

Molecular Modeling and Machine Learning for Small Molecule and Biologic Drug Formulation

Schrödinger

Irene Bechis, PhD – Senior Scientist (Materials Science)

Dan Cannon, PhD – Principal Scientist (Life Science)



MISSION

To improve human health and quality of life
by transforming the way **new medicines**
and **materials** are discovered through
advanced **computational methods**

Our unique commitment to scientific innovation, software development, and support



30+ years of innovation in molecular modeling scientific research and product development



Over 800 employees worldwide; >40% Ph.D.



More than 50% of the company dedicated to research and development



Large scientific support & education teams offer expertise and support in knowledge-transfer



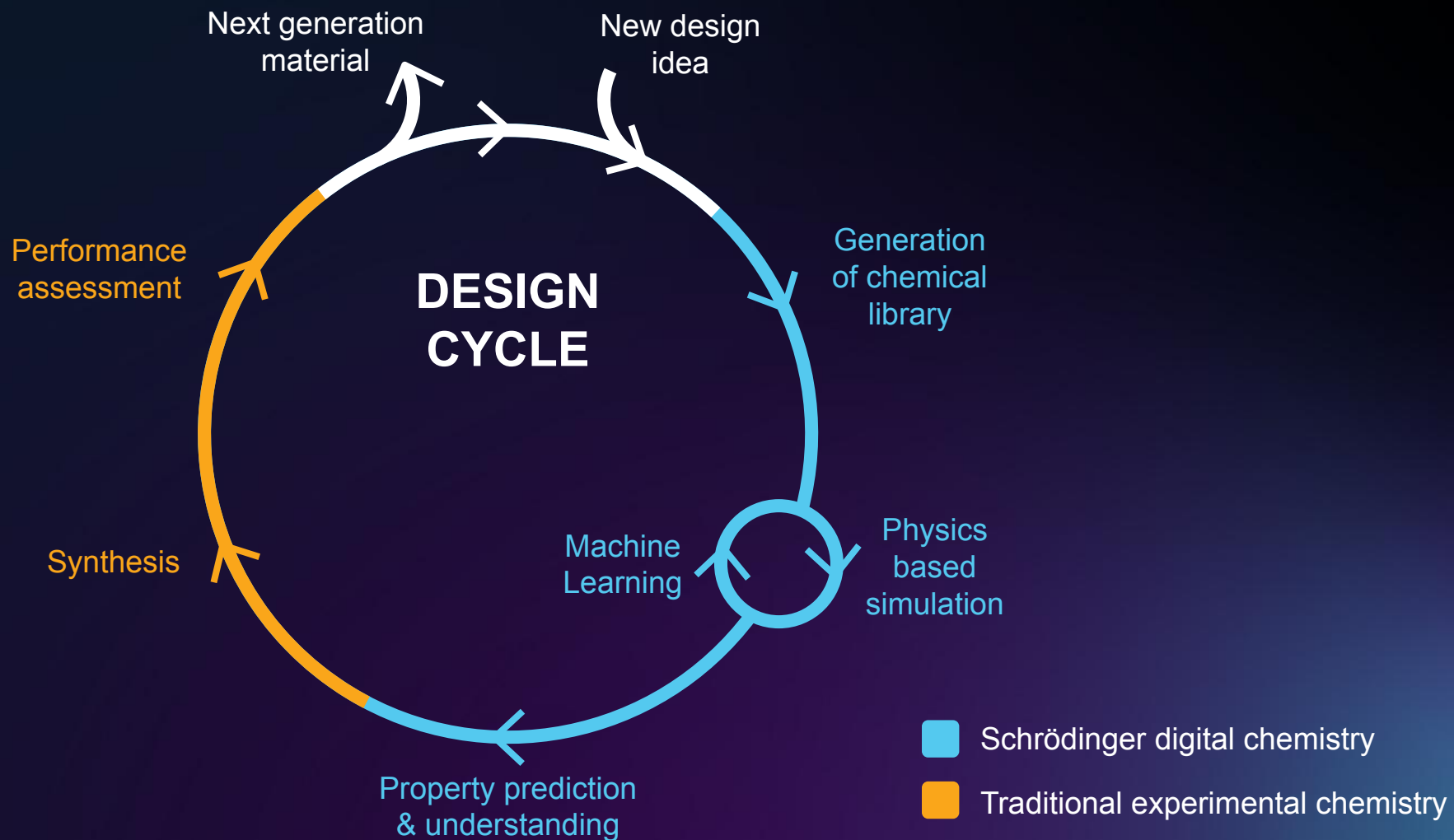
Quarterly software releases with performance and feature improvements

