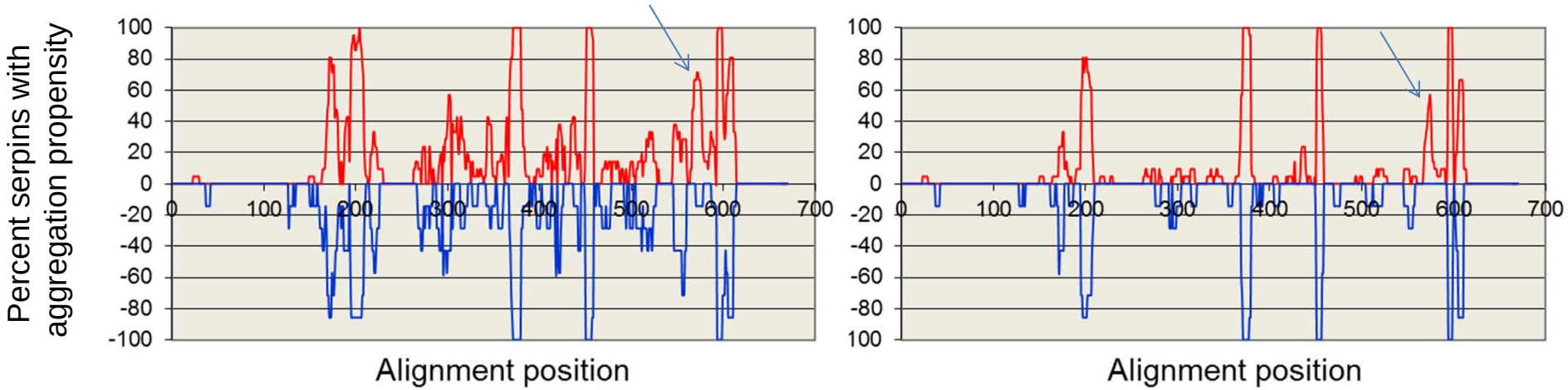
Serpins – APRs conservation



Christine Stuart

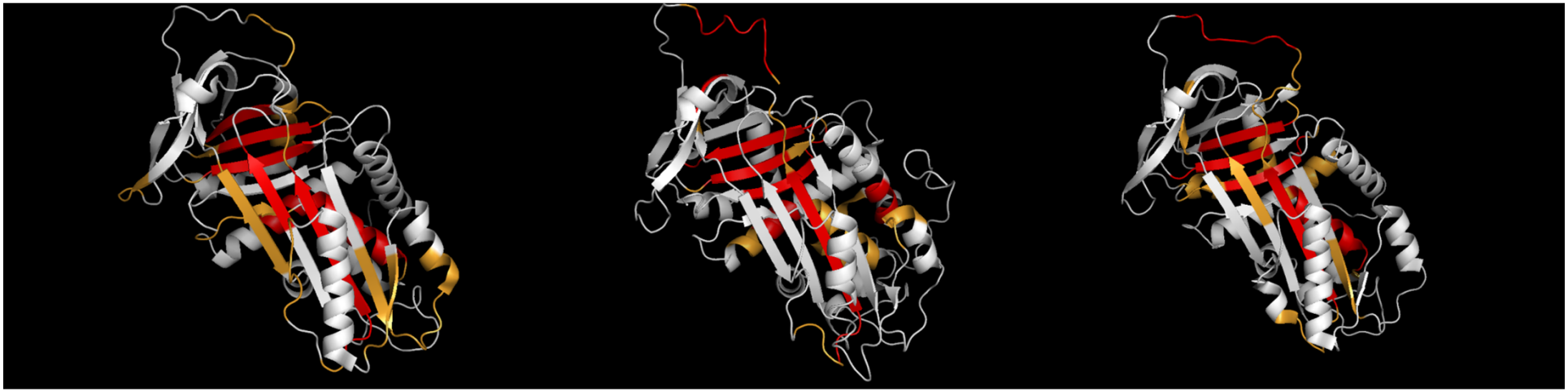
(a) 100% and (b) > 80% of 30 serpins studied
Yellow – all, Red – after threshold (Tango score > 5%)

