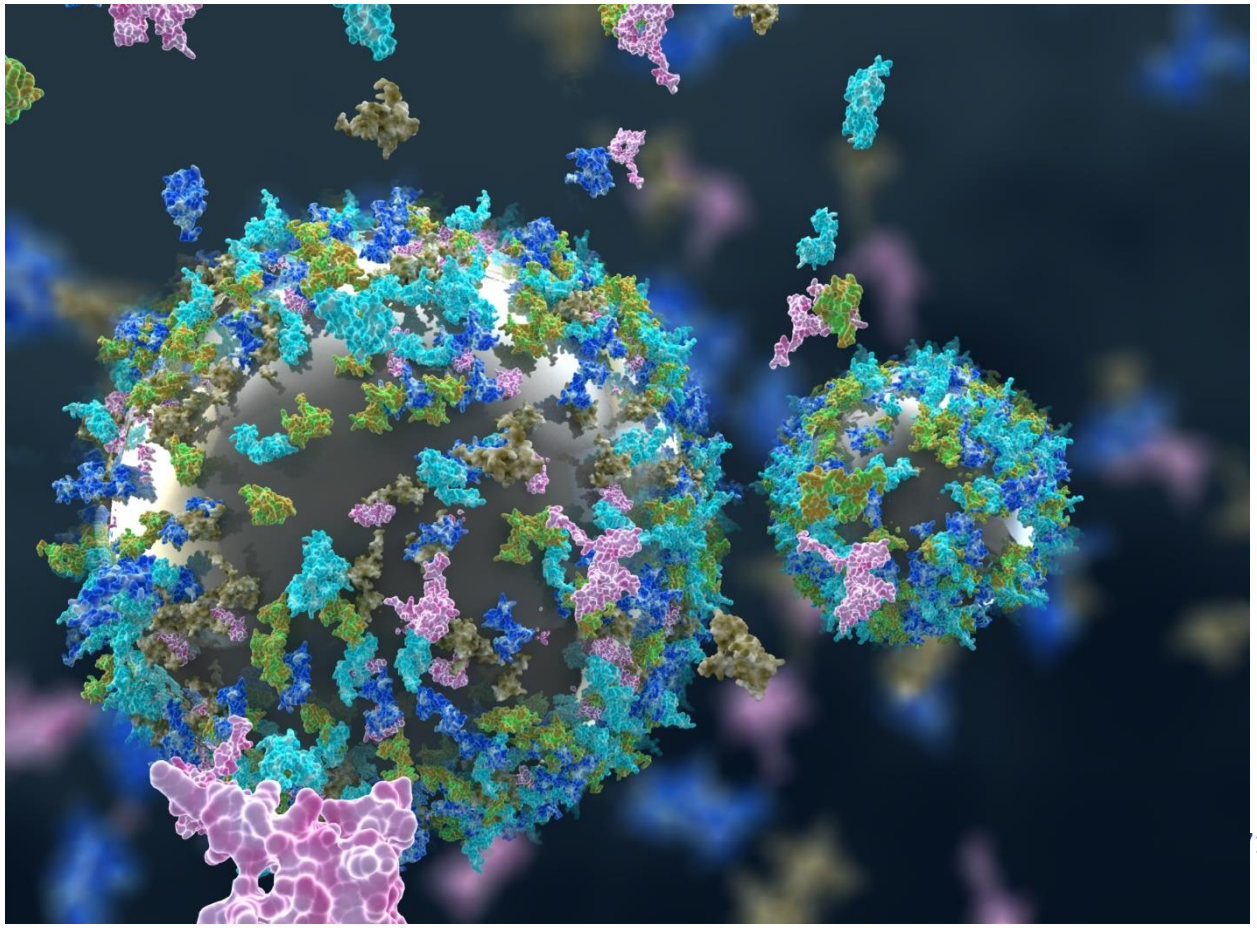
Transitioning to predictive analysis for nanoparticle biocorona studies



1.

Santa Clara University

Korin E. Wheeler, PhD

Associate Professor, Dept. of Chemistry & Biochemistry

Protein or Bio- Corona

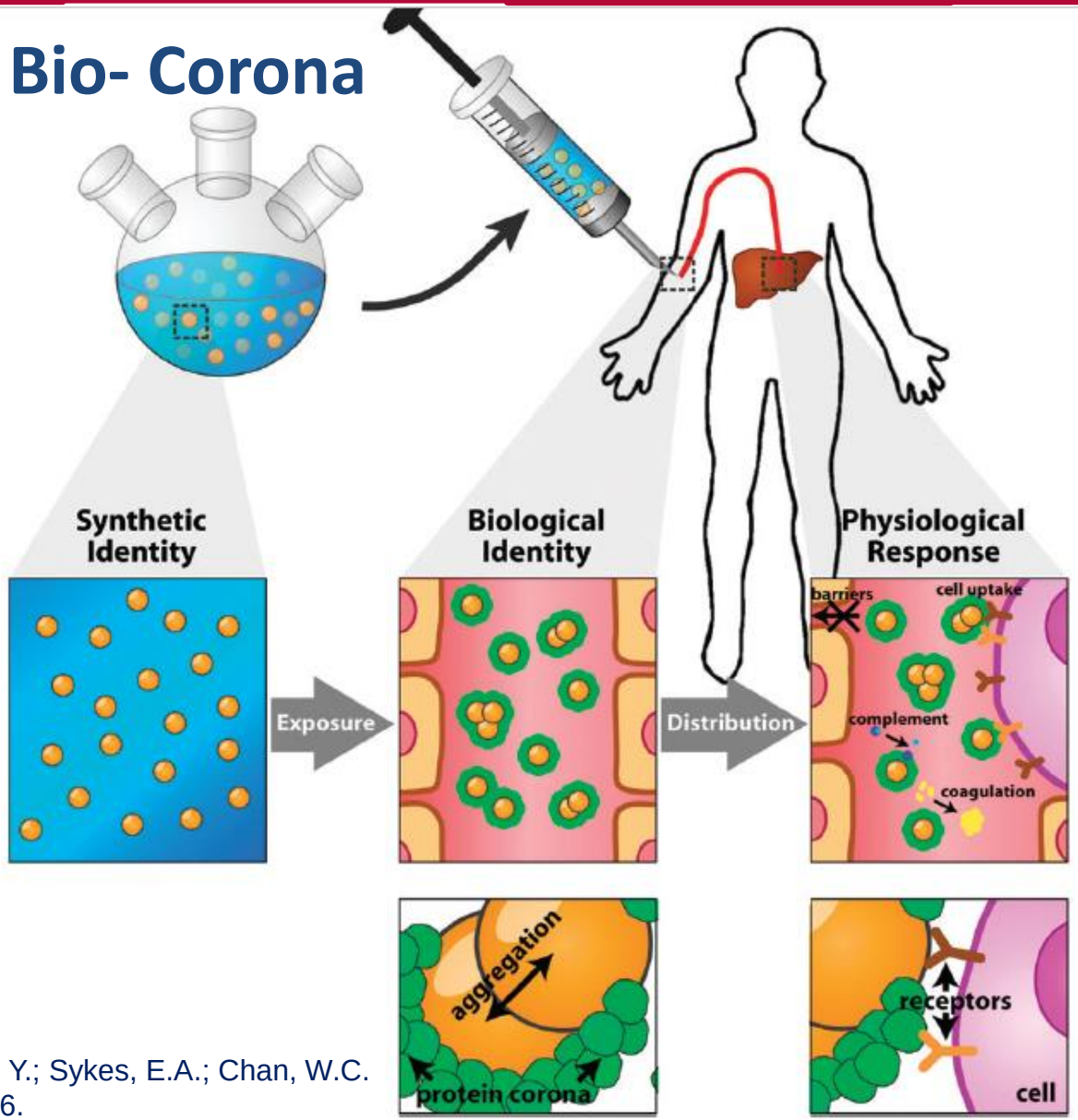


Image: Lazarovits, J.; Yang, Y.; Sykes, E.A.; Chan, W.C. *ChemComm*, 2015, 51, 2756.

State of the field

Reporting Standards

Predictive Analysis

Implications & Next Steps

Particle properties

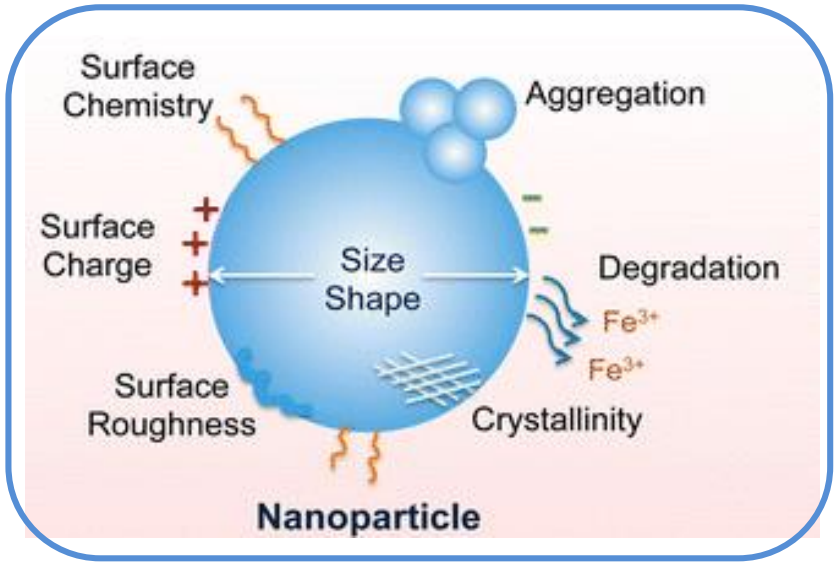
- Core composition
- Size
- Surface properties
- Aggregation state

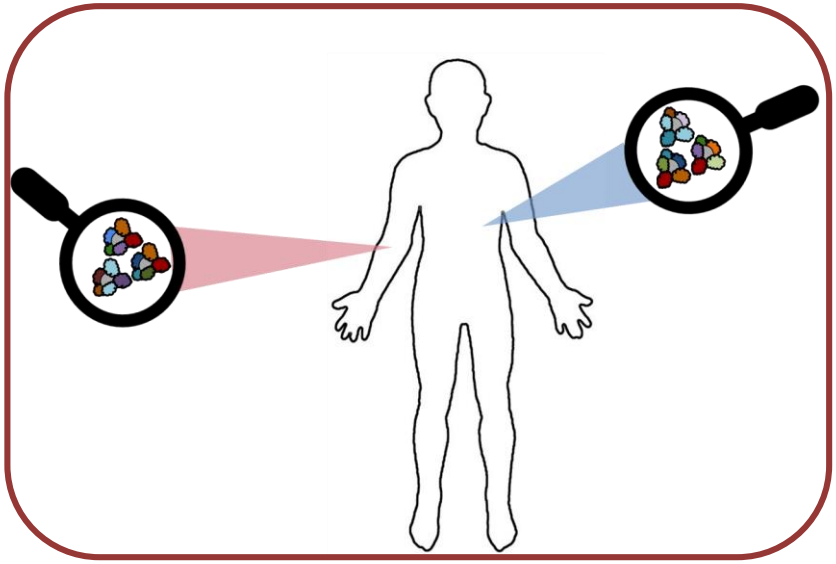
Solution conditions

- pH
- Temperature
- Dissolved solutes

Population of biological molecules

- Protein concentrations
- Organism / organ / organelle
- Protein biophysical properties
- Protein-protein interactions
- Protein PTMs





DOI: 10.1021/cr400295a

https://doi.org/10.1371/journal.pone.0175871

Particle properties

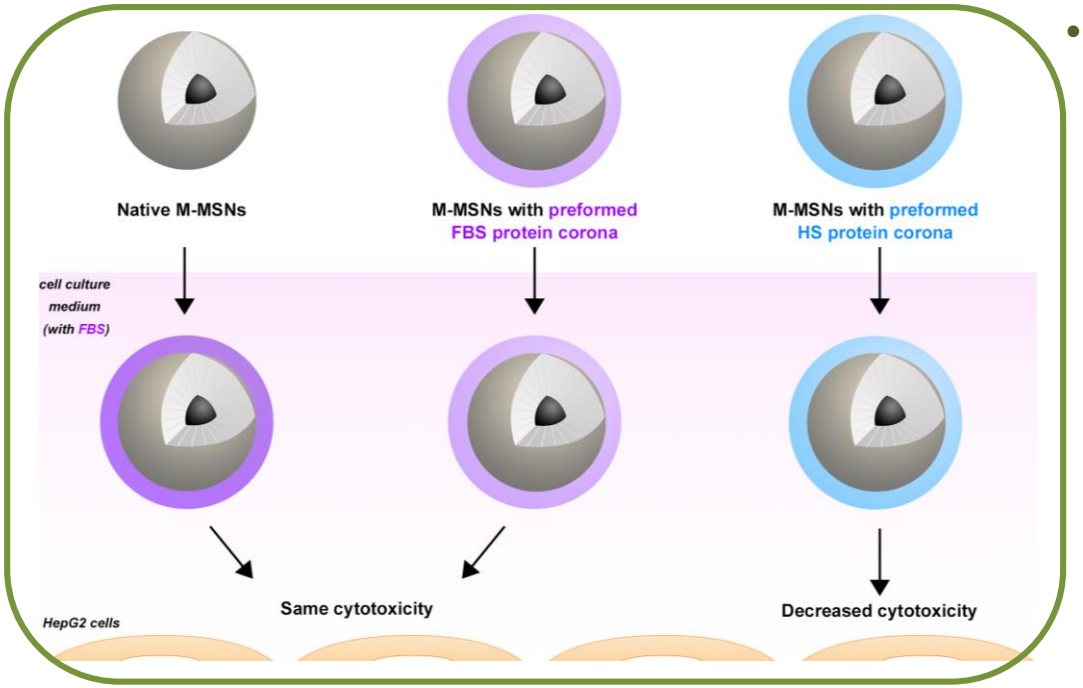
- Core composition
- Size
- Surface properties
- Aggregation state