>800 employees spanning multiple sites across the globe



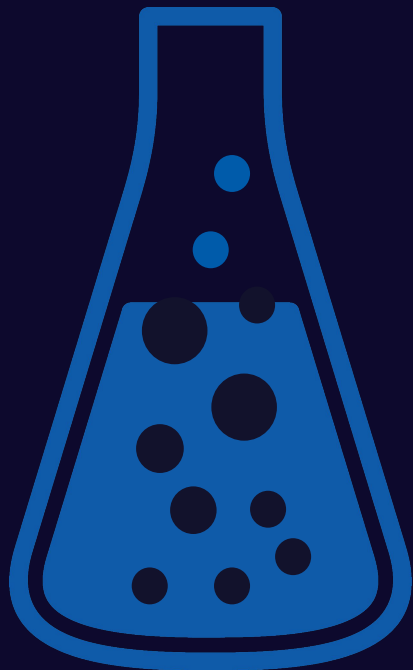
Value of Modeling

Accelerate your design cycle



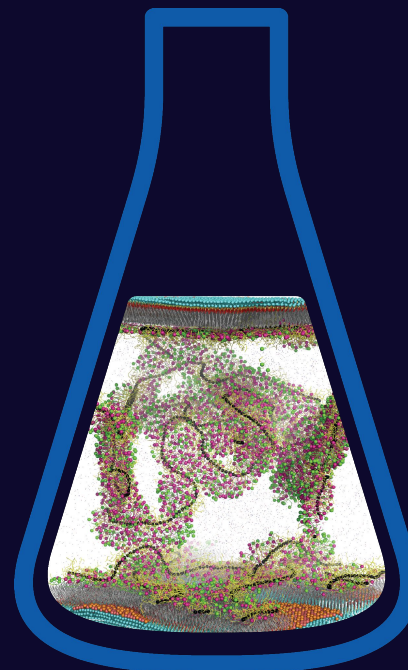
Increased understanding

Traditional R&D



Trial-and-error approach to test inputs and observe outputs

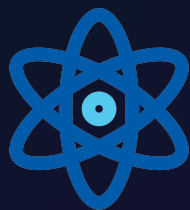
Incorporating Computation



Simulation reveals molecular-level resolution, linking structure to function

Software Platform

Digital chemistry strategy built on three pillars



**Physics-based
modeling**

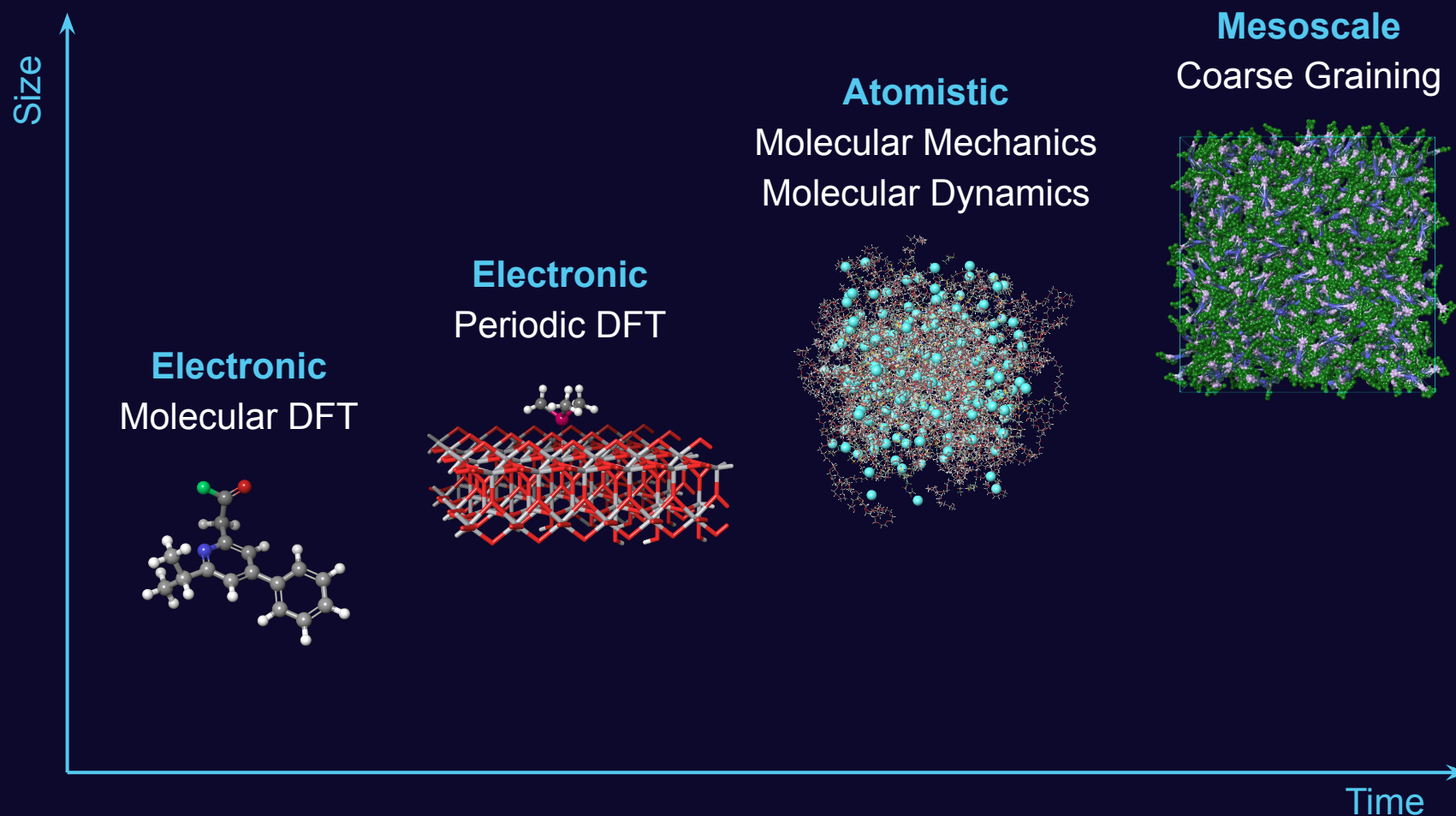


**Machine
learning**



**Collaborative
informatics**

Simulate: access all methods in a single platform



Amplified by



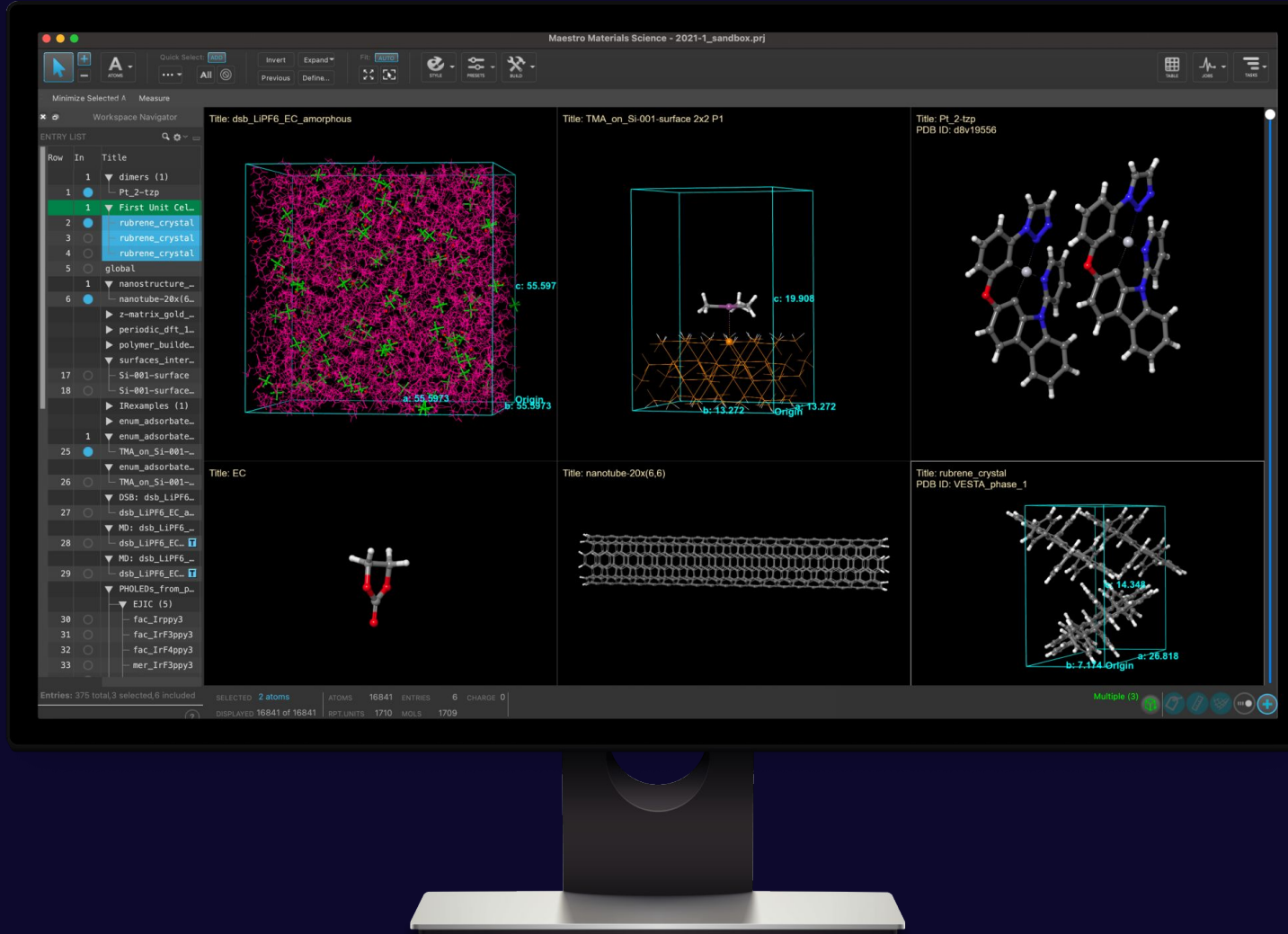
Machine Learning



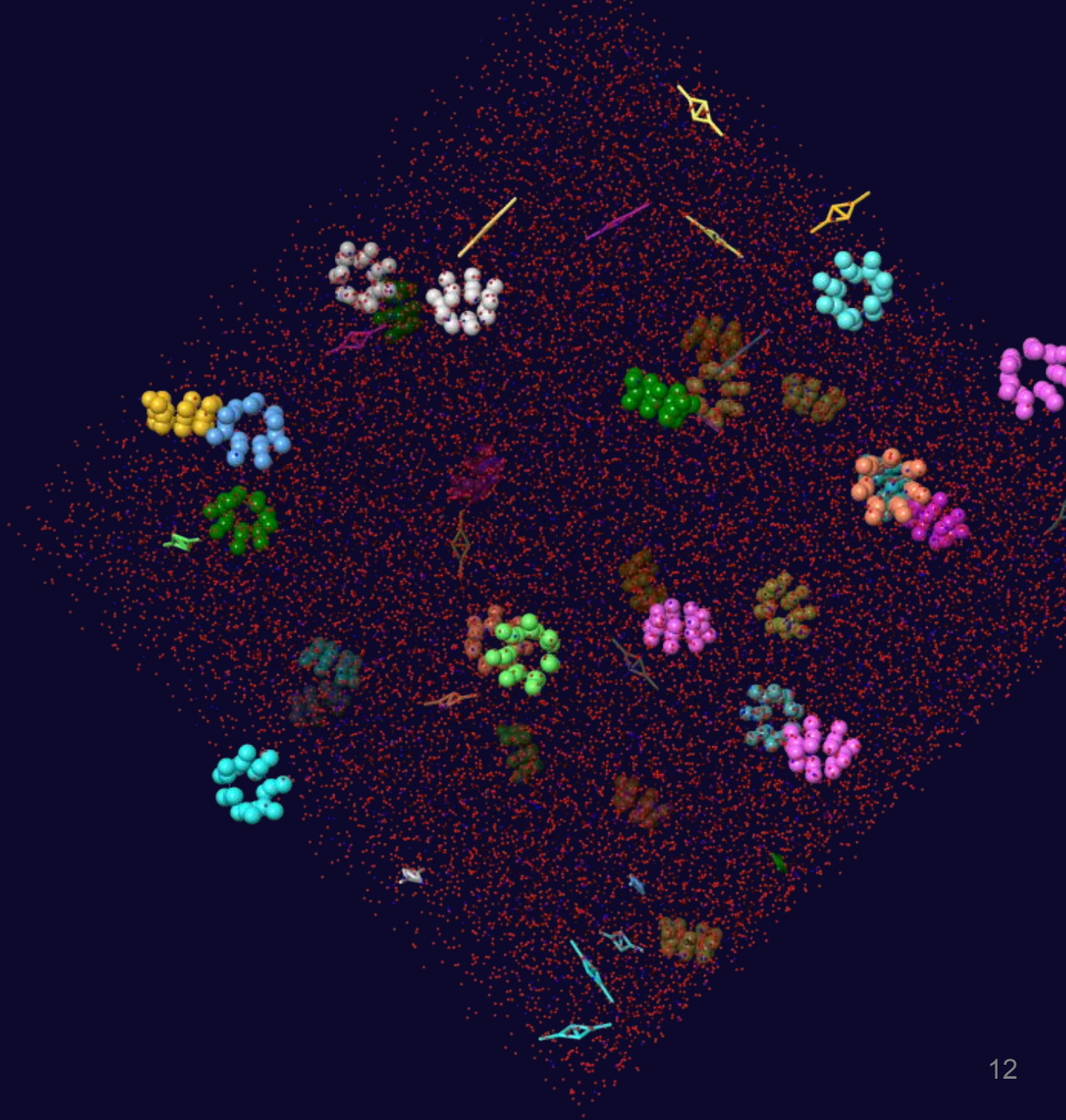
Enumeration /
Library generation

Modern, integrated, user-friendly GUI

MS Maestro



Modeling and Simulation for Pharmaceutical Formulation & Delivery

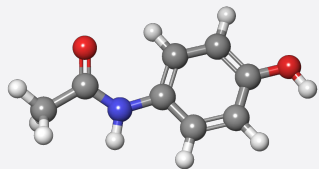


Capabilities for drug characterization and formulation

Application	Capabilities
Characterization	pKa , powder X-ray diffraction (XRPD), crystal morphology, density of crystalline or amorphous phases
Catalysis, Reactivity, Degradation	QM multistage workflow, bond dissociation energy, nanoreactor (API degradation), AutoTS (transition state searching), reaction channel enumeration, Auto Reaction Workflow (catalyst design), conformational search
Spectroscopy	VCD, solution-state NMR, solid-state NMR, IR, UV-Vis
Crystal Structure Prediction	Crystal structure prediction
Formulation and Delivery	Creating machine learning models for formulations , system builders (mixtures, polymers, surfactants, lipids, etc.), aqueous and solvent solubility of amorphous/crystalline API , API aggregation, glass transition temperature, mechanical properties, wettability (contact angle), separation during solvent removal (evaporation), API encapsulation in cyclodextrin, etc., excipient selection and ASD formulation , API solubility and LogP in excipient , solubility parameters , API — excipient mixing enthalpy, ASD separation and dissolution , protein/biologics excipients selection , hygroscopicity (moisture sorption) in amorphous solid dispersions, APIs, tablet coatings, solution viscosity , protein/polymer interactions , lipid nanoparticles , liposomes

Machine Learning for Formulations

Featurization: diverse materials systems

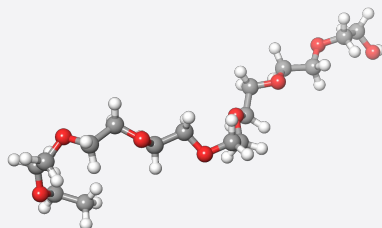


Small molecules

- Physiochemical, topographical descriptors
- Binary fingerprints (RDKit, Canvas)
- Graph-based convolution neural networks