APRs in inhibitory vs. non-inhibitory serpins



Christine Stuart

Inhibitory serpins – APRs conservation

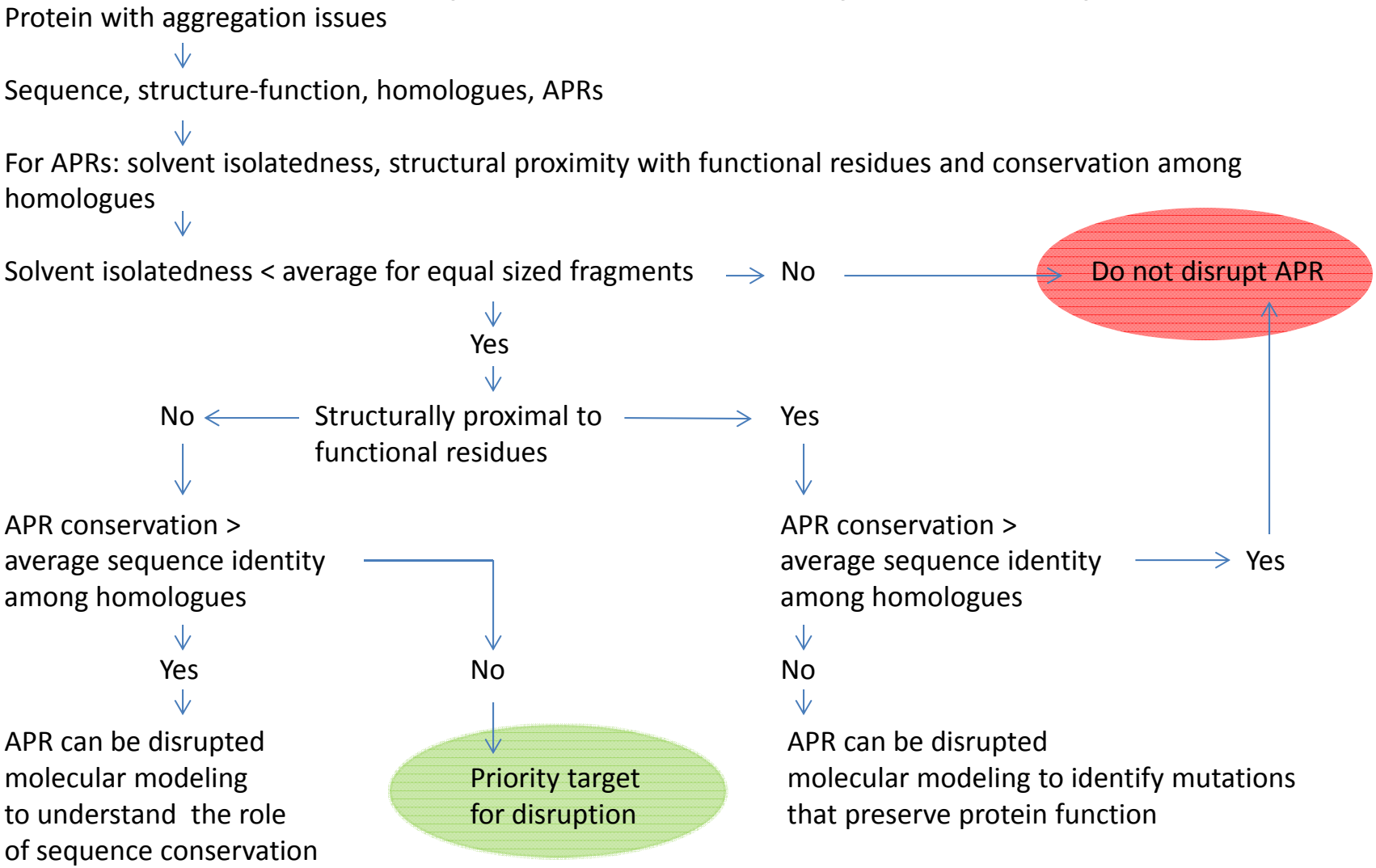


α_1-proteinase inhibitor Antithrombin Plasminogen activator inhibitor-1

Yellow – all, Red – after threshold (Tango score > 5%)