Solution conditions

- pH
- Temperature
- Dissolved solutes

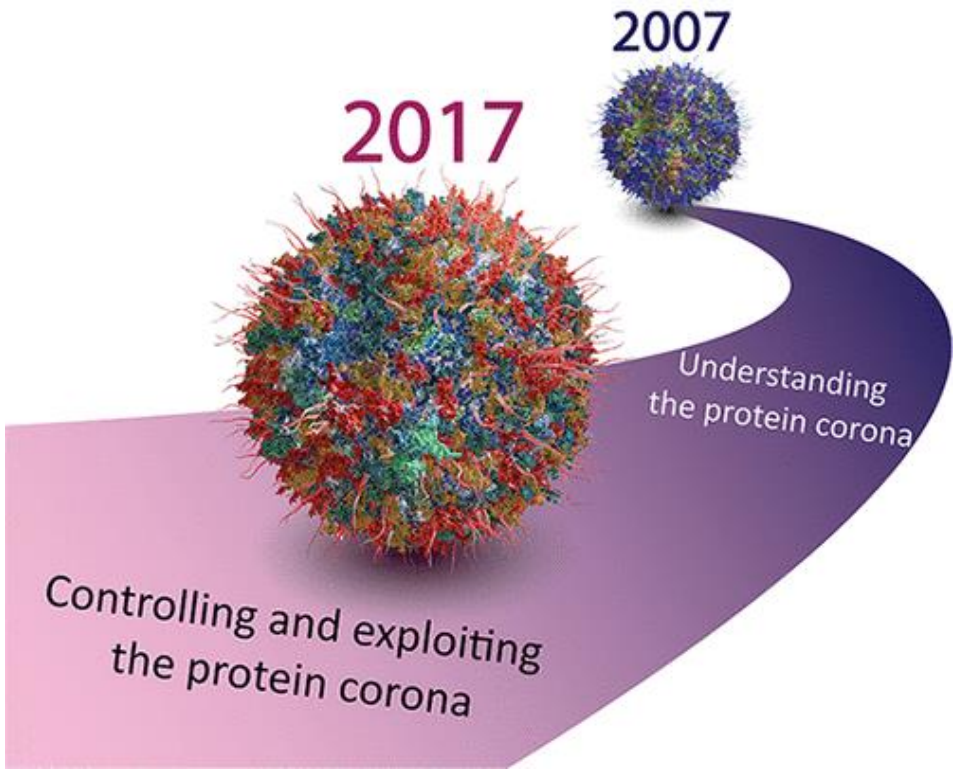
Population of biological molecules

- Protein concentrations
- Organism / organ / organelle
- Protein biophysical properties
- Protein-protein interactions
- Protein PTMs



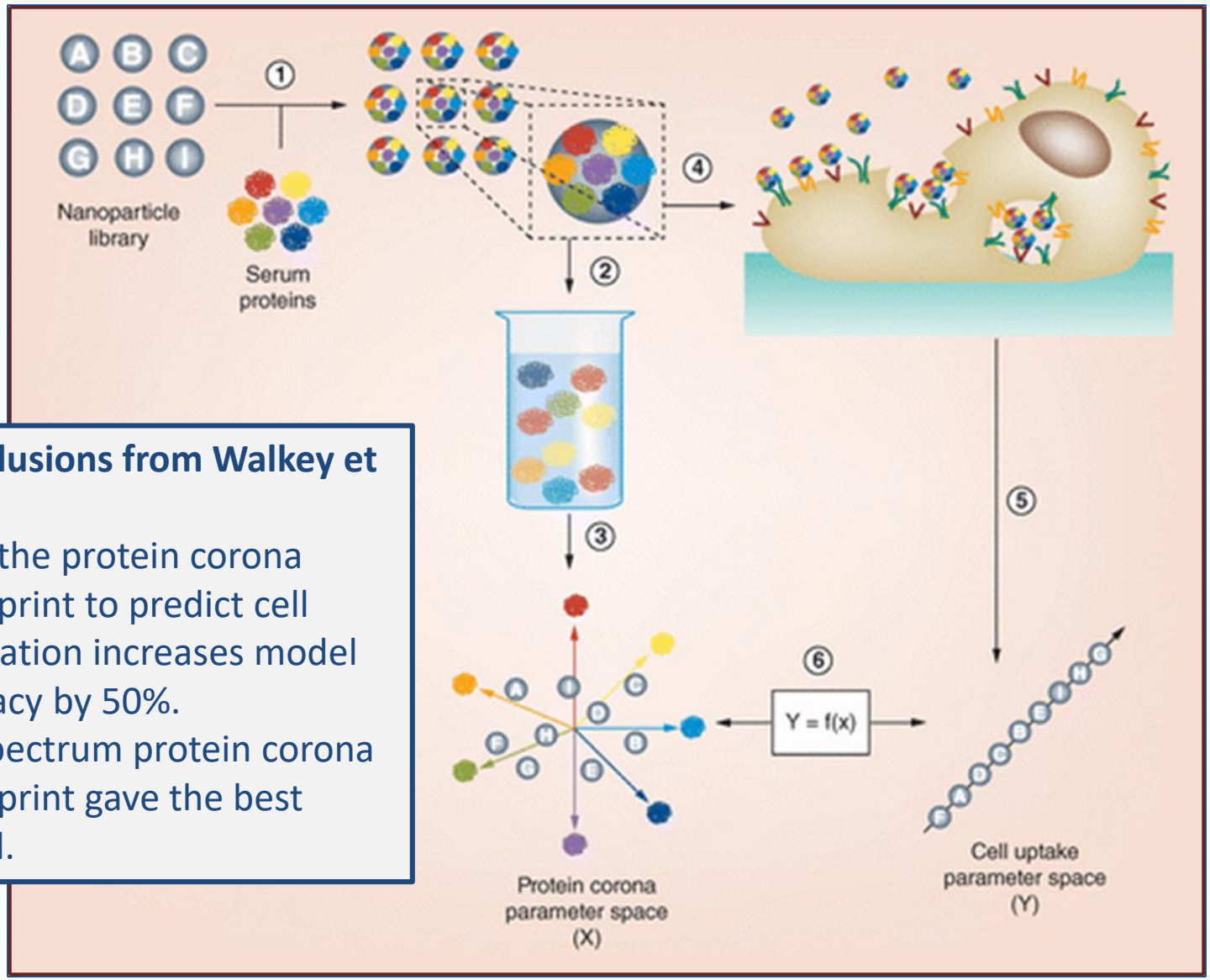
https://doi.org/10.1371/journal.pone.0182906

A decade of the protein corona



My Hypothesis:
If we really understand the protein corona (PC), then we should be able to predict its composition.

ACS Nano, **2017**, *11* (12), pp 11773–11776
DOI: 10.1021/acsnano.7b08008



Key conclusions from Walkey et al:

- using the protein corona fingerprint to predict cell association increases model accuracy by 50%.
- Full spectrum protein corona fingerprint gave the best model.

Image: Walkey, C.D. et al *ACS Nano* 2014, 8(3) 2439-2455

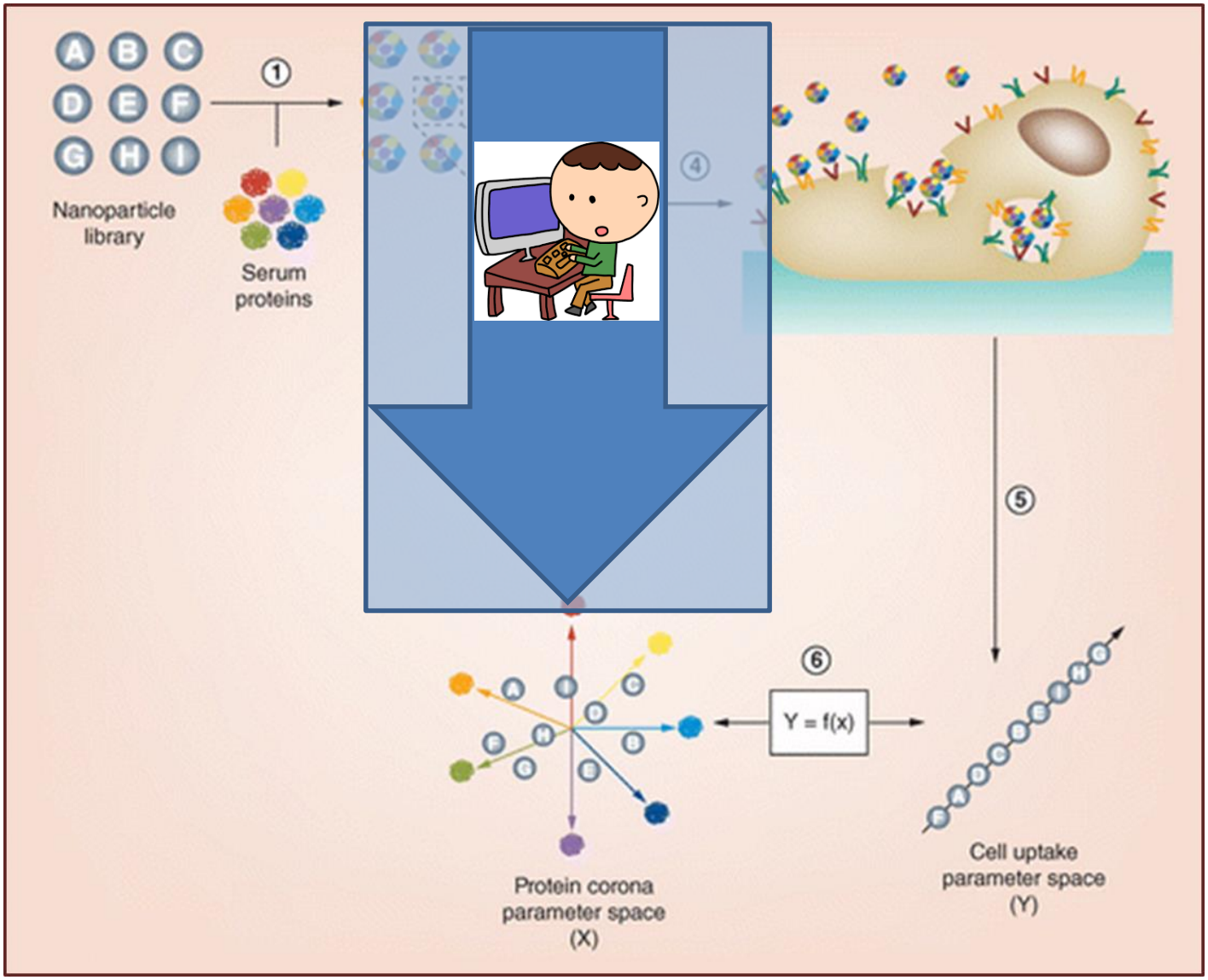
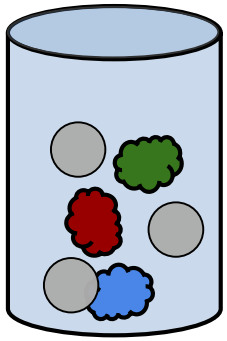