Physics-based descriptors:

- Quantum mechanics
- Auto Reaction Workflow

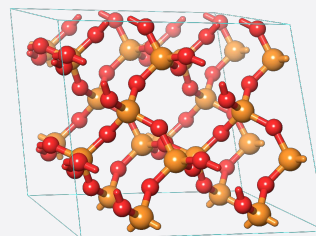


Polymers

- Taking into account connections between repeat units
- RDKit fingerprints + customized descriptors

Physics-based descriptors:

- Molecular dynamics
- Quantum mechanics

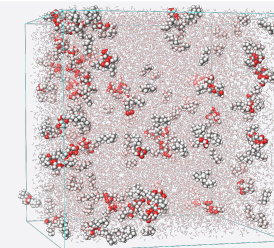


Periodic inorganic solids

- Element
- Lattice structure
- Oxidation state
- Intercalation descriptors
- 3D SOAP (with PCA)

Physics-based descriptors:

- Periodic quantum mechanics



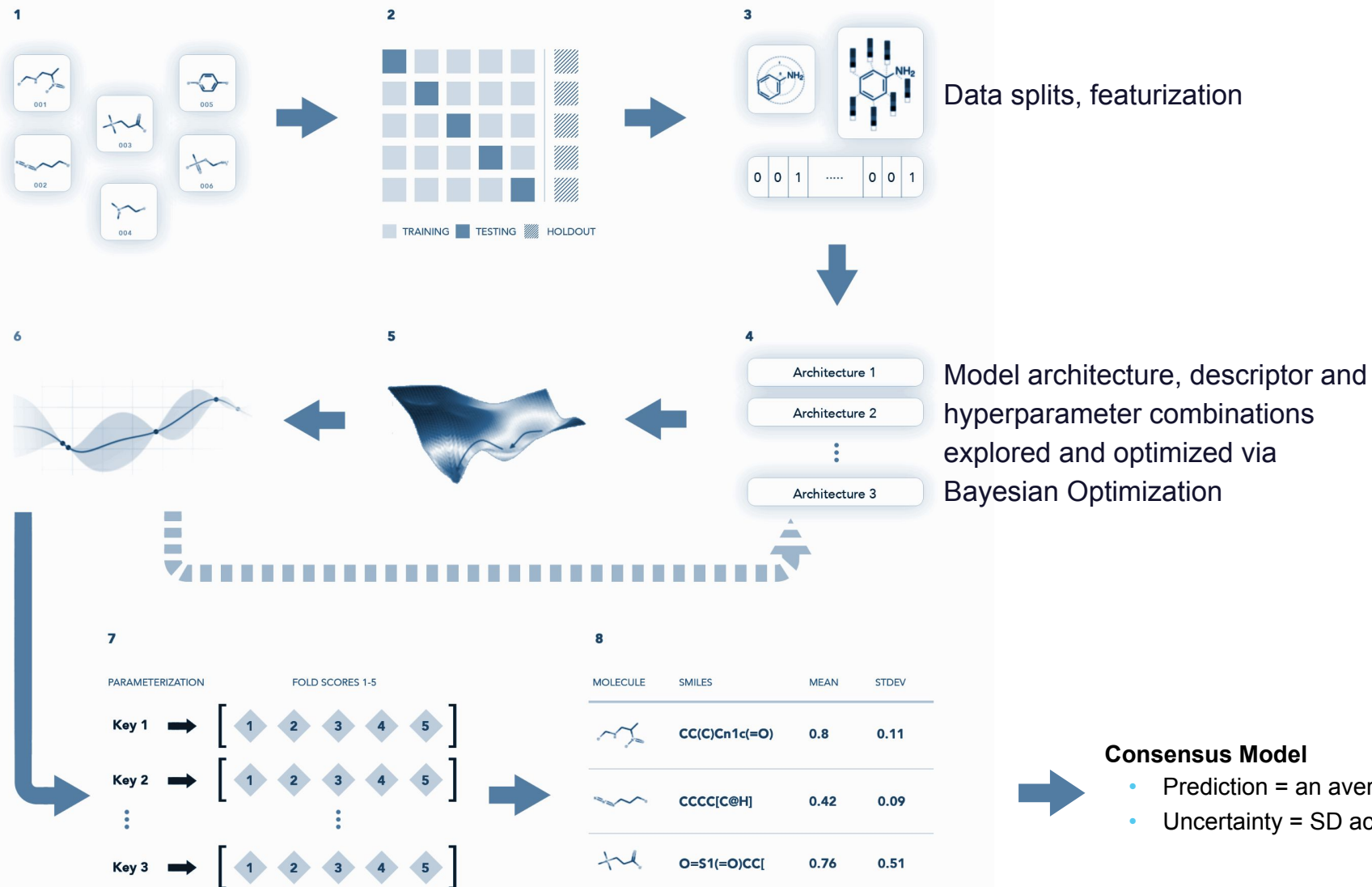
Formulations and mixtures

- Composition
- Chemistry of the components
- Experimental/processing conditions

Physics-based descriptors:

- Molecular dynamics

Automated machine learning: DeepAutoQSAR



Models Sampled

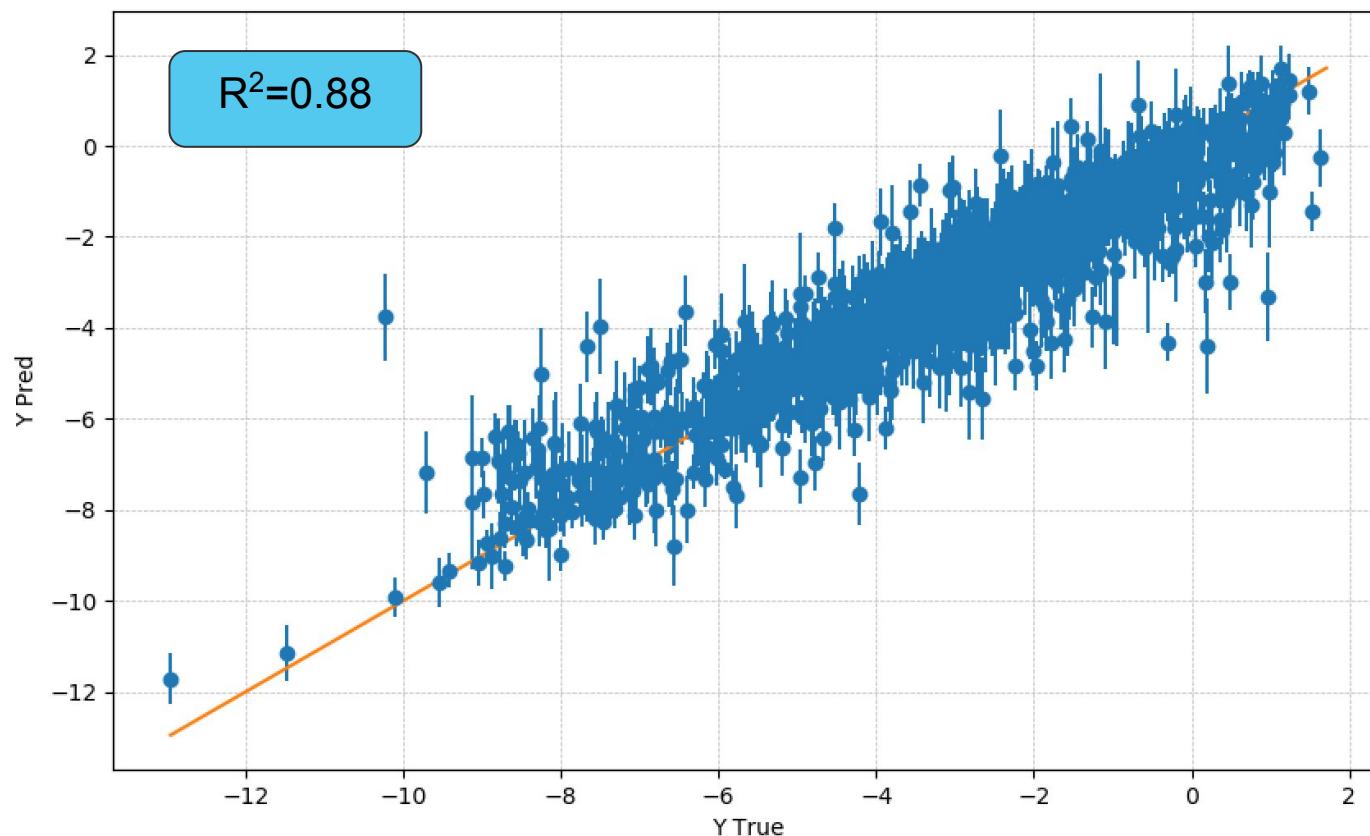
- Dense Neural Network
- Random Forest Regressor
- XGBoost
- TorchGraphConv
- GCN
- GraphSAGE
- GIN
- TopK
- SAGPool
- EdgePool
- GlobalAttention
- Set2Set
- SortPool

Consensus Model

- Prediction = an average of the predictions for 5 best models
- Uncertainty = SD across the 5 predictions

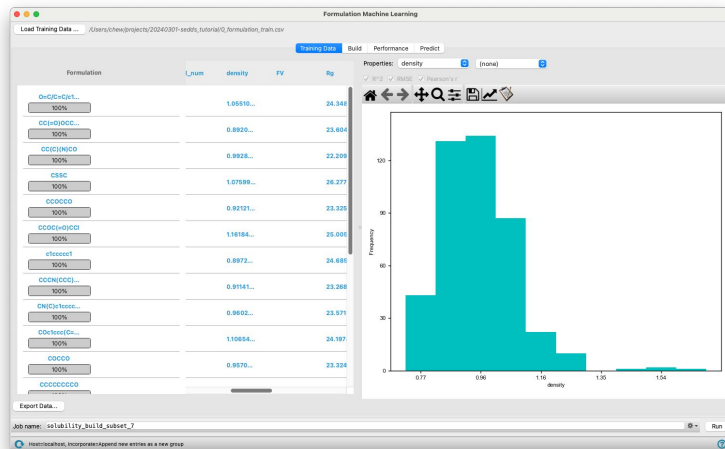
Machine learning for aqueous solubility

- 8,773 molecules spanning 10 elements
- Trained on a GPU for 12 hours using DeepChem/AutoQSAR