Christine Stuart

Strategy for identifying mutation sites to improve protein solubility via disruption of APRs



Adapted from Buck *et al*. (2013) On the Role of Aggregation Prone Regions in Protein Evolution, Stability, and Enzymatic Catalysis: Insights from Diverse Analyses. *PLoS Comput Biol* 9(10)

PLOS | COMPUTATIONAL BIOLOGY

Quantitation and characterization of protein aggregates

Characterization challenges of protein aggregates

- Heterogeneity
  - Size, morphology, secondary and tertiary structure, reversibility/dissociation, covalent modifications, surface chemistry protein and conditions specific
- Instability

Analytical tools - examples

- Aggregates detection, counting, characterization
  - Turbidimetry
  - Light scattering
  - Light obscuration
  - PAGE
  - SEC
  - AUC
  - FFF
  - Microscopy: TEM, AFM, flow microscopy
  - CD, FT-IR, Raman, UV
  - Fluorescence, intrinsic, extrinsic - dyes (thioflavin, ANS, Congo red, polythiophene acetic acid derivatives, PTAAs-amyloids)
  - Conformation-dependent antibodies (e.g., dot blot)

Emerging techniques for profiling protein aggregates

Technique	Size (µm)	Parameters
NMR	N/A	Diffusion coefficient; monomer loss
MS	N/A	Protein unfolding and aggregation
AFM	>0.001	Surface properties, particle size/morphology; unfolding
4D EM	>0.001	Phase transitions (excipient-protein compatibility)
DSLS	N/A	T_{agg}
Image analysis tools	0.05-300	Concentration, diffusion coefficient
NTA	20-1000	Particle size, concentration, polydispersity, morphology
RMM	0.1-50	Particle mass, viscosity, number of particles
Taylor dispersion analysis	0.01-5	Diffusion coefficient (hydrodynamic radius)
Polarization intensity differential scattering (PIDS)	0.01-100	Particle size
HPCE	10-22 kDa	Intrinsic fluorescence, aggregation/fragments, peak tracking
Raman-based methods	N/A	Peptide environment, chemical identity

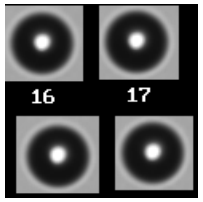
Adapted from Hamrang *et al*. (2013) Proteins behaving badly: emerging technologies in profiling biopharmaceutical aggregation. *Trends Biotech* 31, 448-58

Continues improvements in aggregates metrology and characterization

- New regulatory standards
  - FDA immunogenicity draft guidance, USP chapter <787> and <1787>
- New physical standards, methods, instruments, algorithms
  - NIST EFTE standard
  - DLS resolution
  - PTA repeatability
  - Visible particles monitoring
  - Combining different principles of measurement
    - e.g., microscopy & spectroscopy (IR, Raman)
- Data analysis (visualization)

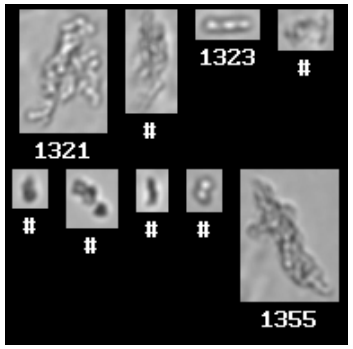
Particle Standards

- Standards → Nice!



  - Homogenous
  - Spherical
  - High optical contrast
  - Instrument calibration (Nanosight, FlowCAM, Archimedes, ect)
    - "Simple NIST-traceable calibration gives true accuracy at the 1% level" (http://affinitybio.com/ , Archimedes)
    - "Accuracy: Better than +/-2% on NIST traceable latex standards" –(http://www.malvern.com/, Zetasizer)
    - ",,the diameter of a well prepared, monodisperse, calibration polystyrene micro-sphere ,,within 1-2 percent of nominal"- (http://www.nanosight.com/ , Nanosight)

- Aggregates → ????