Image: Walkey, C.D. et al *ACS Nano* 2014, 8(3) 2439-2455

1

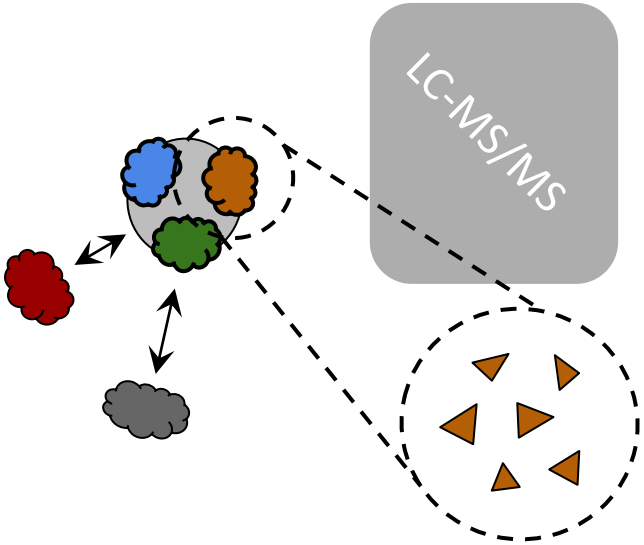


Nanoparticle properties

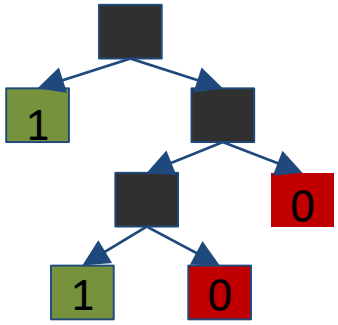
Protein properties

Solvent properties

2



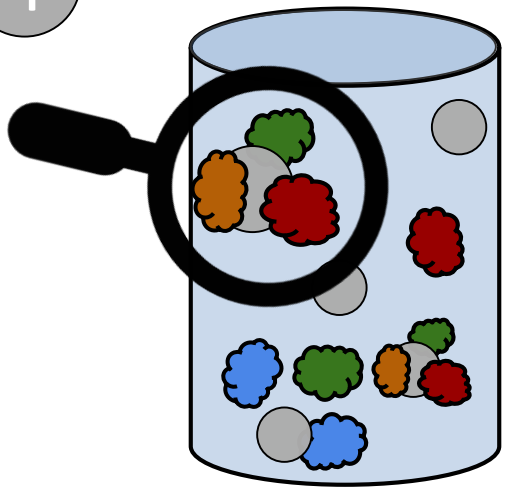
3



Majority Voting

$$P(Bound_{p,prot}) = \frac{1}{N}\sum_{j=0}^{N} decision_j$$

4



Finlay, et al *ES:Nano,* **2018**, 5 (1), 64-71.

Particle properties

- Core composition
- Size
- Surface properties
- Aggregation state

Solution conditions

- pH
- Temperature
- Dissolved solutes

Population of biological molecules

- Protein concentrations
- Organism / organ / organelle
- Protein biophysical properties
- Protein-protein interactions
- Protein PTMs

Our test conditions

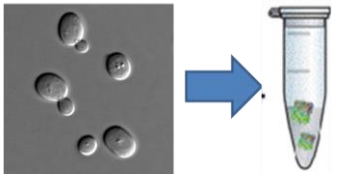
Nanomaterial: silver

- **10 nm citrate coated (-)**
- **10 nm branched poly(ethylene imine) coated (BPEI, +)**
- **100 nm citrate coated (-)**

Solute conditions

- **0.1 mM Cys:** mimics extracellular total thiol.
- **0.8 mM NaCl:** mimics surface fresh water salinity
- **3.0 mM NaCl:** models mitochondrial salinity

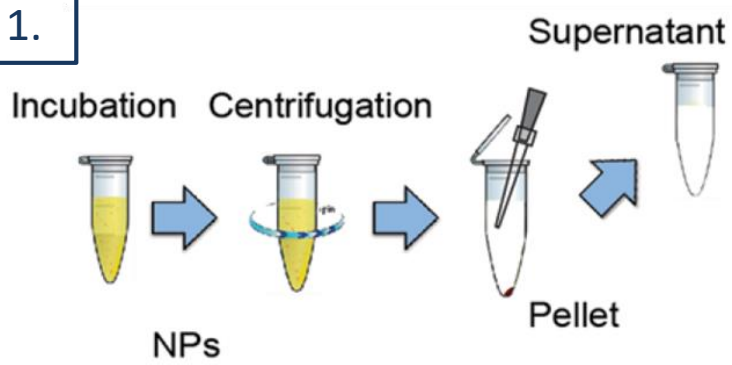
proteins: yeast *Saccharomyces cerevisiae*

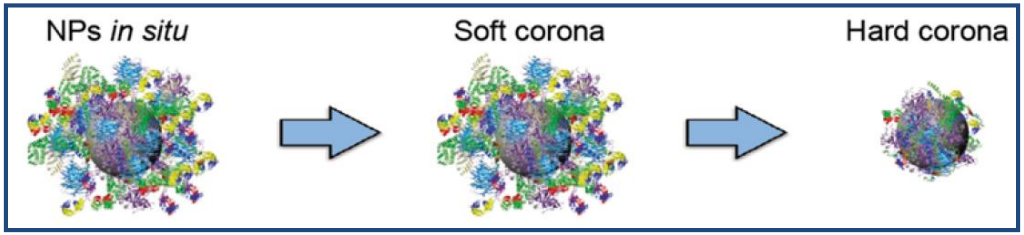


model protein mixture: Yeast cell lysate or yeast protein extract (YPE)

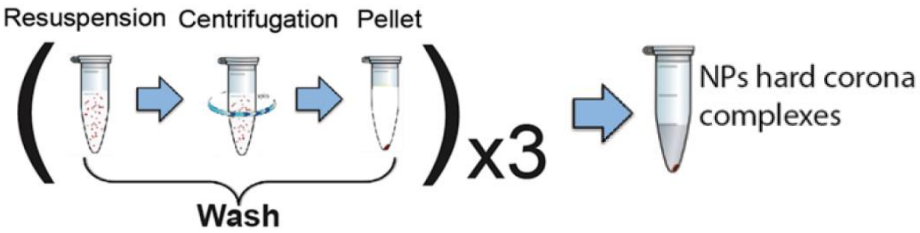
Eigenheer, et al *ES:Nano,* **2014**, 1 (3), 238-247.

1.





2.

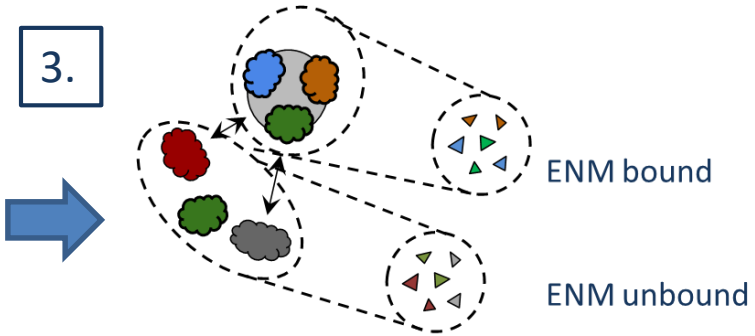


Unique to our approach:

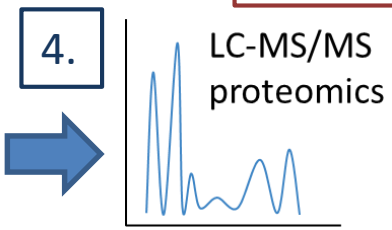
- Directly digest proteins off of the particles.
- Characterize bound proteins AND unbound proteins.
- Define protein enrichment=

$$\frac{\text{[Protein NSAF in bound fraction]}}{\text{[Protein NSAF in unbound fraction]}}$$

3.

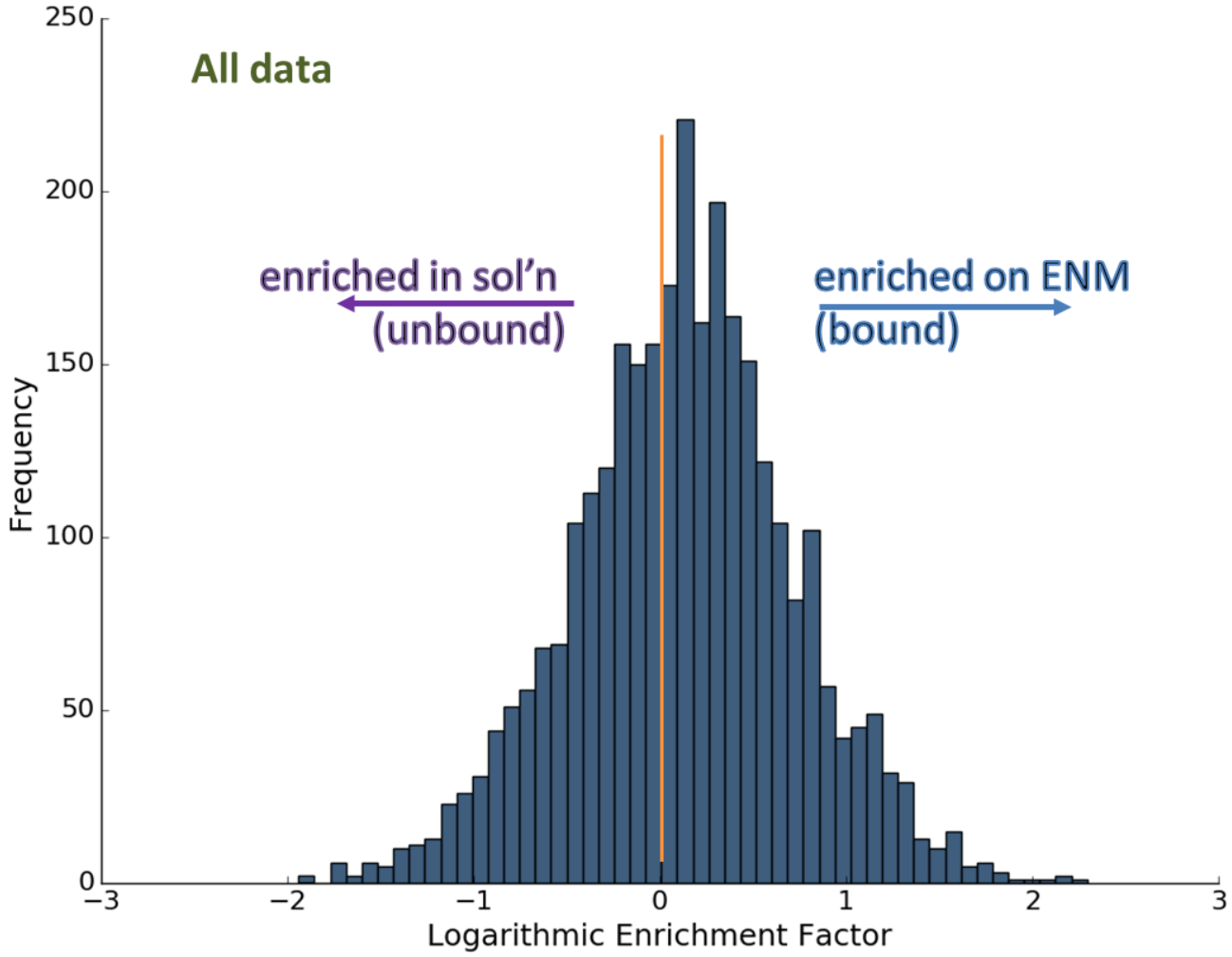


4.



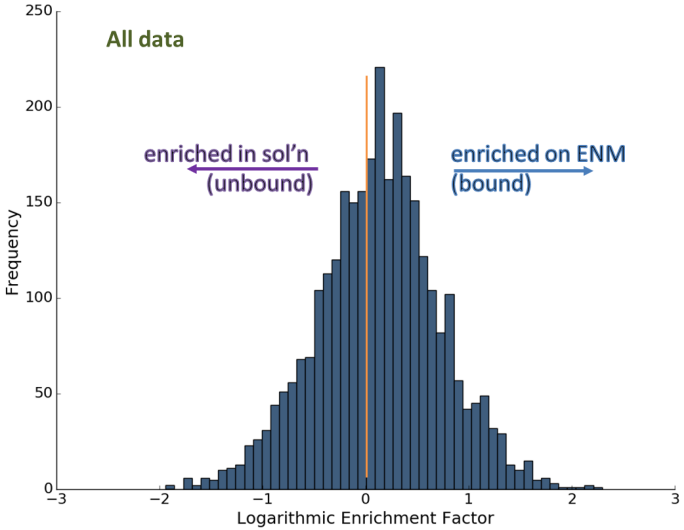
images: P. Bergese and K. Hamad-Schifferli (eds.), *Nanomaterial Interfaces in Biology: Methods and Protocols, Methods in Molecular Biology*, vol. 1025, DOI 10.1007-978-1-62703-462-3 11.