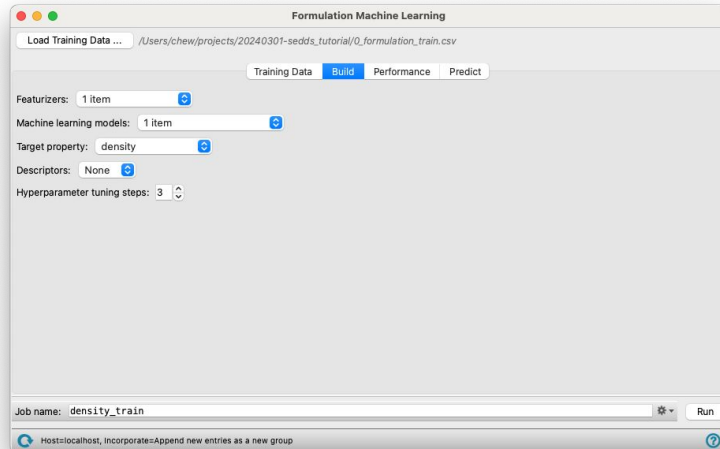


Formulation machine learning

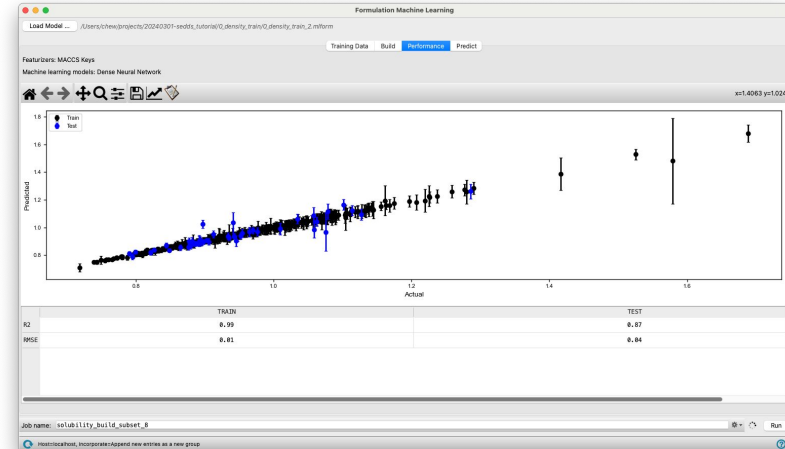
Data loading/visualization



Model training



Model evaluation/prediction



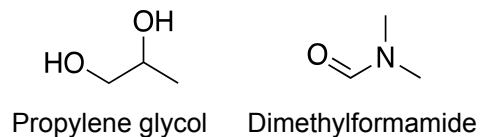
- Build, validate, and apply machine learning models based on chemistry and composition to predict any formulation property (e.g. density, viscosity)
- Allows input of external conditions, such as temperature, pressure, and so on
- Fully automated machine learning approach with automatic hyperparameter tuning

Drug solubility: Bao *et al.* dataset

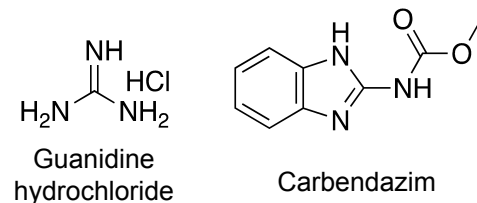
Dataset summary

28,703 total examples: 128 drugs, 44 solvents

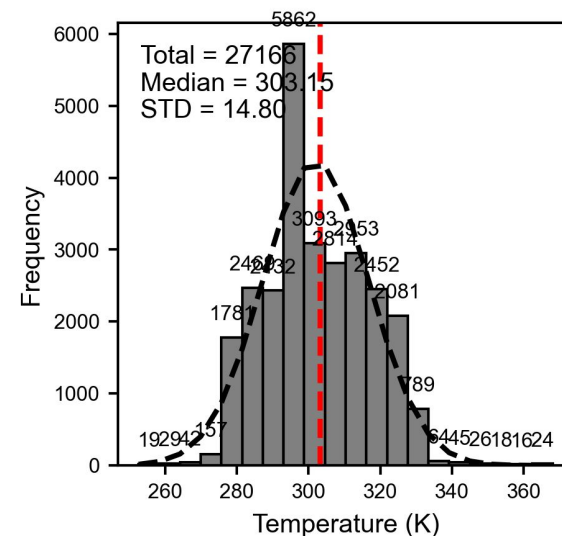
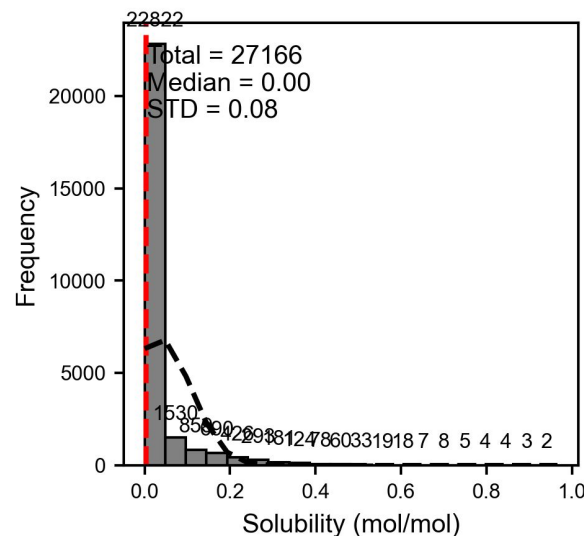
Example of solvents



Example of drugs



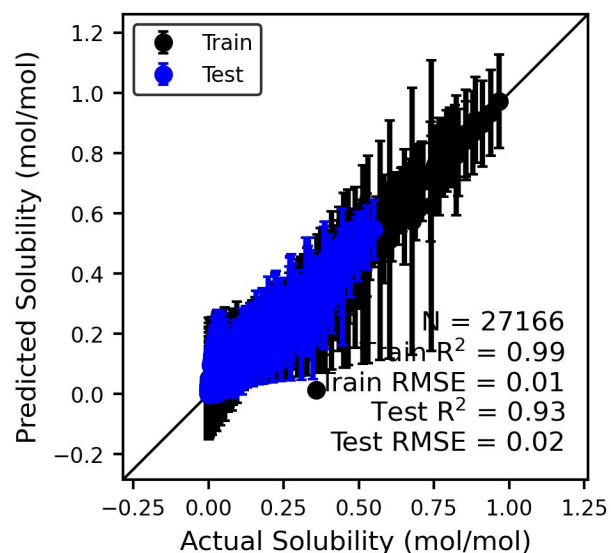
Distribution of solubility and temperature



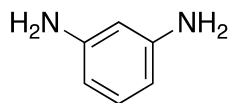
Goal: Predict temperature-dependent solubility of drug in pure/binary solvents

Drug solubility: formulation machine learning

Parity plot
Out-of-sample
90:10 train:test split

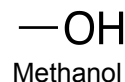


Drug

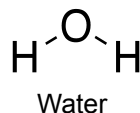


1,3-phenylenediamine

Solvents

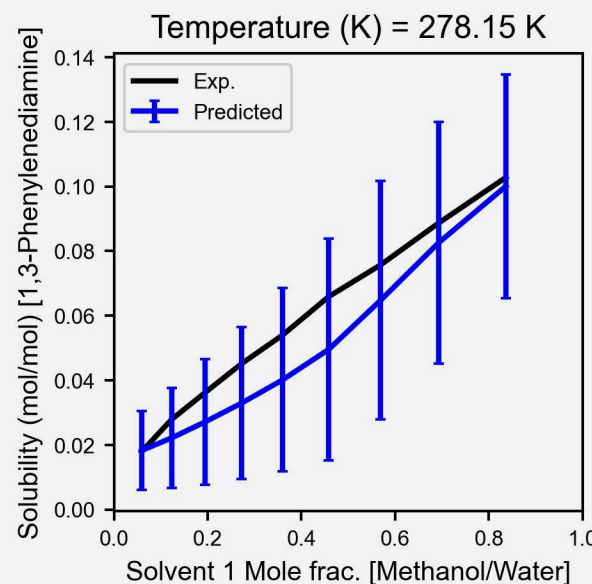


Methanol

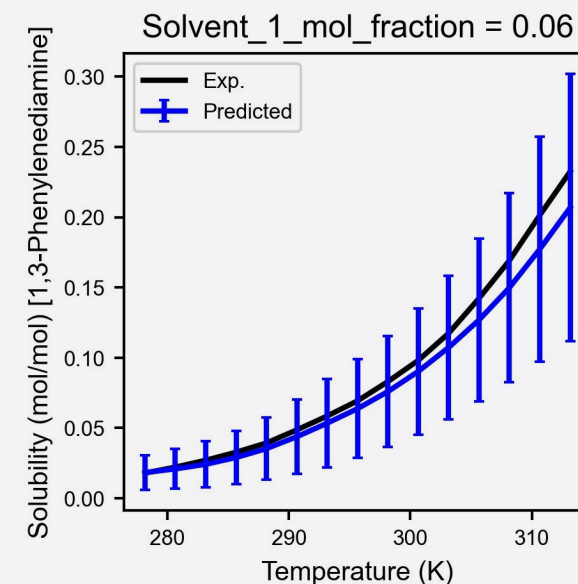


Water

Test set prediction as a function of
composition for binary mixture



Test set prediction as a function of
temperature



- Formulation ML models used to create an accurate model for predicting solubility
- Model achieves an average test set R^2 of ~0.93
- Model enables tunability in both composition and temperature

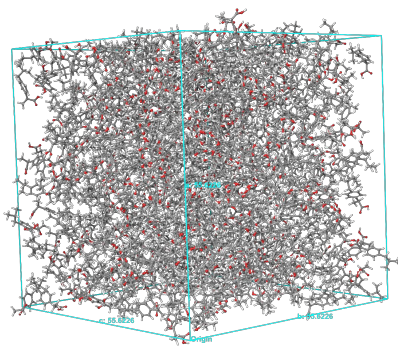
Small molecules formulation and delivery