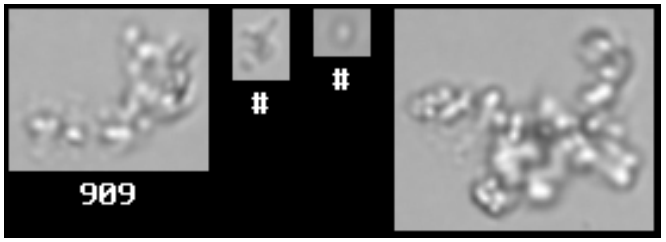
  - Heterogeneous
  - Low aspect ratios
  - Low optical contrast

How well can standards facilitate cross technique comparison?
Can they be more representative of protein aggregates?

Curtesy Joey Kotarek

Particle Standards

- EFTE standards- NIST
  - (tetrafluoroethlyene/ethylene copolymer)
  - Refractive index ~1.4
  - Stable, durable,
  - morphologically closer to protein aggregates

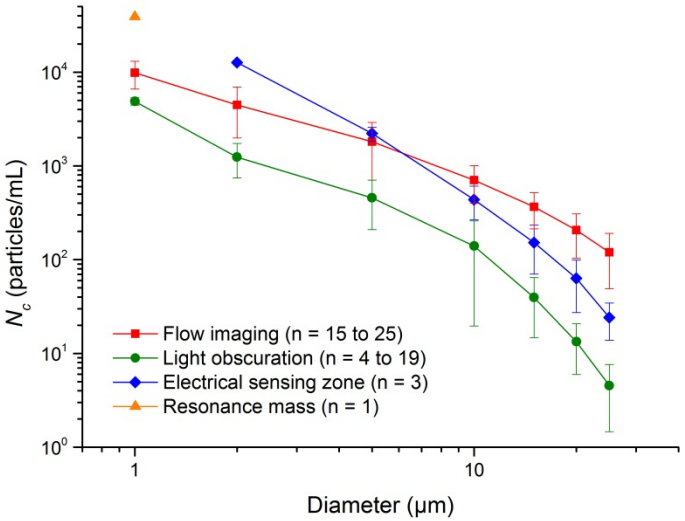


Courtesy Joey Kotarek

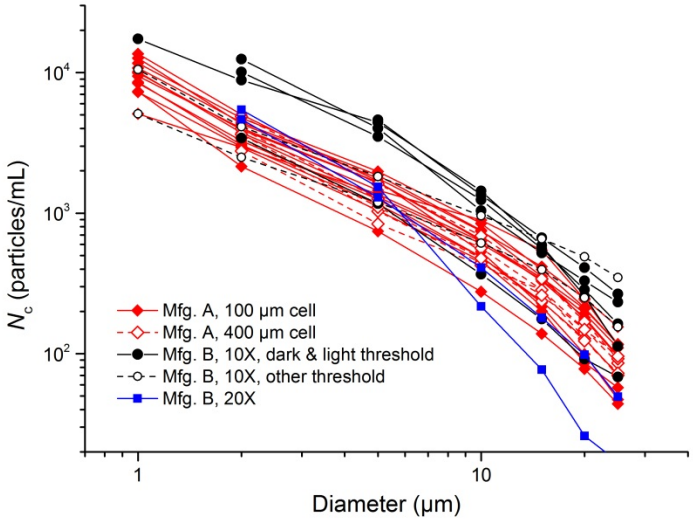
Method Comparison

How well can instruments measure a more realistic standard?