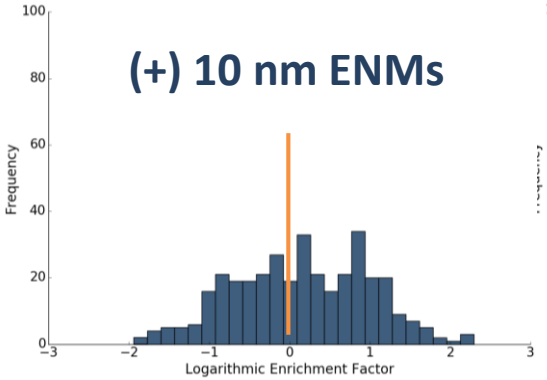
Unique approach: defining an enrichment factor

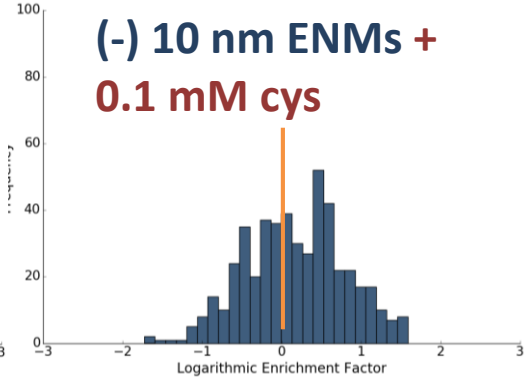


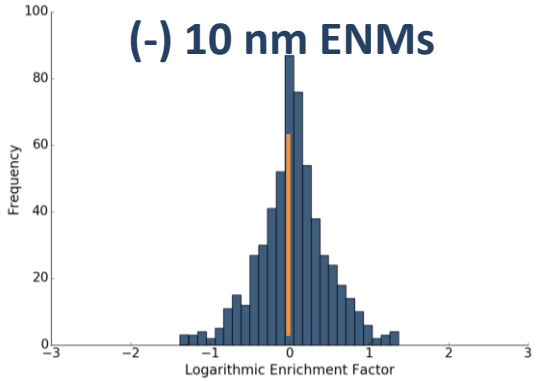
Finlay, et al *ES:Nano,* **2018**, 5 (1), 64-71.

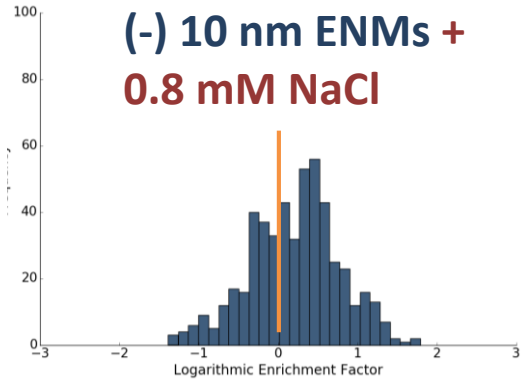
Variability of enrichment factors

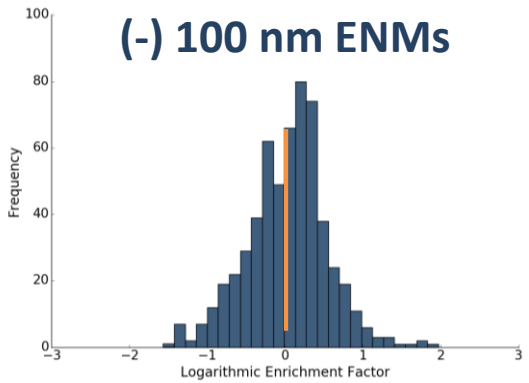


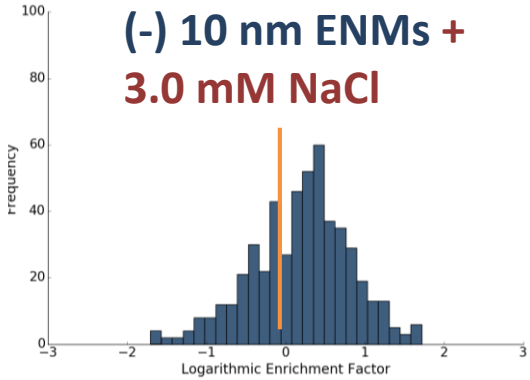








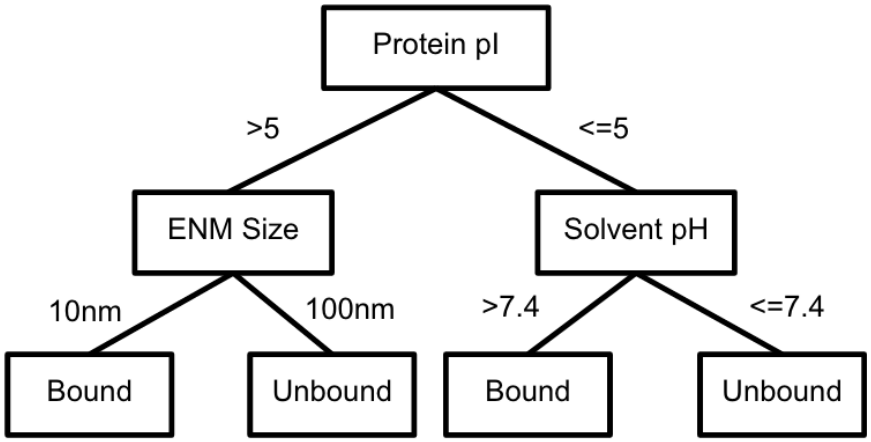




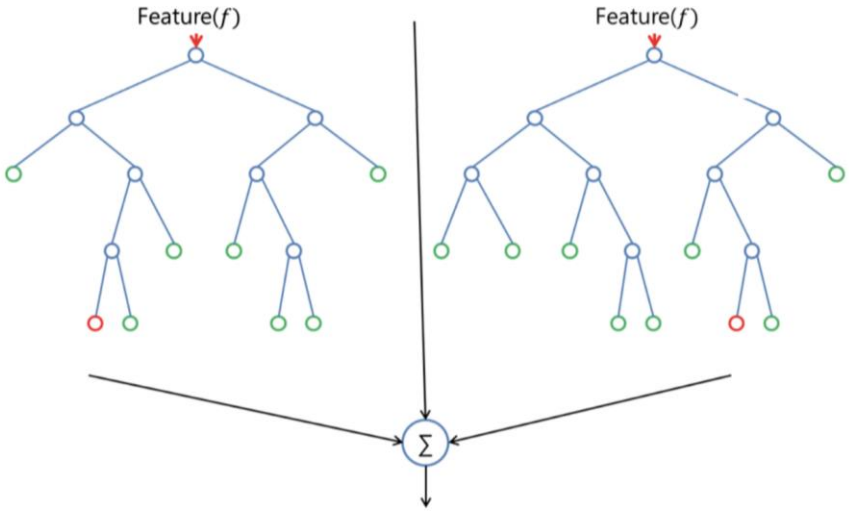
Findlay, M.R. *et al* ES: Nano 2018,5, 64-71.

Random forest classification
builds multiple decision trees and merges them together to get a more accurate and stable prediction.

Example decision tree



merged trees give prediction



Findlay, M.R. *et al* ES: Nano 2018,5, 64-71.

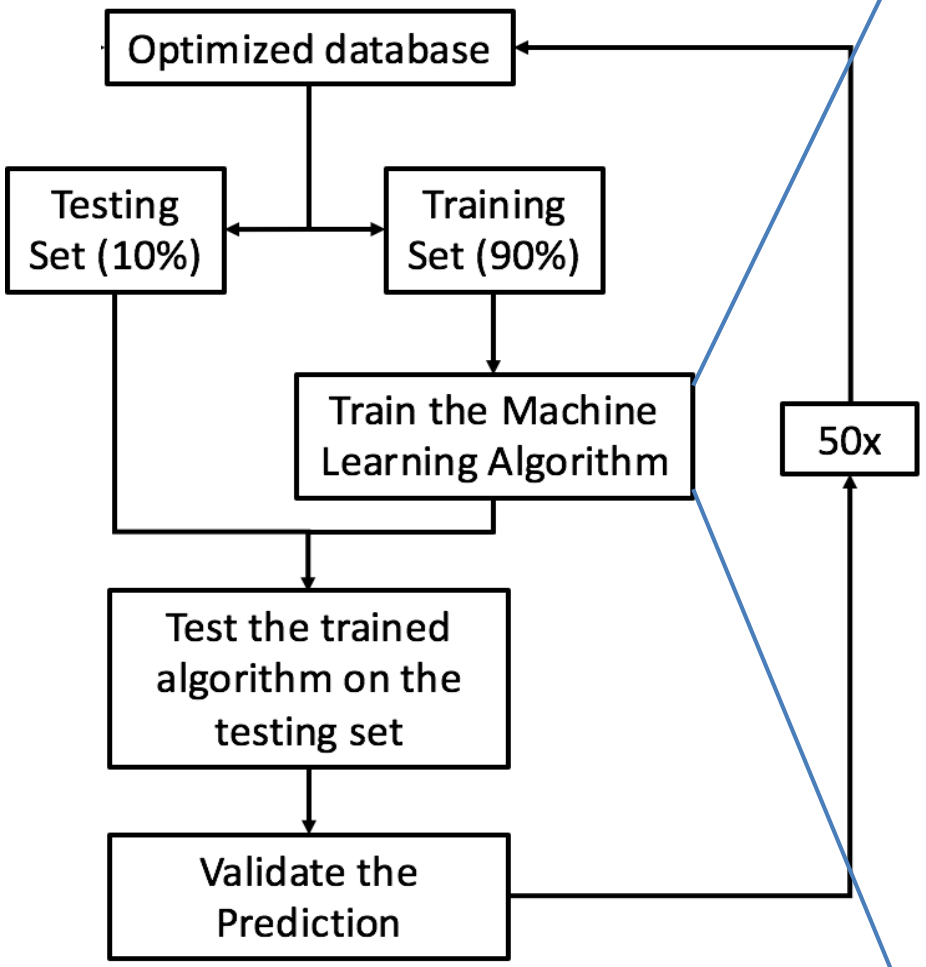

Protein pI, MW ENM size, coating solute ID, [solute]

SPE in 10 mM NaPi +

1. (−) 10 nm ENMs
2. (+) 10 nm ENMs
3. (-) 100 nm ENMs
4. (−) 10 nm ENMs
   + 0.1 mM Cys
5. (−) 10 nm ENMs
   + 0.8 mM NaCl
6. (−) 10 nm ENMs
   + 3.0 mM NaCl

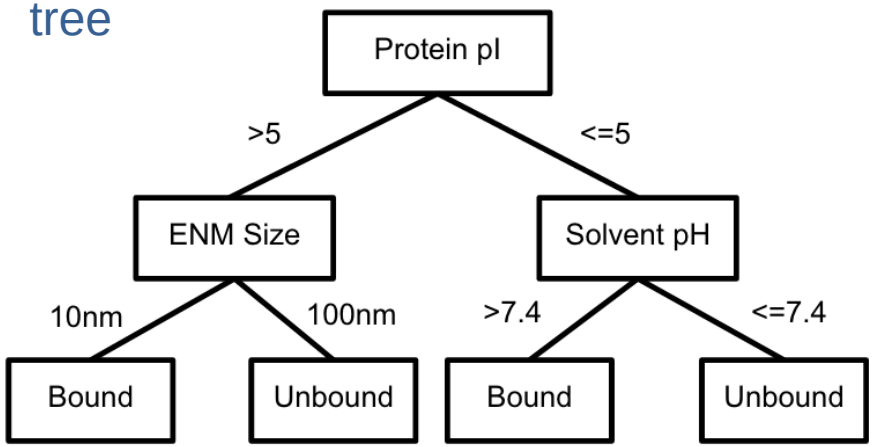
Observations	feature 1	feature 2	…	feature m	Classification Target
Experiment 1	Value or Class	Value or Class		Value or Class	Bound/Unbound
Experiment 2					
.					
.					
.					
Experiment n					

Finlay, et al *ES:Nano,* **2018**, 5 (1), 64-71.