Solubility parameters

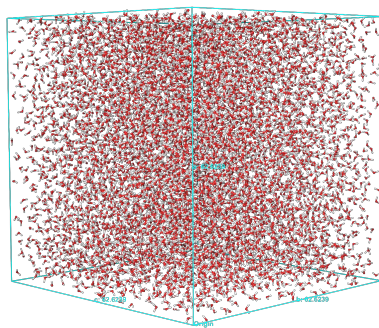
Calculation of cohesive energy density allows for prediction of Hansen/Hildebrand solubility parameter (δ).
Smaller differences in δ ($< \sim 7 \text{ MPa}^{1/2}$) indicate miscibility

- Fast miscibility screening for API / solvent combinations
- Screening of 100s-1000s of solvents enables construction of a miscibility matrix
- API input can be amorphous and/or crystalline

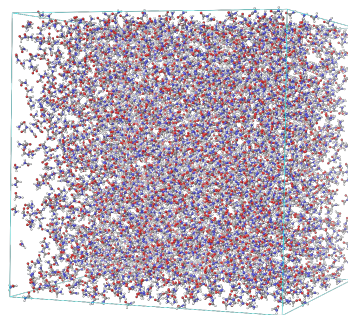
Drug	Solvent	$\Delta\delta \text{ (MPa)}^{1/2}$ (predicted)	Miscible? (experiment) ¹
Ibuprofen	Water	25.4	No ($1.5 \times 10^{-6} \text{ M}$)
	Formamide	14.9	No ($1.4 \times 10^{-3} \text{ M}$)
	Cyclohexane	6.1	Yes (0.15 M)
	Ethyl Acetate	2.9	Yes (0.33 M)



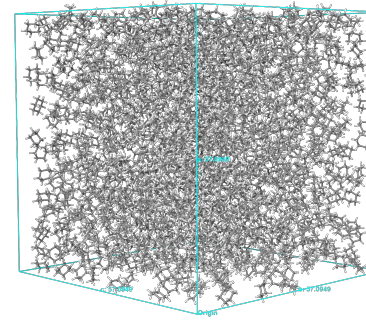
ibuprofen



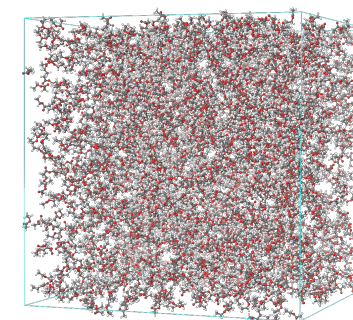
water



formamide



cyclohexane

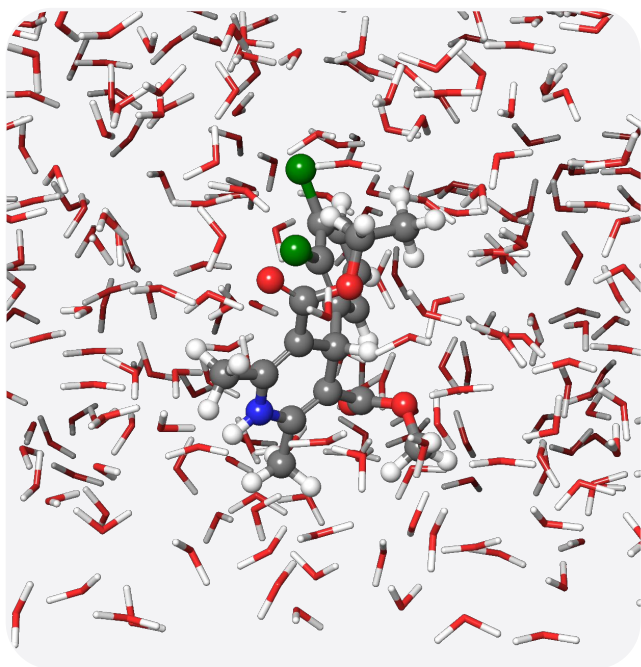


ethyl acetate

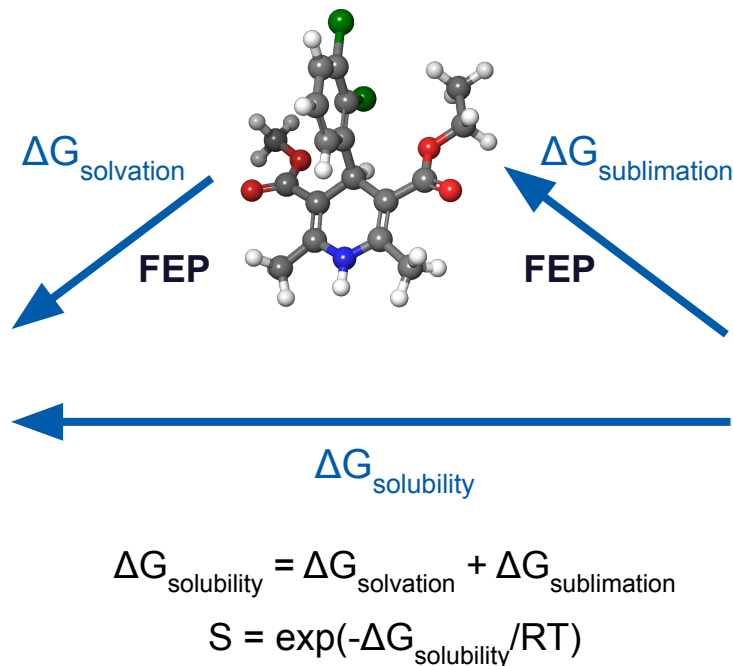
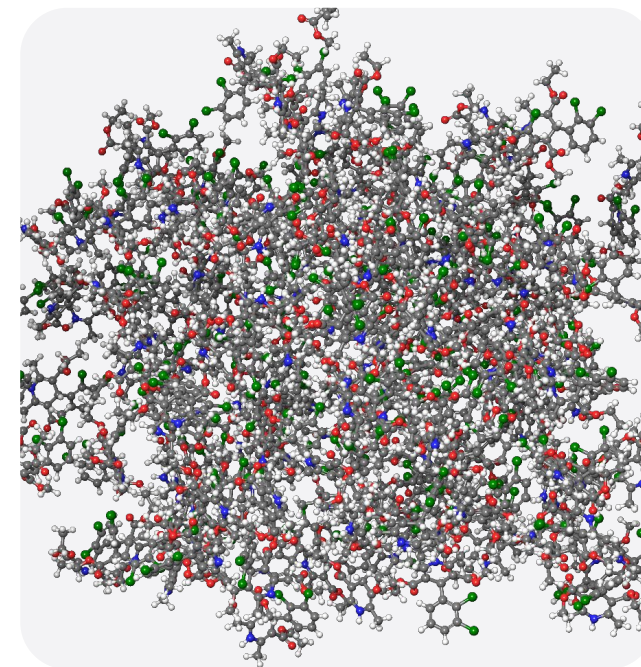
FEP Solubility: amorphous API into water

Calculate the free energy to move molecule from the amorphous into the solvent via constructing a thermodynamic cycle

Solvated (1 M)

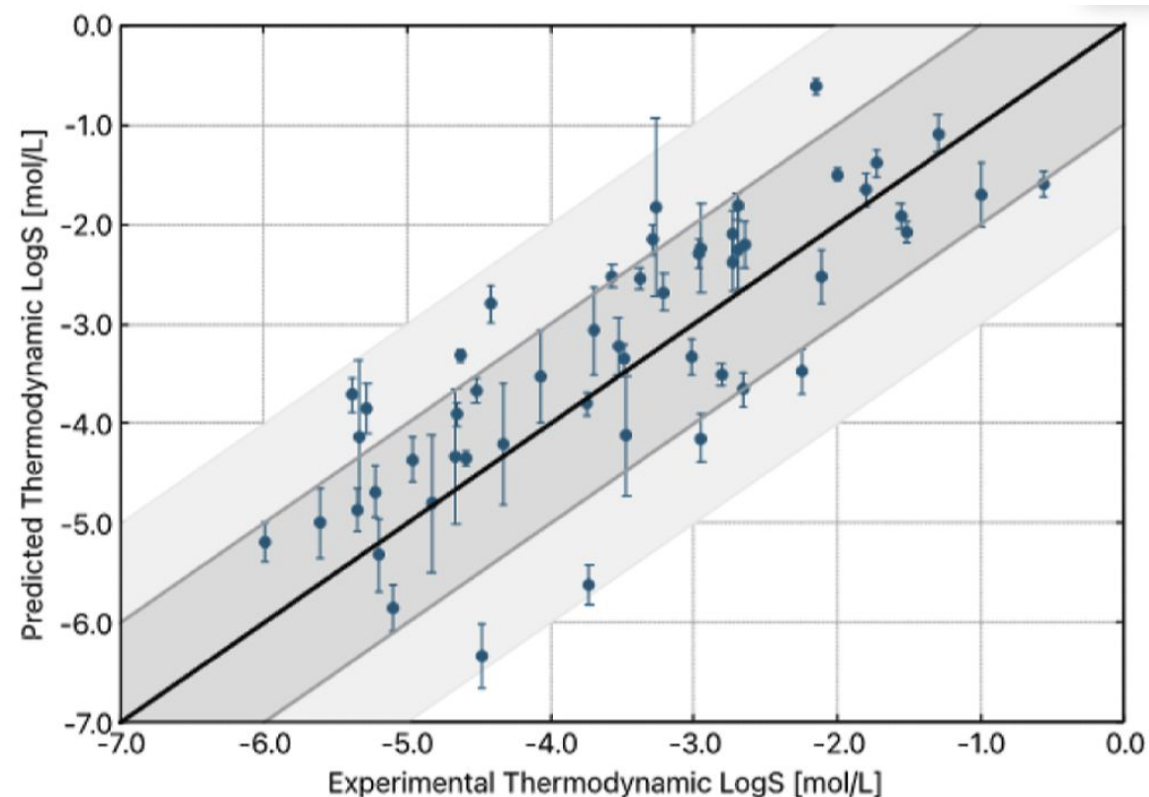


Aggregate of small molecules



FEP Solubility: crystalline API into water

- Dataset includes drug-like molecules from:
 - DLS100 and Sol Challenge datasets (53 data points)
 - AbbVie proprietary compounds (17 datapoints)
- Molecular weight range: 250 – 445 g/mol



FEP Solubility: API into solvent

Input:

- SMILES, 2D structure, .cif, etc., for API and solvents list

Output:

- solvent solubility/water solubility ratio or solute solubility

Timeline:

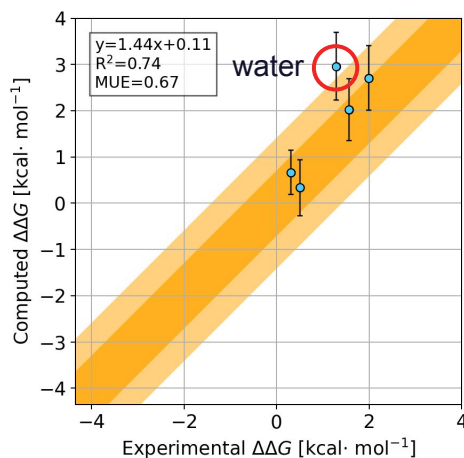
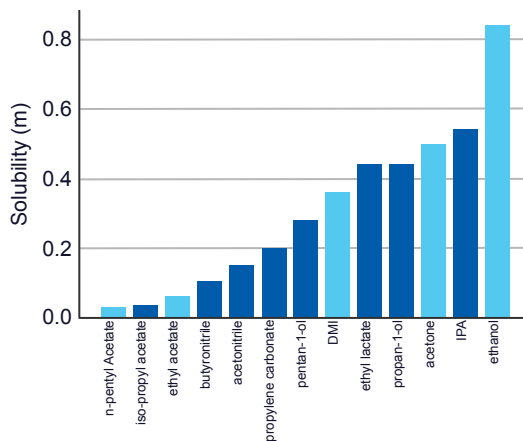
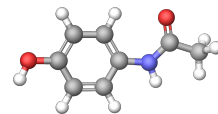
- ~2 days per API/solvent combination

$$\underbrace{-2.303RT \log\left(\frac{S_1}{S_2}\right)}_{\text{Experimental } \Delta\Delta G} = \underbrace{(\Delta G_{\text{solvation},1} - \Delta G_{\text{solvation},2})}_{\text{Computed } \Delta\Delta G}$$

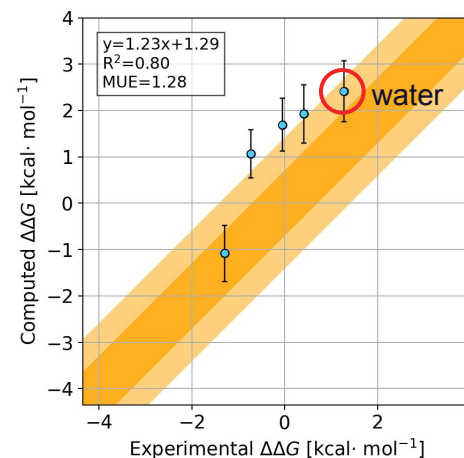
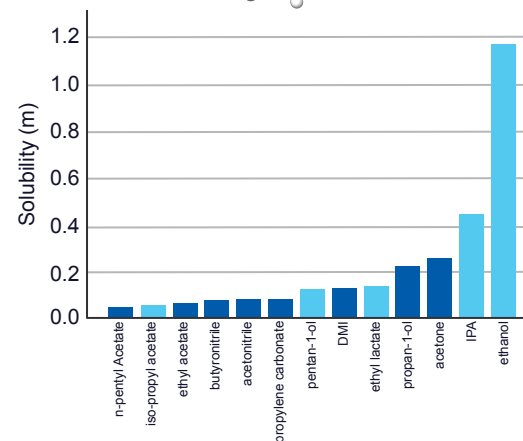
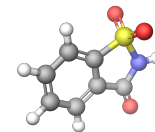
S_2 : experimental aqueous solubility

$\Delta G_{\text{solvation},2}$: solvation free energy of solute in ethanol

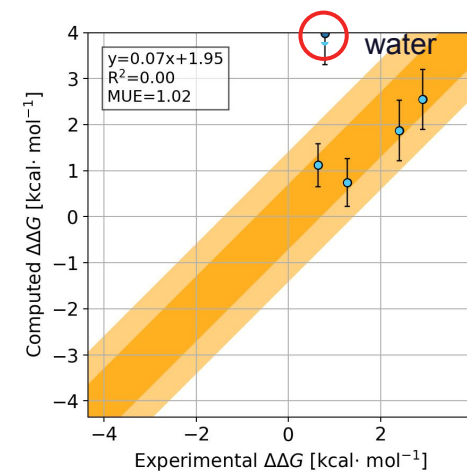
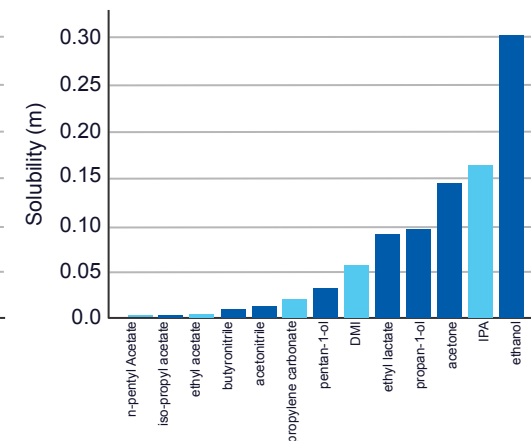
Paracetamol



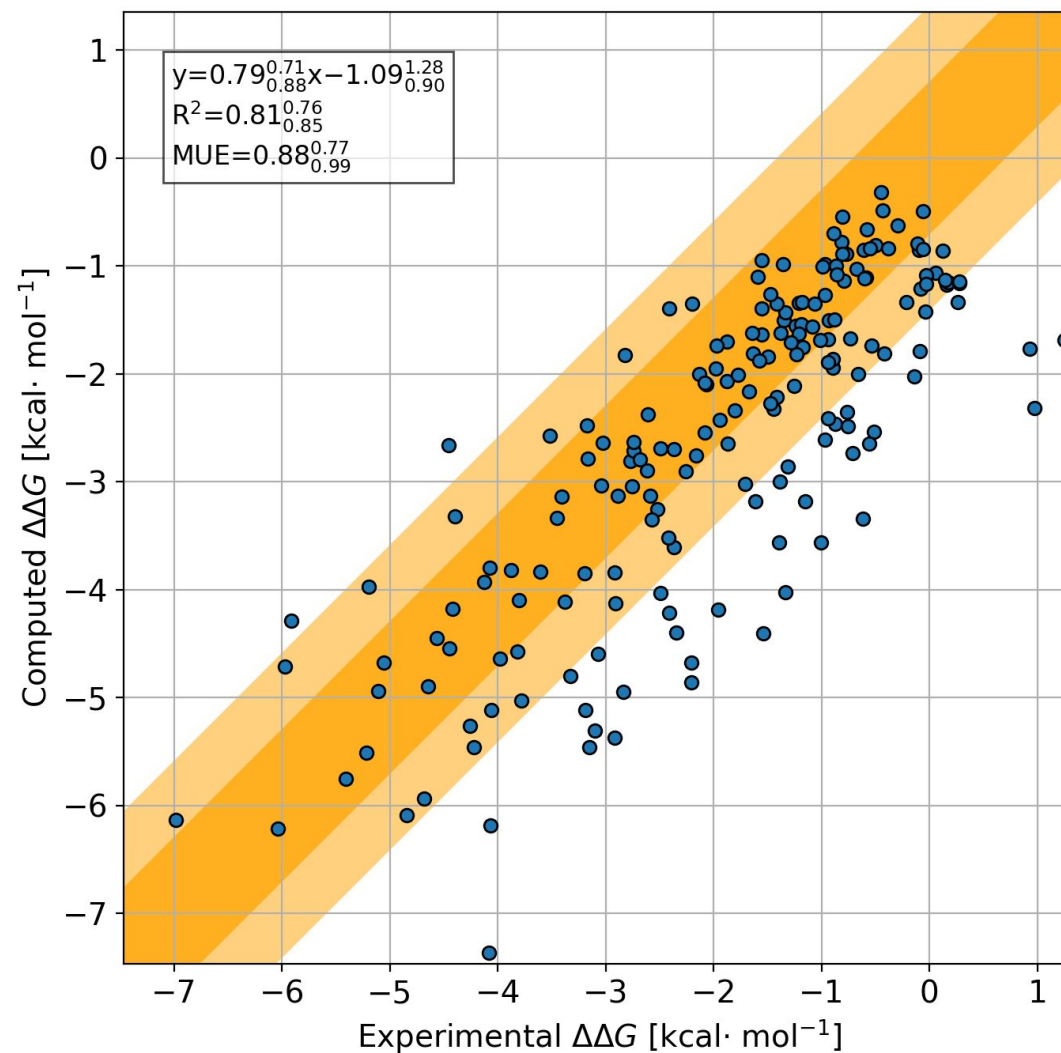
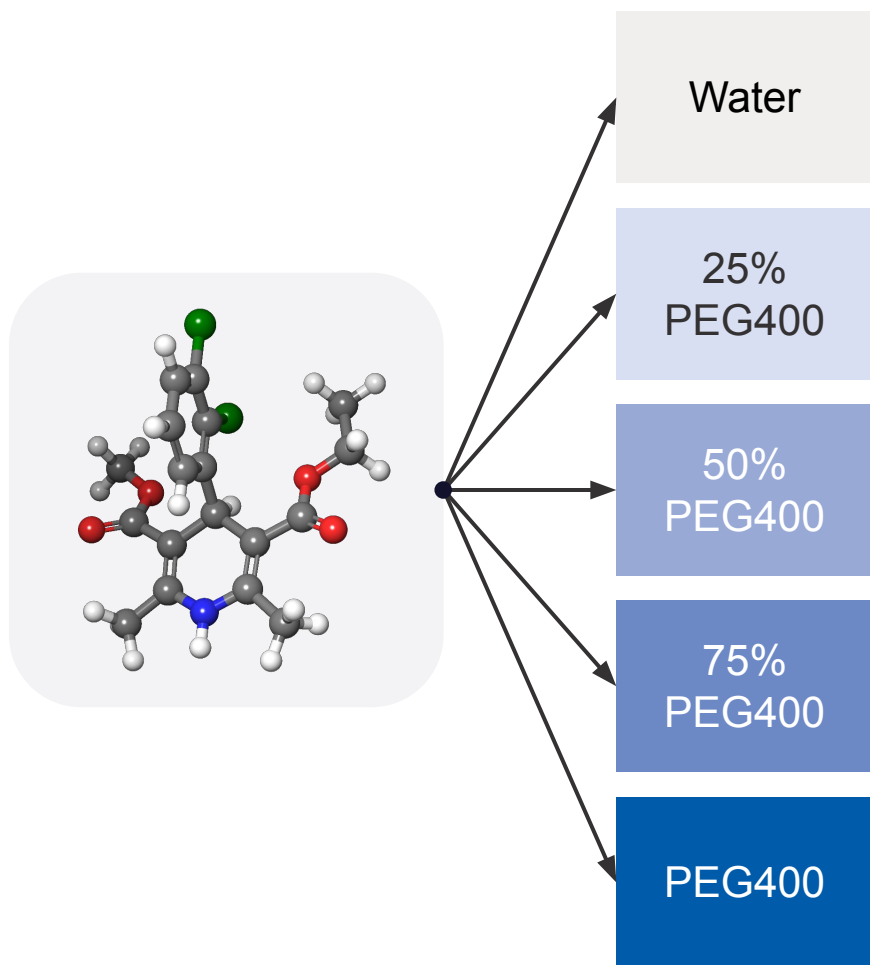
Saccharin