Round robin, multi-lab study



Different techniques
Different sensitivity

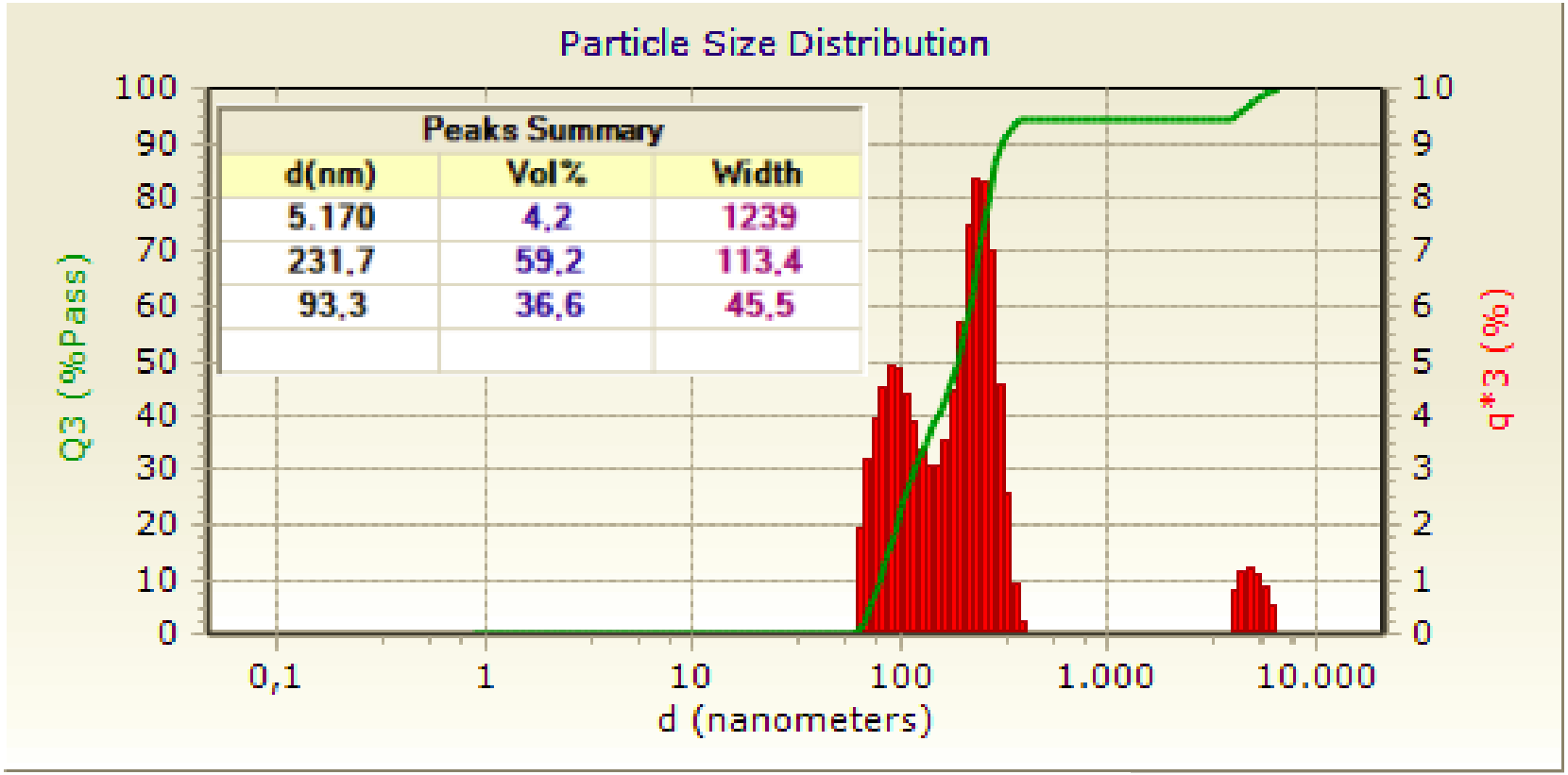


Different instruments (all flow imaging!)
Different sensitivity

`