Random Forest Classification

- m=log(n) experiments randomly selected with replacement from database with size n
- m experiments used to grow a decision tree

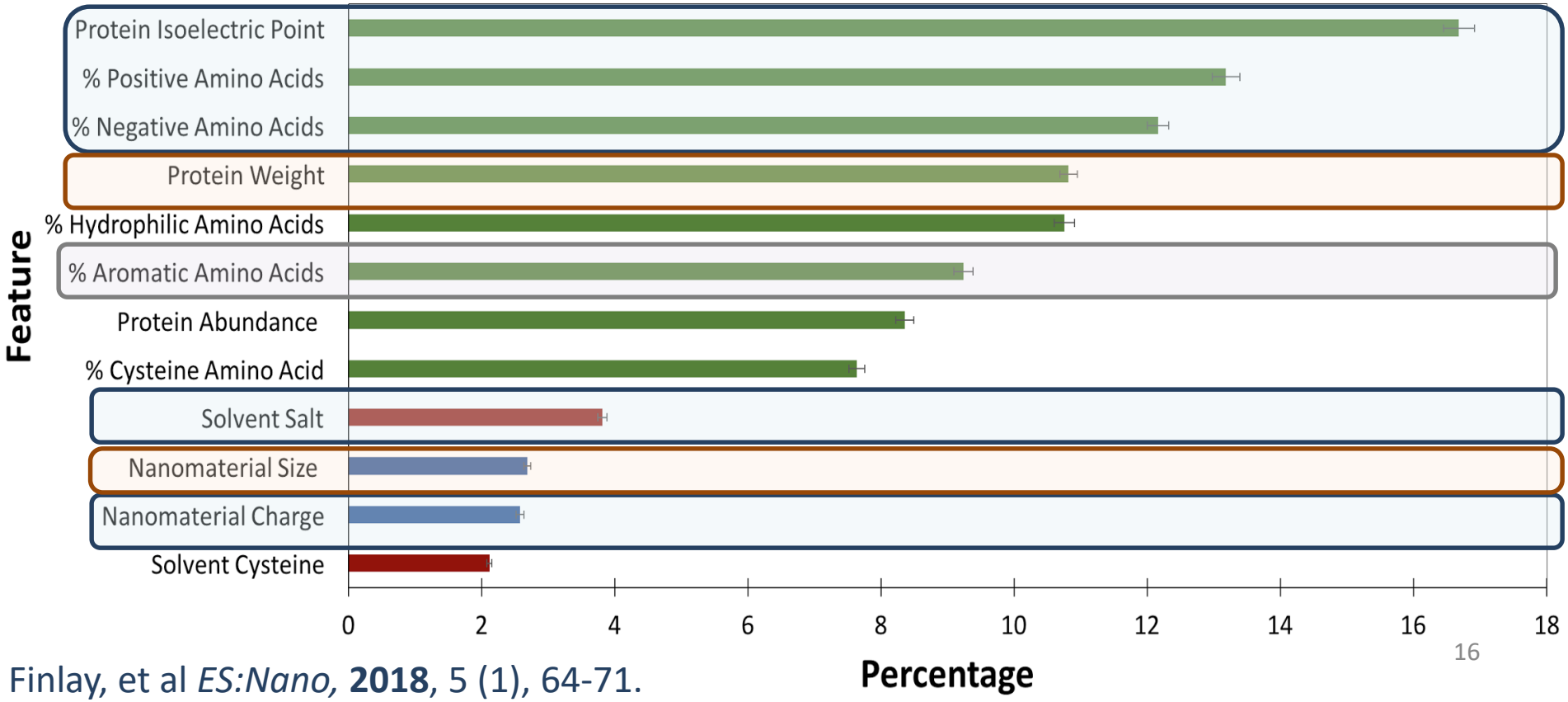


- Repeat 2500 times
- Voting between trees to decide if protein is within the ENM corona
- Probability of ENM protein corona assignment outputted

Finlay, et al *ES:Nano,* **2018**, 5 (1), 64-71.

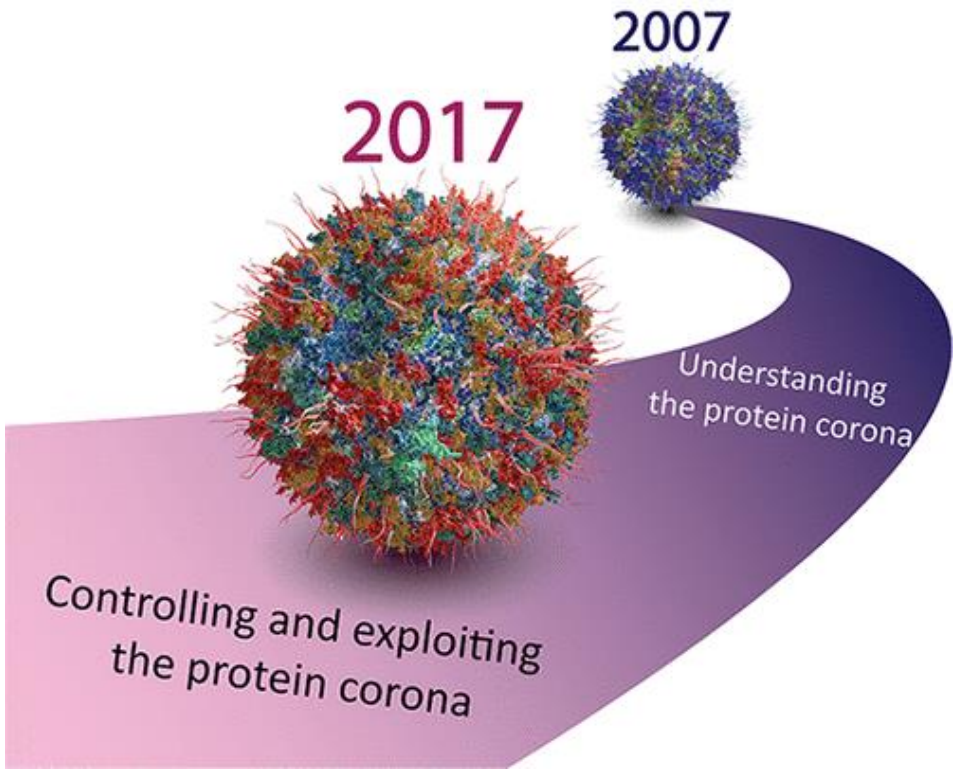
Our Random Forest Classification analysis gives:

- Accuracy 76 %
- F1-score 0.81

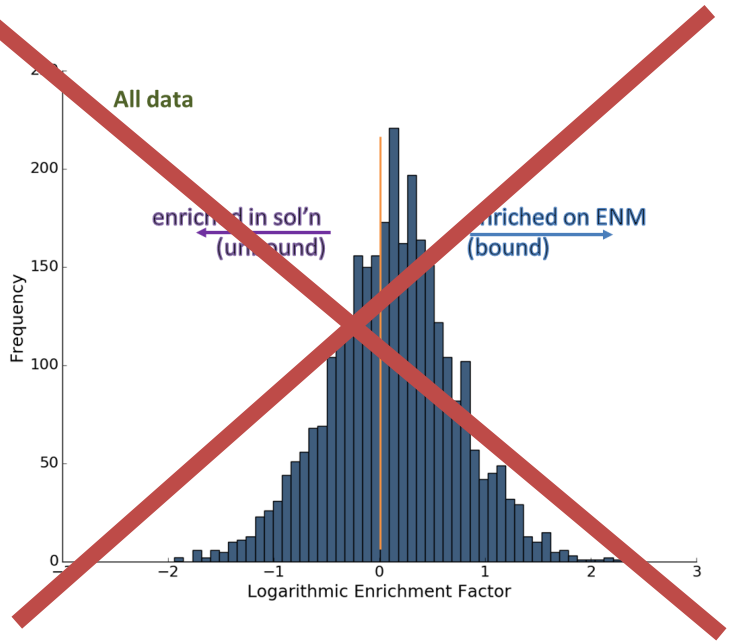


Finlay, et al *ES:Nano,* **2018**, 5 (1), 64-71.

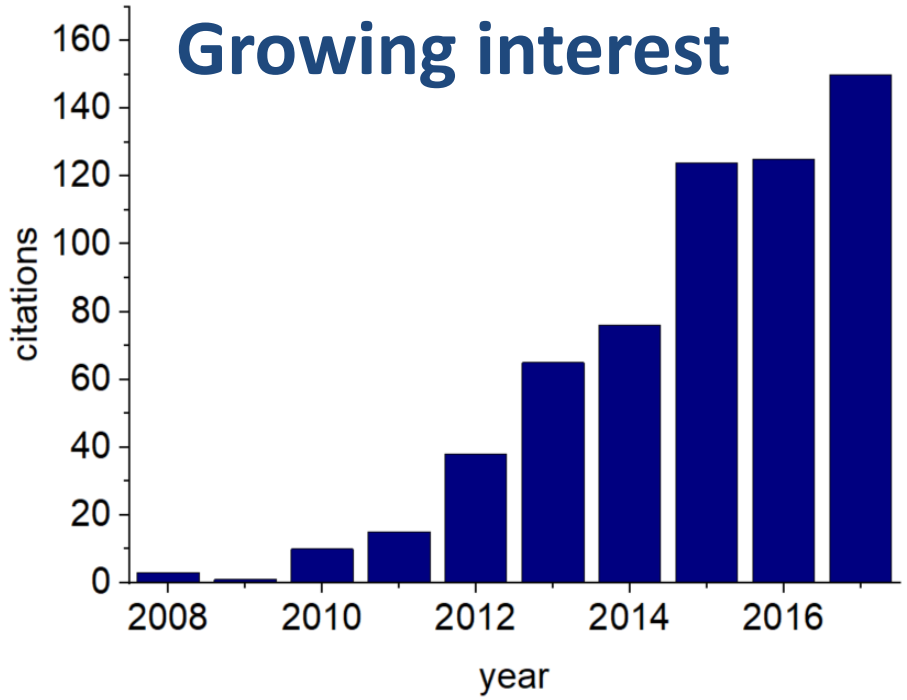
A Decade of the protein corona



ACS Nano, **2017**, *11* (12), pp 11773–11776
DOI: 10.1021/acsnano.7b08008



No. To model PC formation, information on 'unbound' proteins is necessary.



PubMed search for “Protein Corona”, Sept 1, 2018.

Chem Soc Rev

REVIEW ARTICLE

View Article Online

The nanoparticle biomolecule corona: lessons learned – challenge accepted?

Environmental Toxicology and Chemistry, Vol. 36, No. 12, pp. 3181–3193,

Printed in the

Critical Reviews

-NANO INTERFACE AND ENVIRONMENT: A CRITICAL REVIEW

LIDO-REYES,[a,b,*] FRANCISCO LEGANES,[a] FRANCISCA FERNÁNDEZ-PIÑAS,[a] and ROBERTO ROSAL[b]

[a]Departamento de Biología, Facultad de Ciencias, Universidad Autónoma de Madrid, Madrid, Spain

oxicology

, Volume 89, Issue 4, pp 519–539 | Cite as

rds understanding of nanoparticle–protein corona

Authors and affiliations

n Tian, Yuliang Zhao, Chunying Chen, Ruhong Zhou, Zhifang Chai

ACSNANO

Cite This: *ACS Nano* 2017, 11, 11773–11776

www.acsnano.org

A Decade of the Protein Corona

Pu Chun Ke,[†] Sijie Lin,[‡] Wolfgang J. Parak,[§] Thomas P. Davis,[*,†,⊥] and Frank Caruso[*,∥]

[†]ARC Centre of Excellence in Convergent Bio-Nano Science and Technology, Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Parade, Parkville, Victoria 3052, Australia

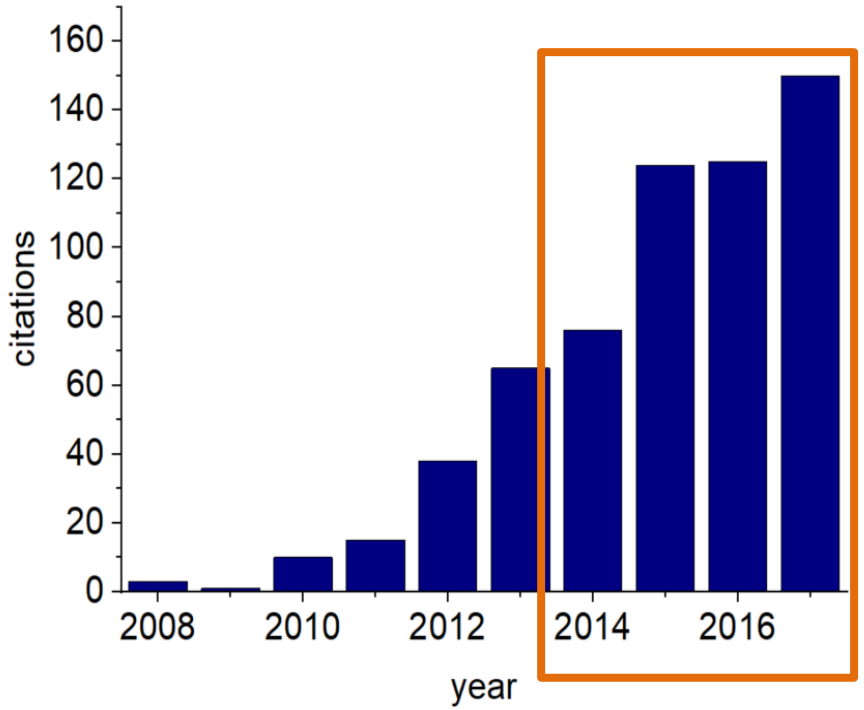
[‡]College of Environmental Science and Engineering, State Key Laboratory of Pollution Control and Resource Reuse, Tongji University, 1239 Siping Road, Shanghai 200092, China

[§]Fachbereich Physik und Chemie and CHyN, University of Hamburg, 22607 Hamburg, Germany

[⊥]Department of Chemistry, University of Warwick, Gibbet Hill, Coventry CV4 7AL, United Kingdom

[∥]ARC Centre of Excellence in Convergent Bio-Nano Science and Technology, and the Department of Chemical Engineering, The University of Melbourne, Parkville, Victoria 3010, Australia

Can we mine the PC literature?



PubMed search for "Protein Corona", Sept 1, 2018.