Cimetidine



FEP Solubility: API into excipient



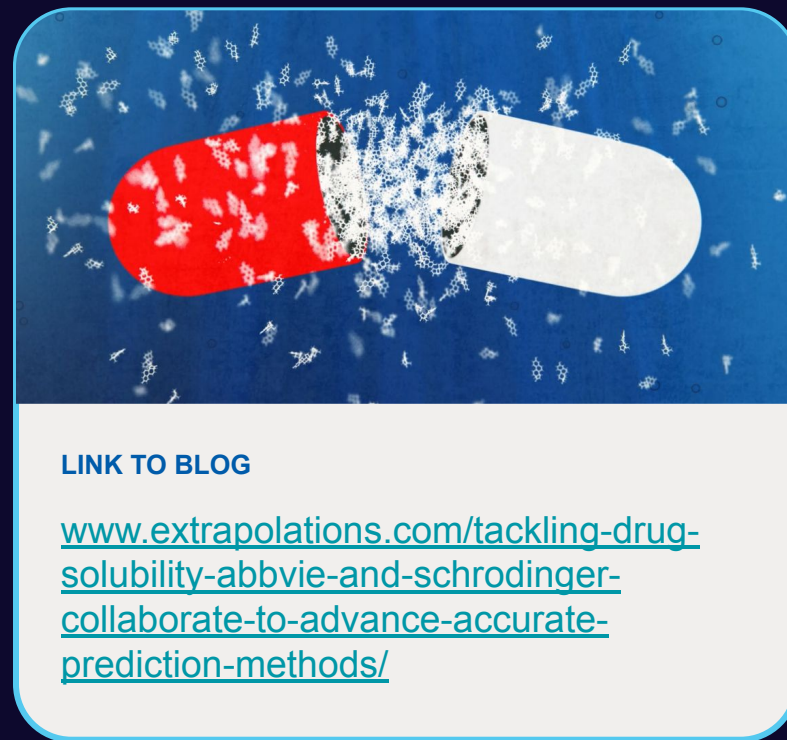
Tackling Drug Solubility: AbbVie and Schrödinger Collaborate to Advance Accurate Prediction Methods



“

Predicting crystalline solubility at an early stage not only enables the identification of potential risks but also aids chemists in prioritizing synthesis and molecular design.

— Richard Hong, AbbVie



LINK TO BLOG

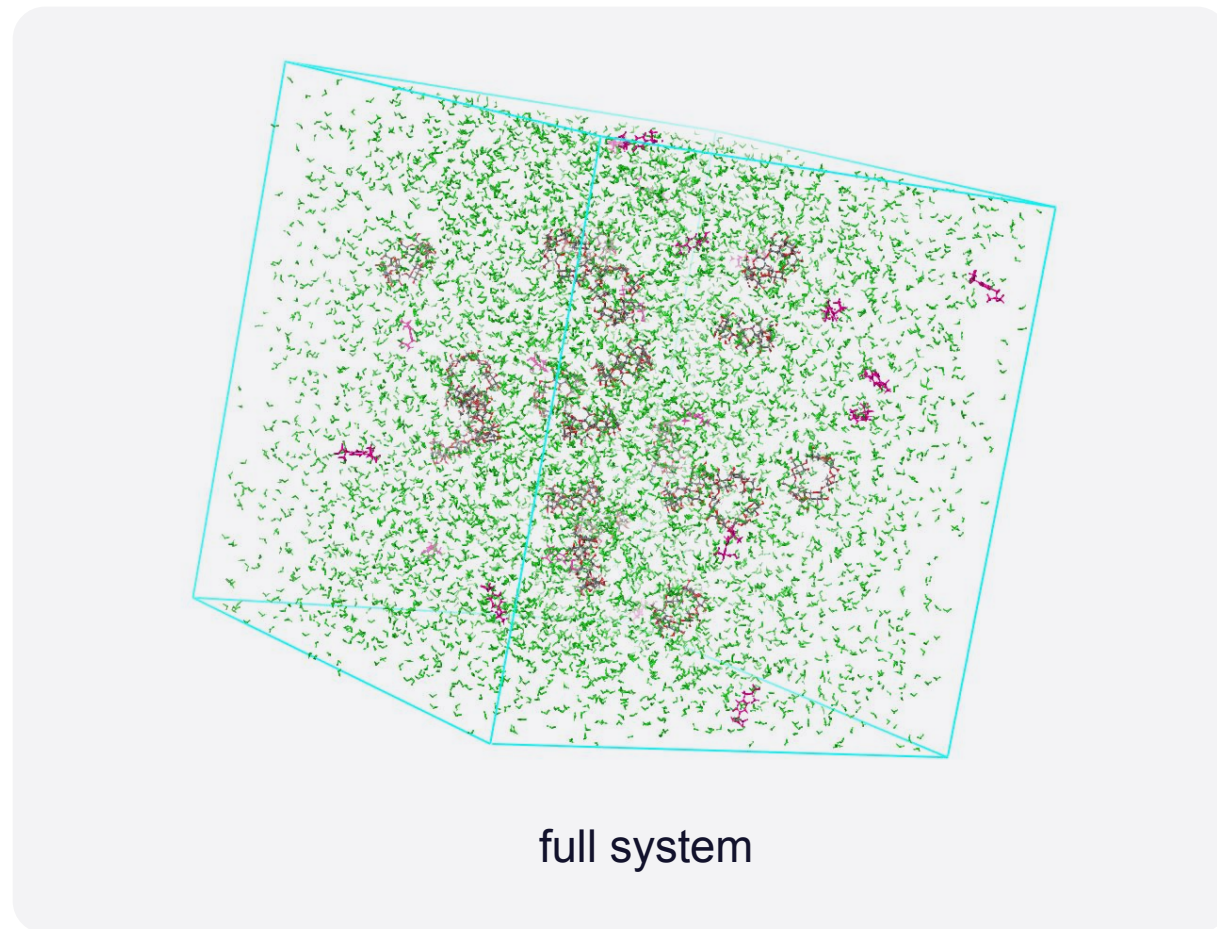
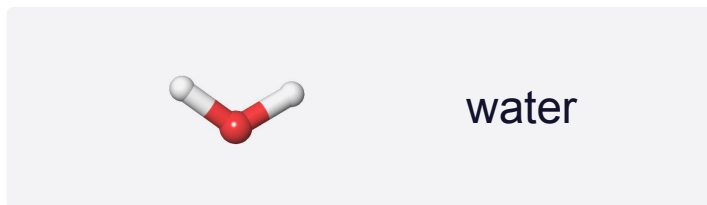
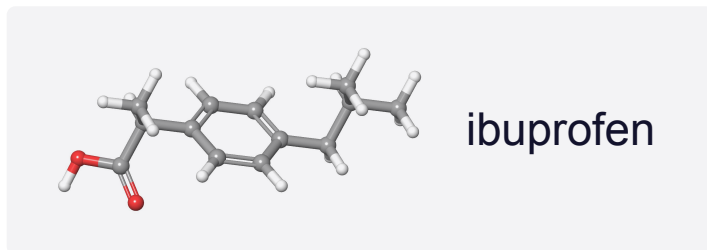
www.extrapolations.com/tackling-drug-solubility-abbvie-and-schrodinger-collaborate-to-advance-accurate-prediction-methods/

LINK TO ARTICLE

<https://pubs.acs.org/doi/10.1021/acs.jmedchem.3c01339>

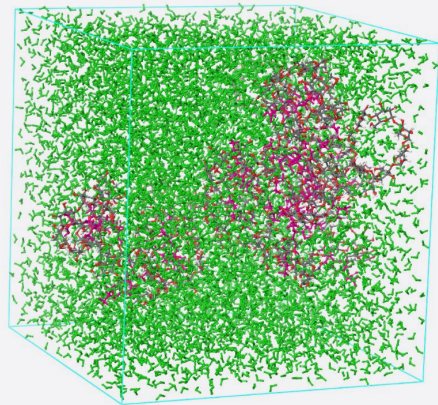
β -cyclodextrin / ibuprofen inclusion complexes

All-atom system building: β CD, ibuprofen and water

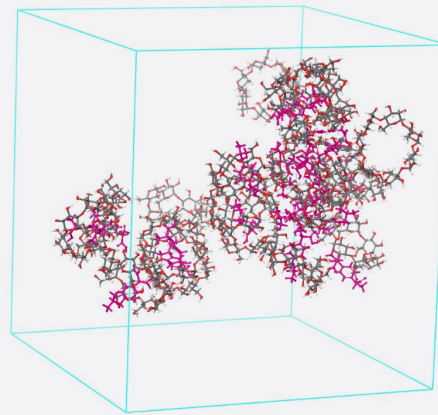


β -cyclodextrin / ibuprofen inclusion complexes

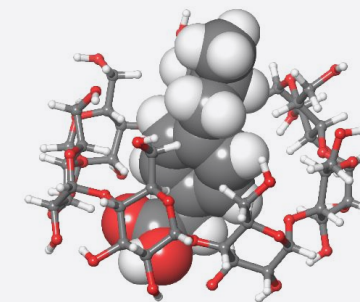
All-atom molecular dynamics simulation of β CD, ibuprofen and water reveals clustering and the formation of inclusion complexes