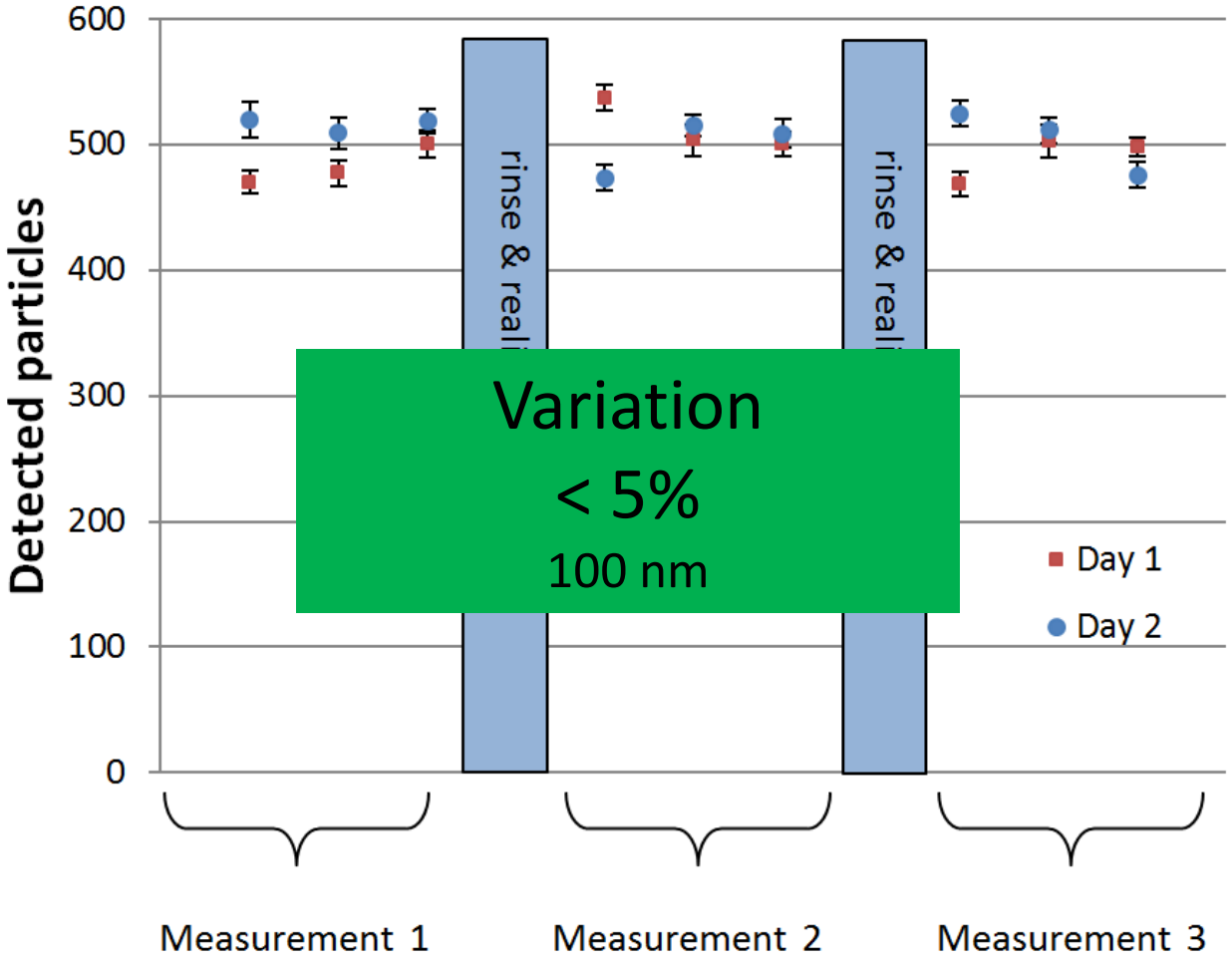
Courtesy Dean Ripple, NIST, and Joey Kotarek

180° DLS size
100 nm & 200 nm PS Mix



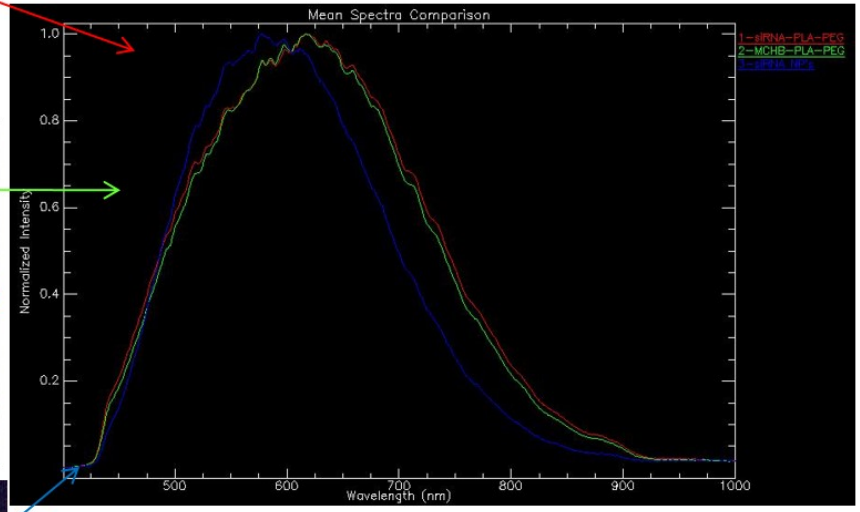
From: Dr. Hanno Wachernig, Particle Metrix