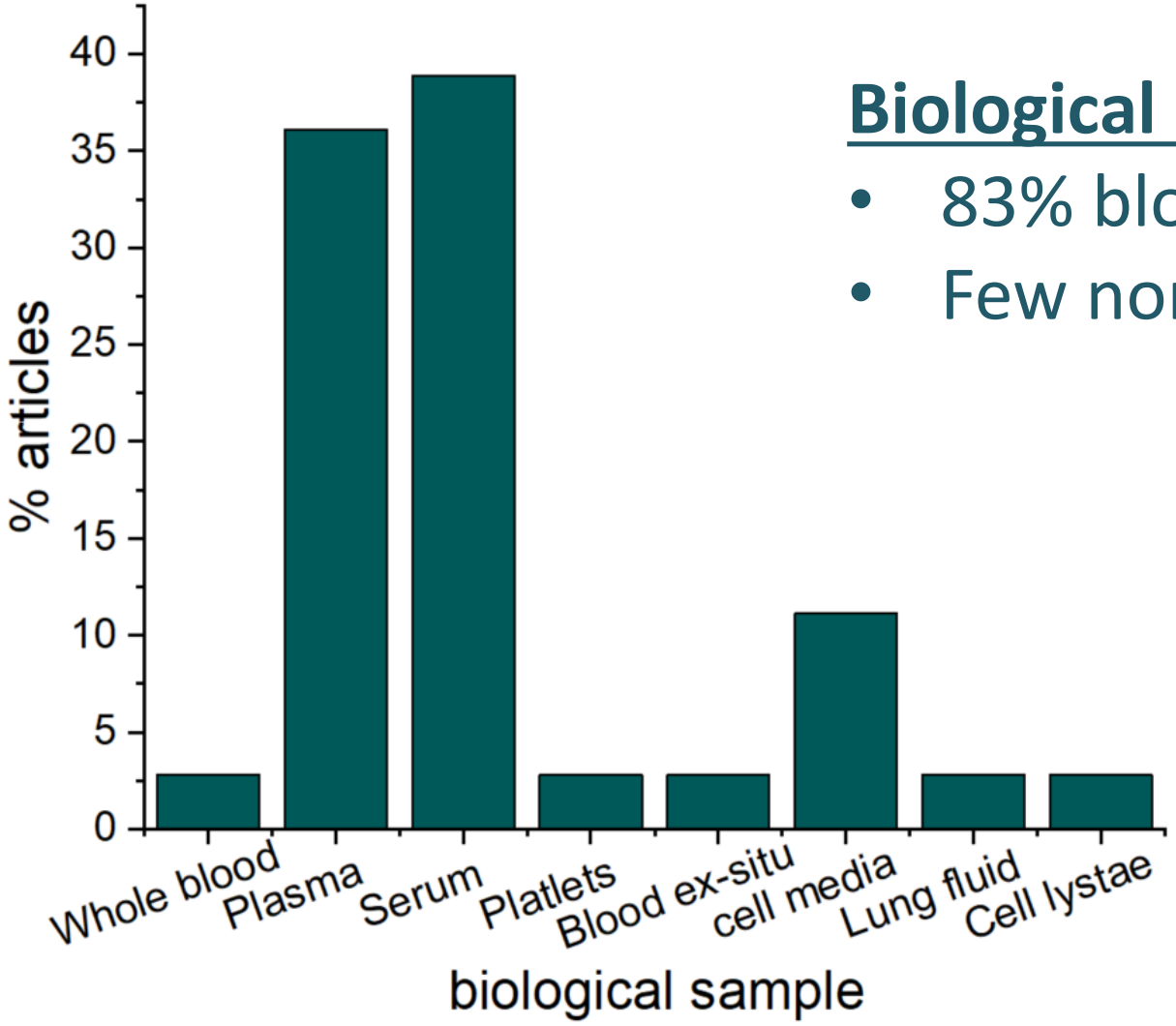
To assess the first decade

Top 110 cited "protein corona" papers from 2014-2017

Reviewed for:

1. diversity of systems
2. data reporting & preproducibility

1. Diversity of Systems

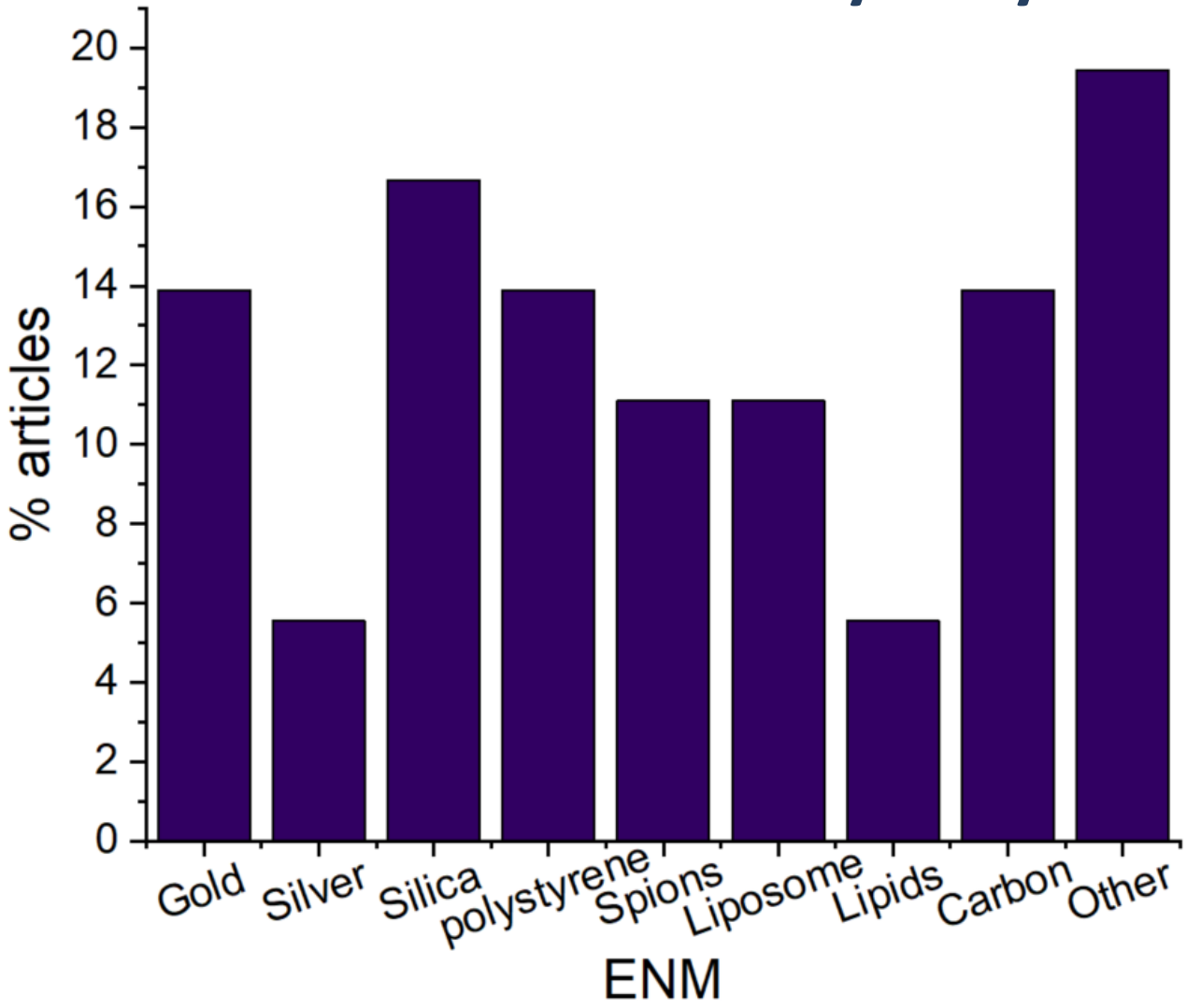


Biological samples

- 83% blood products.
- Few non-human studies

Chetwynd, et al *submitted.*

1. Diversity of Systems

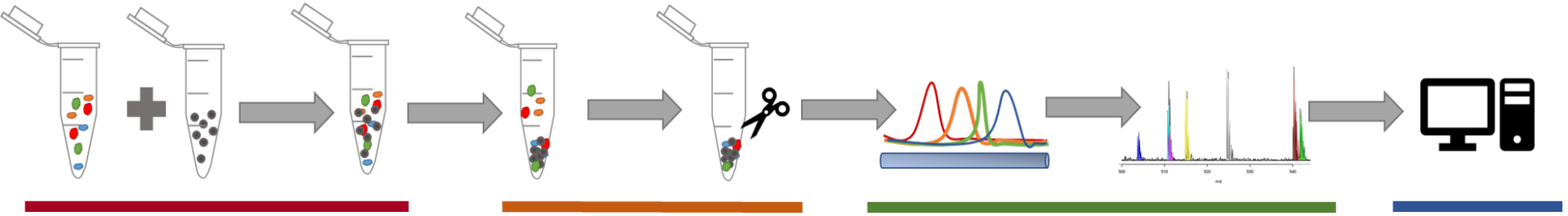


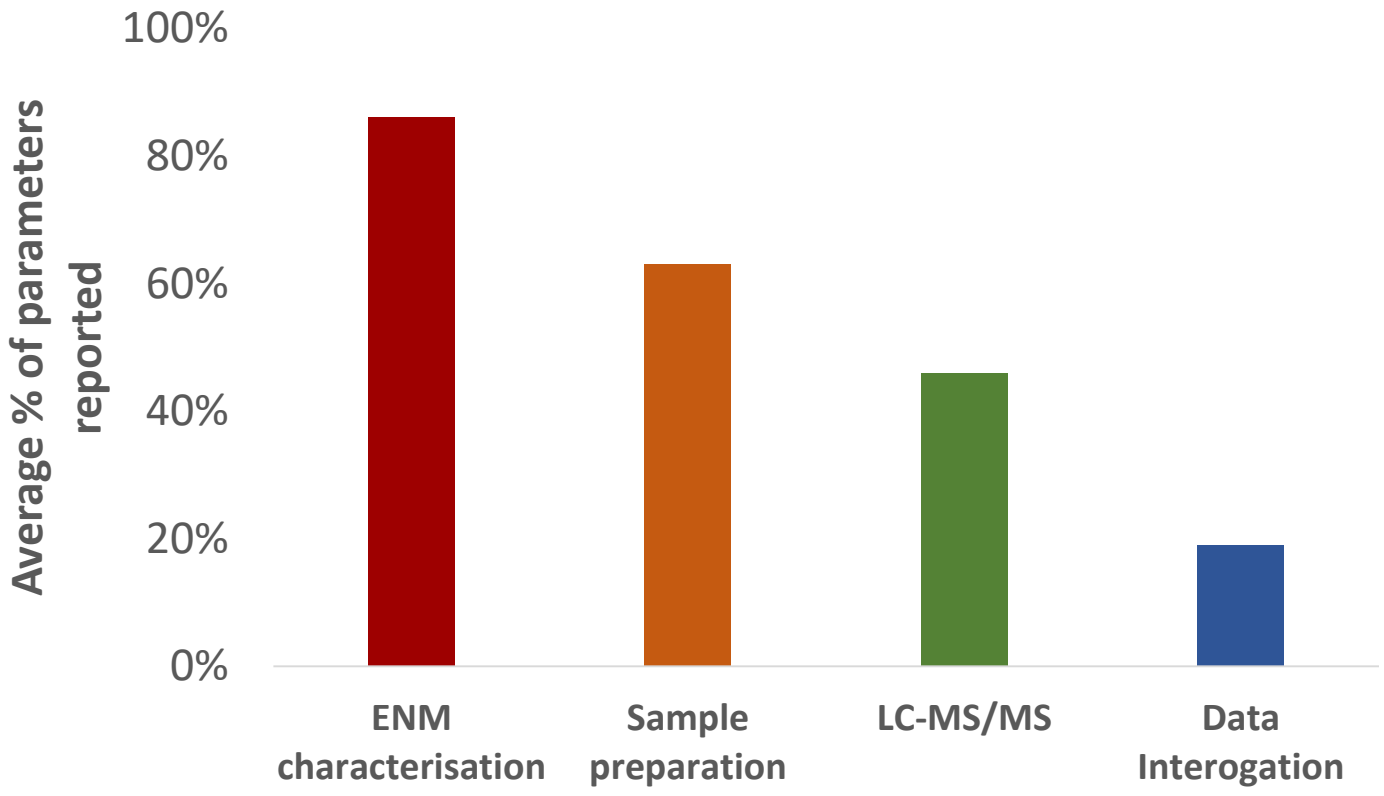
Engineered Nanomaterial

- 20 % metal ENMs
- 30 % silica & polystyrene
- No "next generation" ENMs

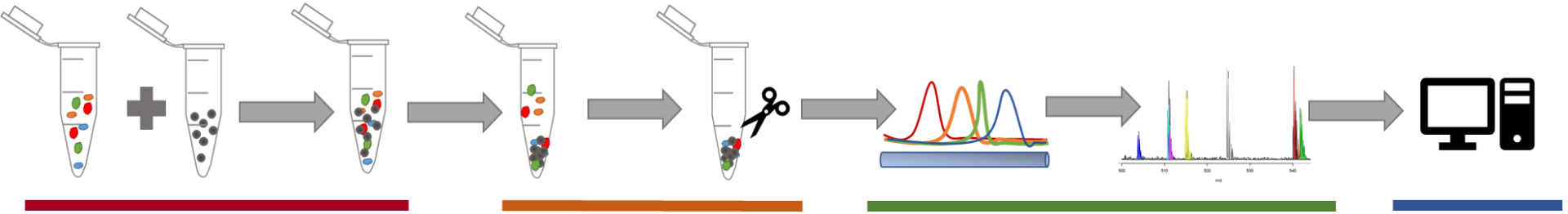
Chetwynd, et al *submitted*.