Particle Tracking Analysis: Intra day and inter day study



From: Dr. Hanno Wachernig, Particle Metrix

Combined techniques
Hyperspectral microscopy and IR



Mean normalized spectra from all three samples
Note: spectra from samples #1 and #2 are very similar with sample #3 displaying significant and quantitative spectral differences.

From: Byron Cheatham, Cytoviva

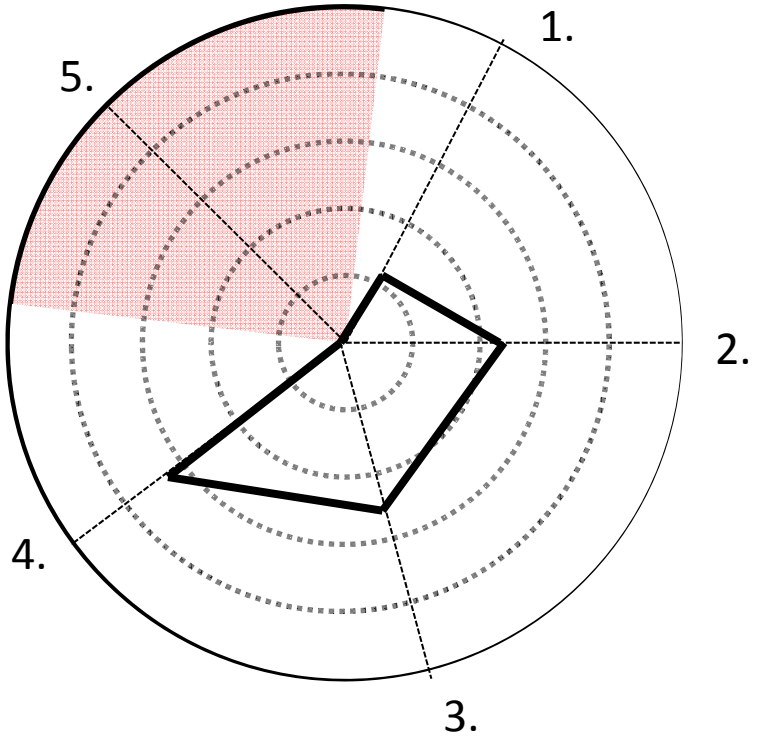
Automated visible particle counting

Run Number	Glass Beads in 65% Glycerol	Polystyrene Spheres in Water
1	51	9
2	54	8
3	52	8
4	59	9
Manual Count	50-60 particles	8-9 particles

From Freeslate Application Note “High Throughput Visible Particle, Turbidity, and Color Measurement of Biologics Formulations”

Data visualization

- Radar chart array analysis



MFI:
Axes represent size bins
Value: number (aspect ratio)
Polygon perimeter: concentration
Pink: insufficient data

Adapted from Kalonia *et al*. (2013) Radar Chart Array Analysis to Visualize Effects of Formulation Variables on IgG1 Particle Formation as Measured by Multiple Analytical Techniques. *J Pharm Sci* 102, 4256-67

Final thoughts

- The ultimate goal – product safety and efficacy
  - Protein aggregates content should be minimized
  - Protein aggregates should be measured and characterized
  - Relevant criteria for product testing should be established
- The task is difficult due to heterogeneity and instability of protein aggregates and the lack of good safety prediction models
- Nevertheless, the progress in the field is remarkable

Acknowledgements

Dorothy Scott

Christine Stuart

Joey Kotarek