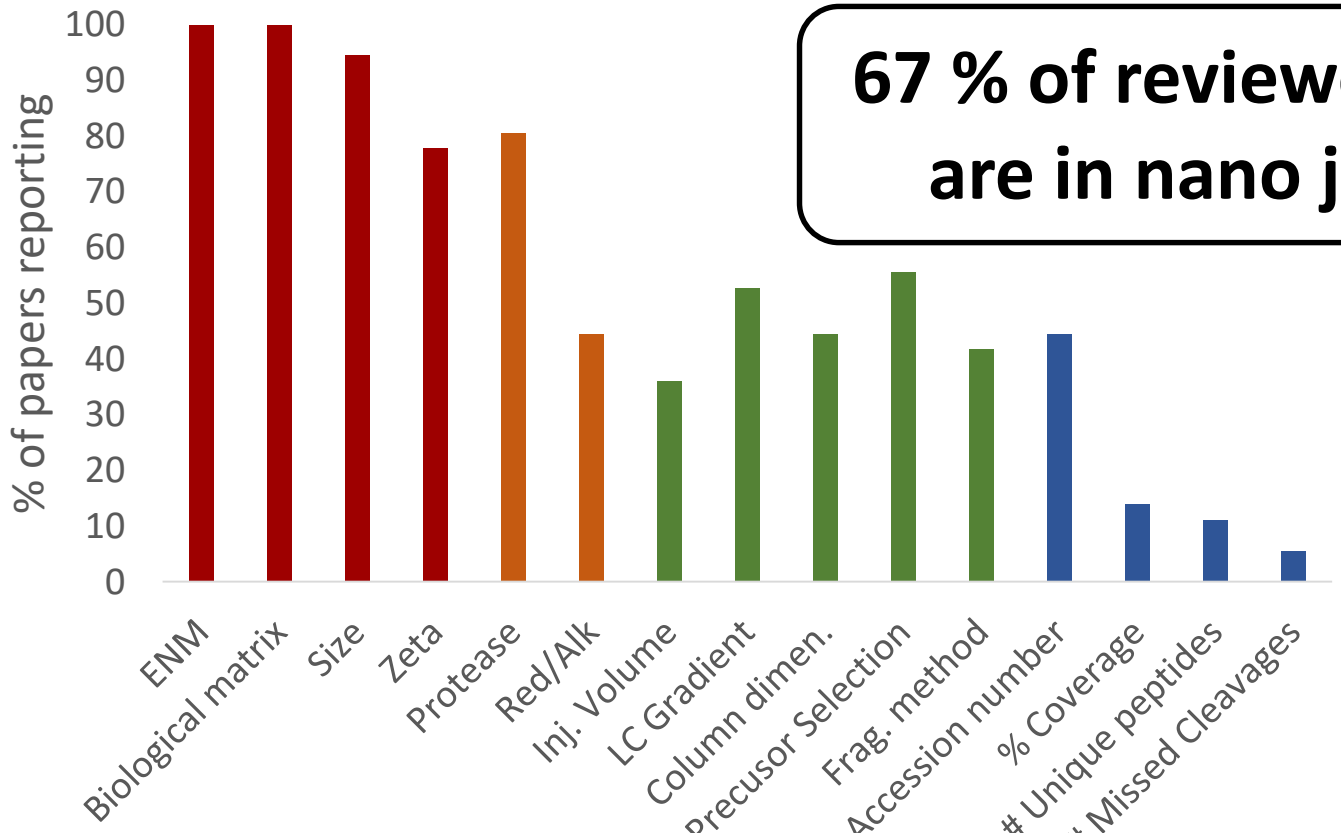
2. Assessing preproducibility





Chetwynd, et al *submitted.*

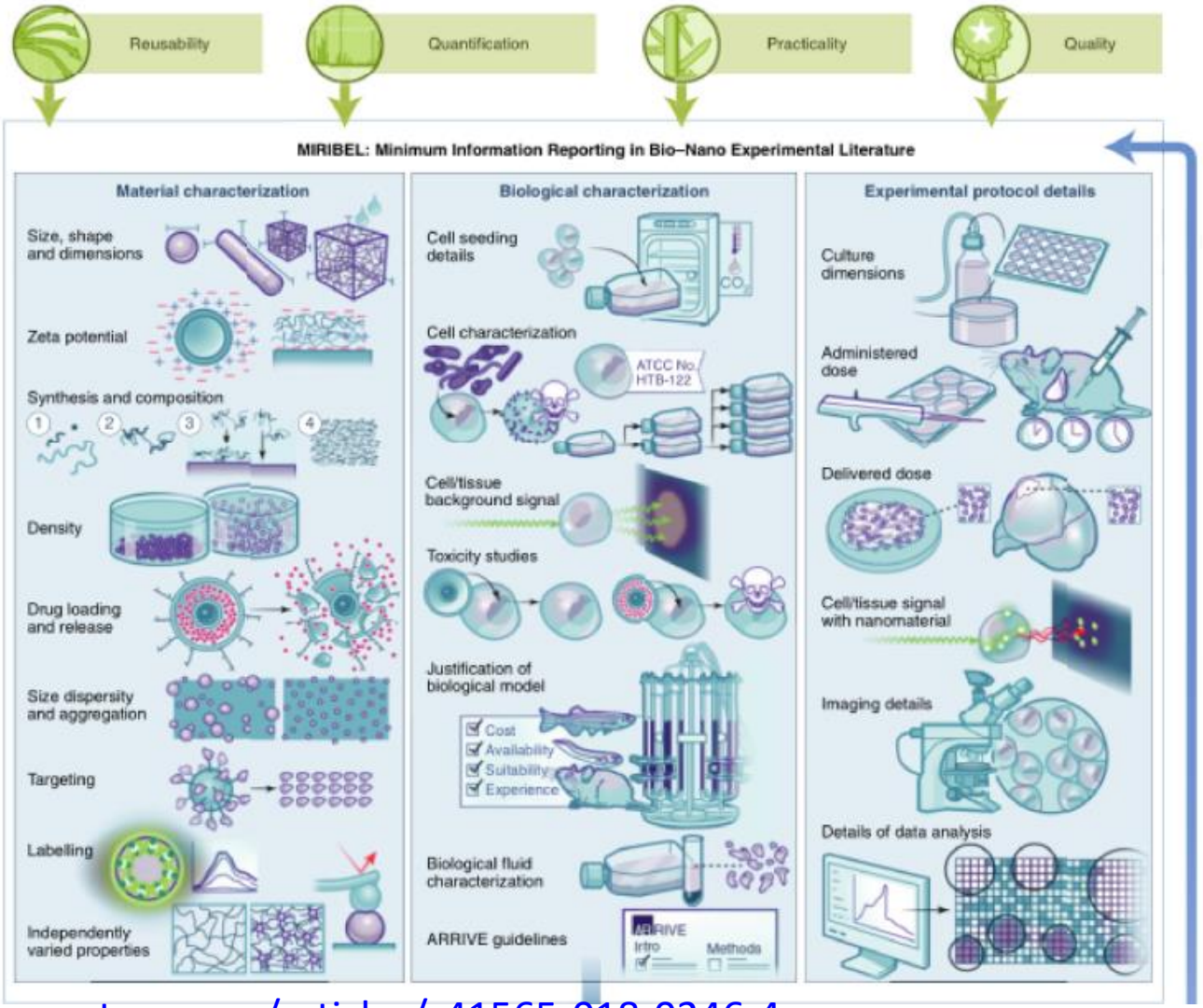




67 % of reviewed papers are in nano journals

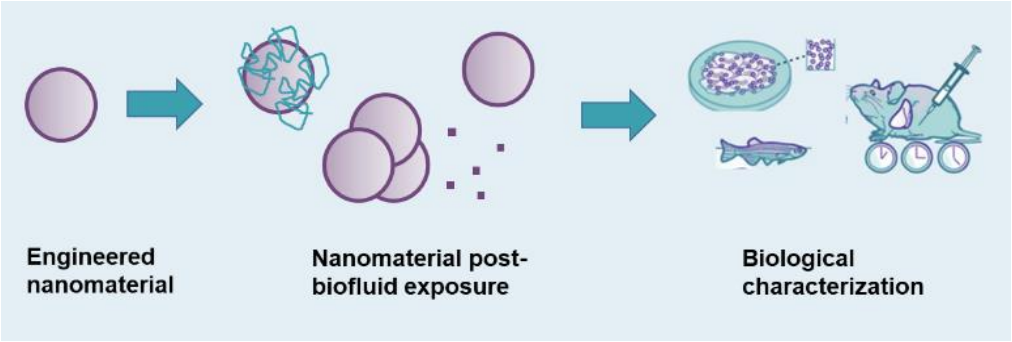
Chetwynd, et al *submitted.*

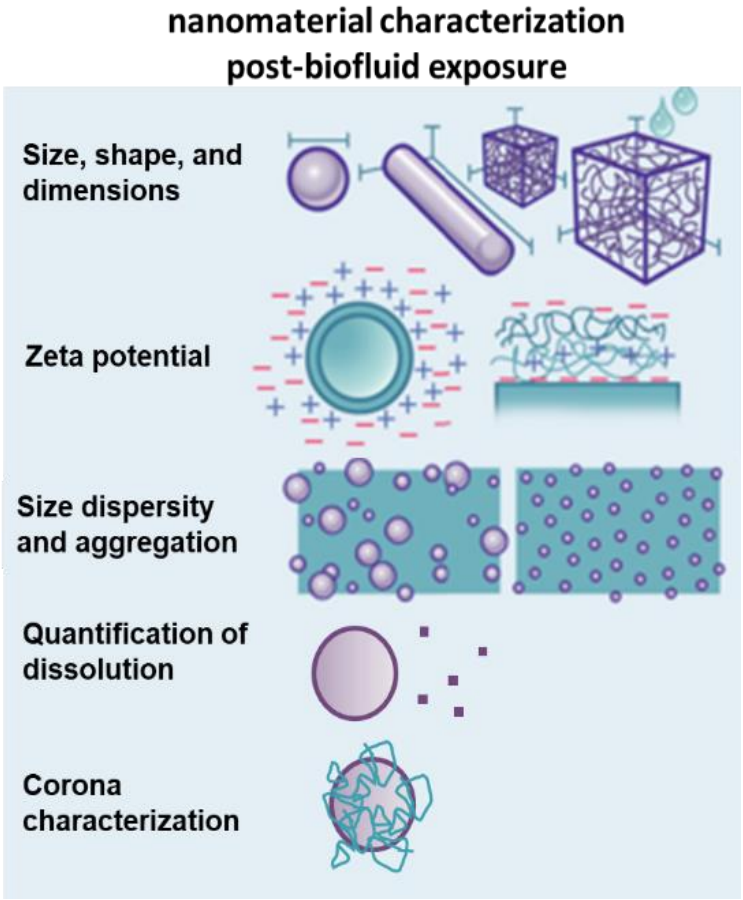
NATURE NANOTECHNOLOGY PERSPECTIVE



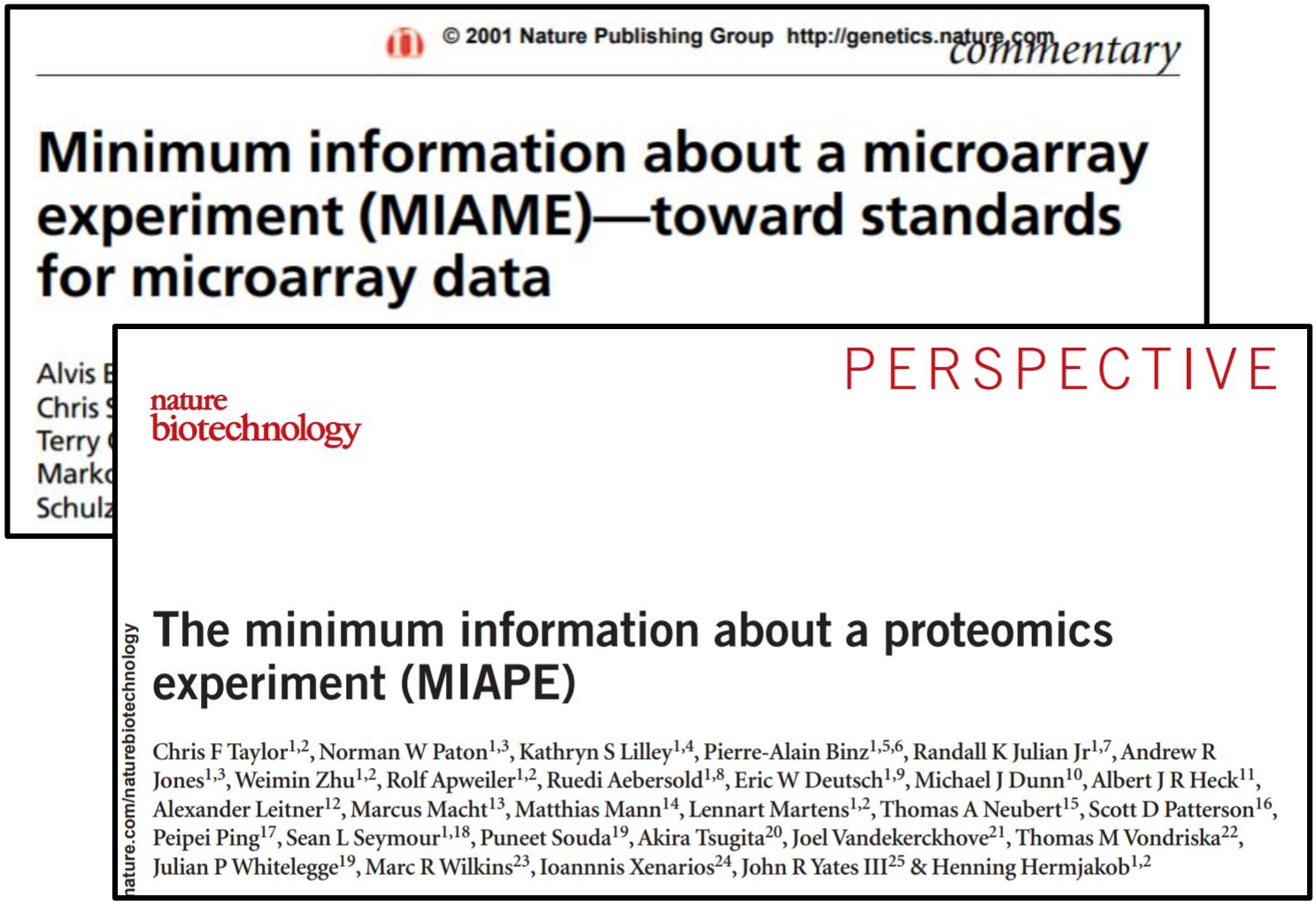
https://www.nature.com/articles/s41565-018-0246-4

But nanomaterial transform in a biological system…





Chetwynd, et al *submitted.* Figures modified from https://www.nature.com/articles/s41565-018-0246-4



commentary

Minimum information about a microarray experiment (MIAME)—toward standards for microarray data

Alvis B
Chris S
Terry (
Marko
Schulz

PERSPECTIVE

nature biotechnology

The minimum information about a proteomics experiment (MIAPE)

Chris F Taylor[1,2], Norman W Paton[1,3], Kathryn S Lilley[1,4], Pierre-Alain Binz[1,5,6], Randall K Julian Jr[1,7], Andrew R Jones[1,3], Weimin Zhu[1,2], Rolf Apweiler[1,2], Ruedi Aebersold[1,8], Eric W Deutsch[1,9], Michael J Dunn[10], Albert J R Heck[11], Alexander Leitner[12], Marcus Macht[13], Matthias Mann[14], Lennart Martens[1,2], Thomas A Neubert[15], Scott D Patterson[16], Peipei Ping[17], Sean L Seymour[1,18], Puneet Souda[19], Akira Tsugita[20], Joel Vandekerckhove[21], Thomas M Vondriska[22], Julian P Whitelegge[19], Marc R Wilkins[23], Ioannnis Xenarios[24], John R Yates III[25] & Henning Hermjakob[1,2]

nature.com/naturebiotechnology

Minimum Information about Nanomaterial Biocorona Experiments (MINBE)

Experimental design checklist for biocorona characterization

Phase 1: ENM exposure to biofluids and ENM characterization

- How can the biofluid be handled and stored to best reflect biological conditions?
- What is the most relevant and appropriate dosage of ENMs in the chosen biofluid or hypothesis?
- Does the study aim for analysis of the biocorona pre- or post-equilibrium? How was the timeline for equilibrium evaluated?
- Are replicates included? And, where possible, do plans include parallel processing of samples for proteomics and additional characterisation?

Phase 2: Isolation of absorbed biomolecules and preparation for analysis

- Is the biocorona separation thorough enough to remove all unbound biomolecules?
- Is the separation technique likely to alter the chemical structure of the biocorona, or the profile of biomolecules?
- Will the biocorona clean-up steps affect the profile of your biomolecules?
- What clean-up methods are likely to least alter the biocorona, e.g. through selective or pervasive protein loss?

Phase 3: separation of biomolecules and spectroscopic characterization

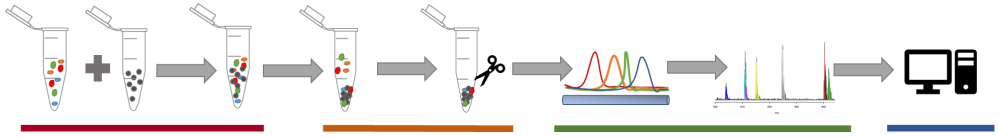
- What type of separation will be best for these samples (in gel, on-particle etc.)?
- How can the protocol be modified to best maintain consistency across samples?
- What instruments are best for these samples and quantification?
- What controls can be included to check for protein identification and quantification?

Phase 4: Informatic identification of the biocorona population and analysis.

- What informatics database will be used for identifying the biomolecules?
- What kind of post-processing is required for quantification?
- What statistical analysis of the dataset can be used to assess data quality?
- How will the full dataset of biomolecules and their characteristics be organized, analysed, and presented?

Chetwynd, et al *submitted.*

Minimum Information about Nanomaterial Biocorona Experiments (MINBE)



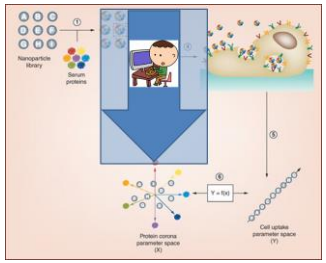
Protein identification and quantification [compare to MAIPE-MSI]	
Database	List the database used for protein identification, including version and any restrictions applied in the
Accession number	Accession numbers provide a unique identifier for each protein from the queried database and serve as the foundation for bioinformatic analyses. With the accession code, readers can query databases for any additional bioinformatic information required.
Confidence in protein identification	• % Coverage: The percentage of the protein sequence covered. • Number of peptides: Total number of peptides detected for each protein, ideally 2 or more peptides for a confident identification. • Number of unique peptides: Number of peptide sequences that are unique to the identified protein. At least one required for a high degree of confidence in the identification. • Missed cleavages: Number of missed cleavages in the protein or peptide sequence. A good indicator of digestion efficiency, missed cleavages, should be kept to 2 or less for a confident identification. • Protein probabilities and scores: Calculated probabilities or scores to give a sense of confidence in a protein identification.
Validation	Statistical analysis or comparisons of replicates should be performed to assess data quality.
Quantification	Details on both the normalisation and quantification method required to enable accurate reproducibility between experiments. Often, label free quantification includes calculation of a normalized spectral abundance factor (NSAF), which includes both the spectral count and length of each protein.
Additional outputs	Additional outputs often included: subcellular localization of the ENMs, amino acid sequence, molecular weight, function, and other biophysical parameters for each protein.

Chetwynd, et al *submitted.*

Technical challenges

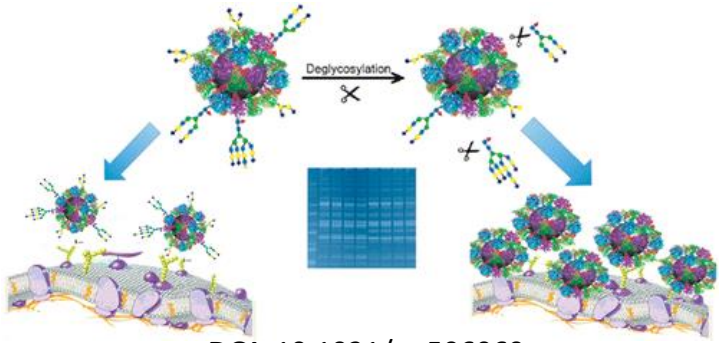
1. modeling

- Comparative analysis
- Predictions of PC formation & dynamics.



2. PTMs / small molecules

- Expansion to challenging molecules
- Deeper insight into chemistry of the PC

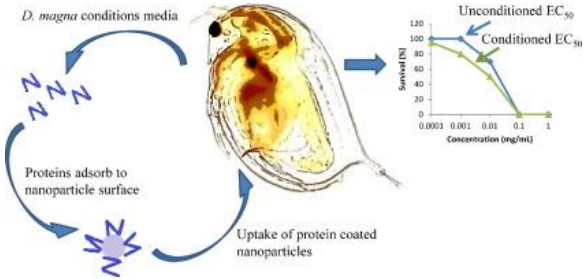


DOI: 10.1021/nn506060q

Complex systems

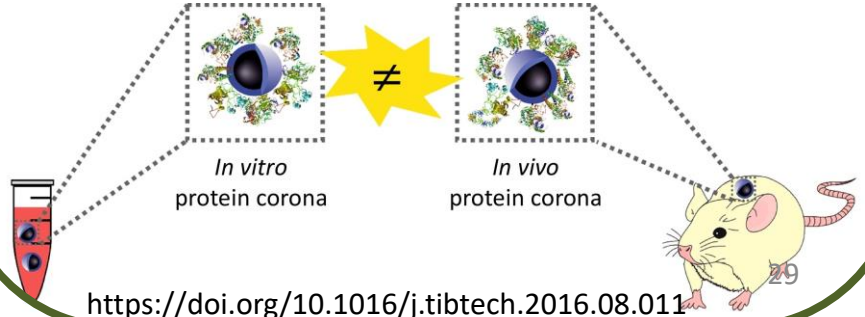
- Wider variety of species / proteins.
- Dynamic / fluidic variations.
- Small molecules.
- Personalized protein corona.

1. Eco-corona

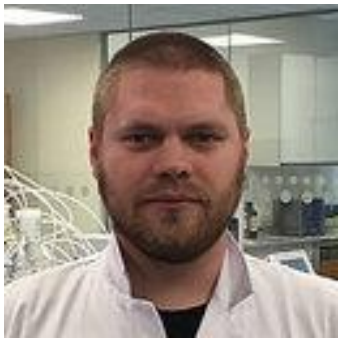


https://doi.org/10.1016/j.jprot.2015.09.005

2. In vivo studies



https://doi.org/10.1016/j.tibtech.2016.08.011



The MINIBE team

UNIVERSITY OF BIRMINGHAM

Andrew Chetwynd

Iseult Lynch

ACEnano

Analytical and Characterisation Excellence

SCU proteomics / modeling team

Maryam Mobed-Miremadi (Bioengineering)

Rich Eigenheer (UC-Davis Proteomics)

Santa Clara University 1851

Daniel Freitas

Matthew Findlay

Erick Castellanos

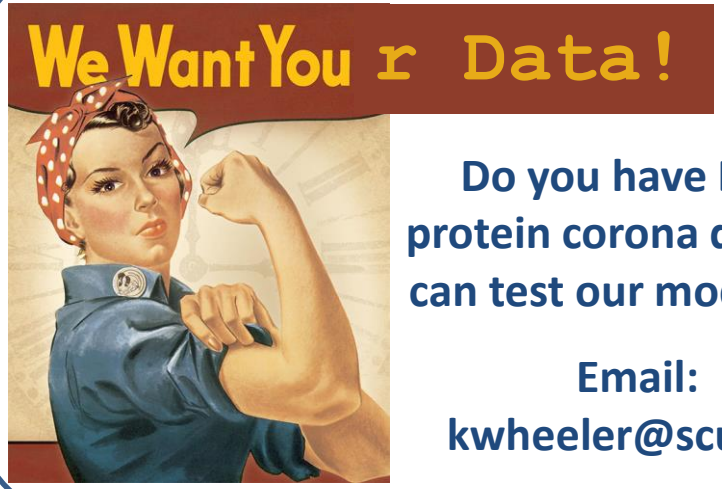
Alyssa Lampe

NIH NIEHS

ETH zürich

Kris McNeill's Lab

Thank you!



Do you have ENM protein corona data we can test our model on?

Email: kwheeler@scu.edu