Equilibrated structure

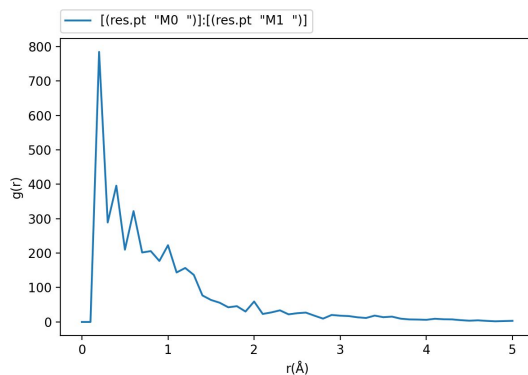


Waters hidden

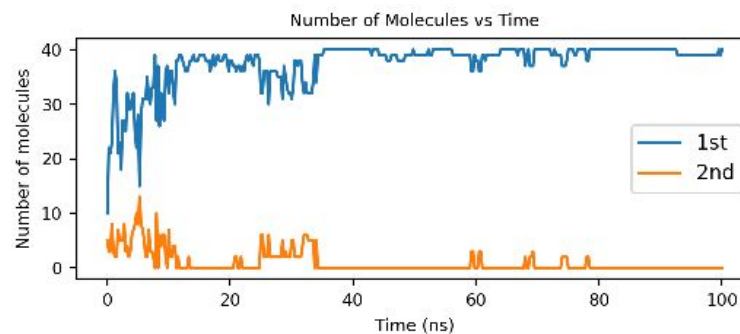


Inclusion complex

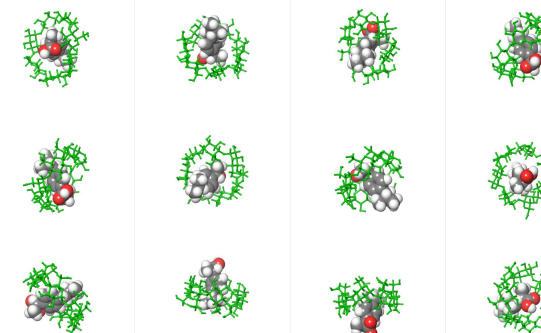
Radial Distribution Function



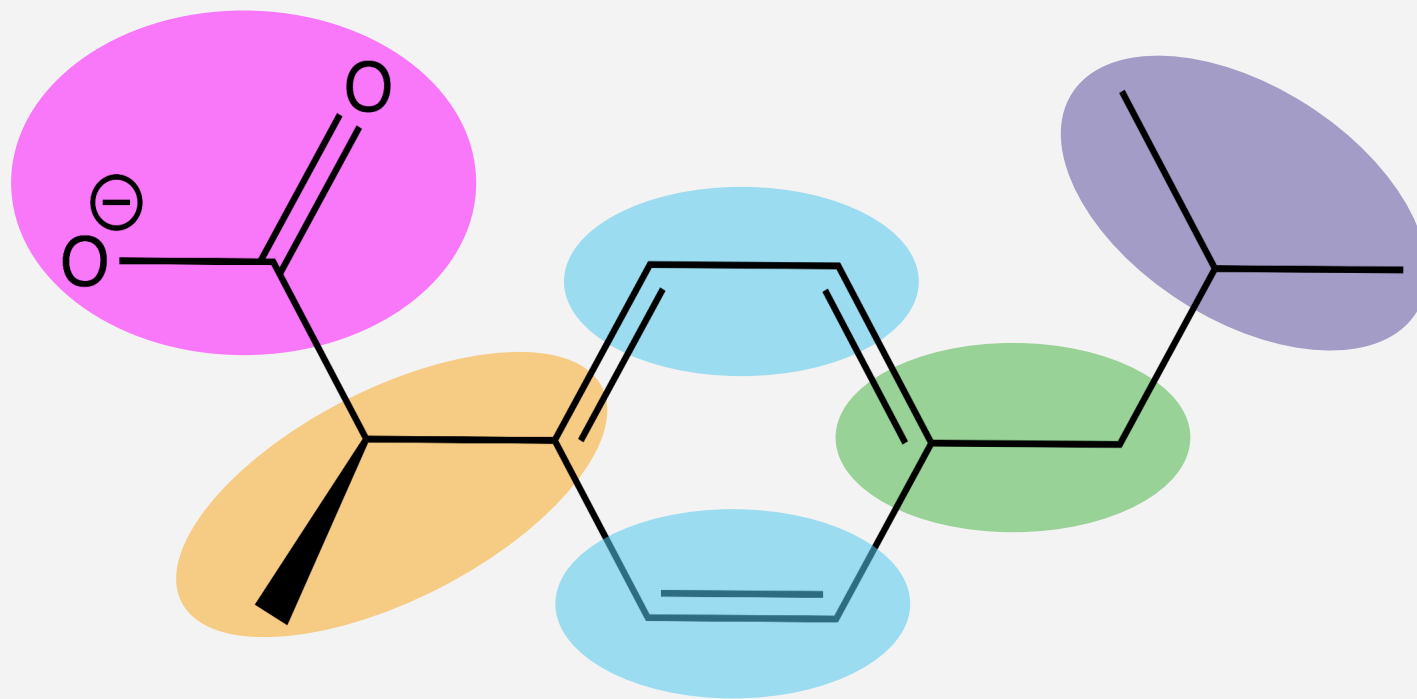
Cluster Analysis



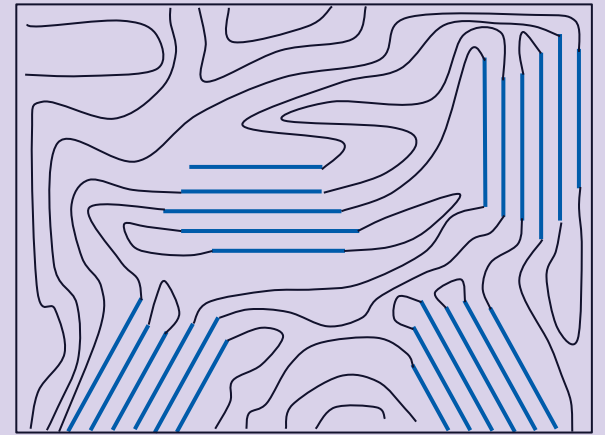
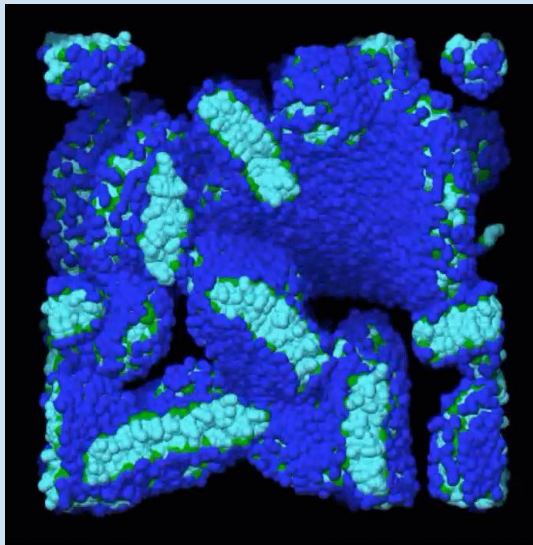
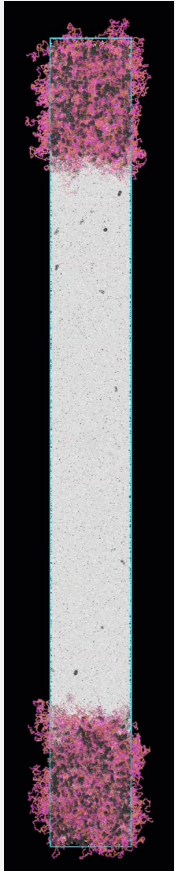
Extract Clusters



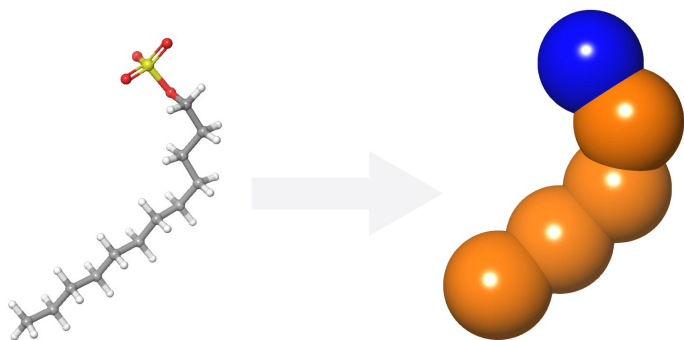
From all-atom to particle representations



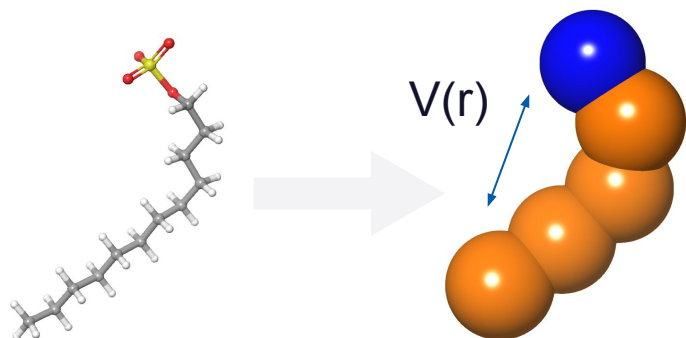
Accessing larger and longer simulations with coarse-graining



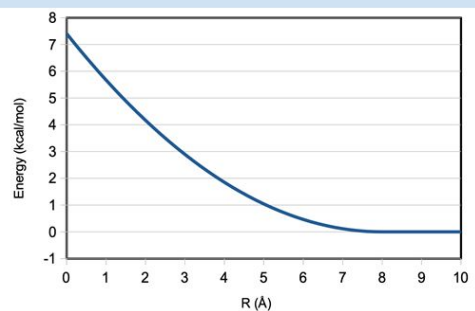
Automated DPD procedure



Automated DPD procedure

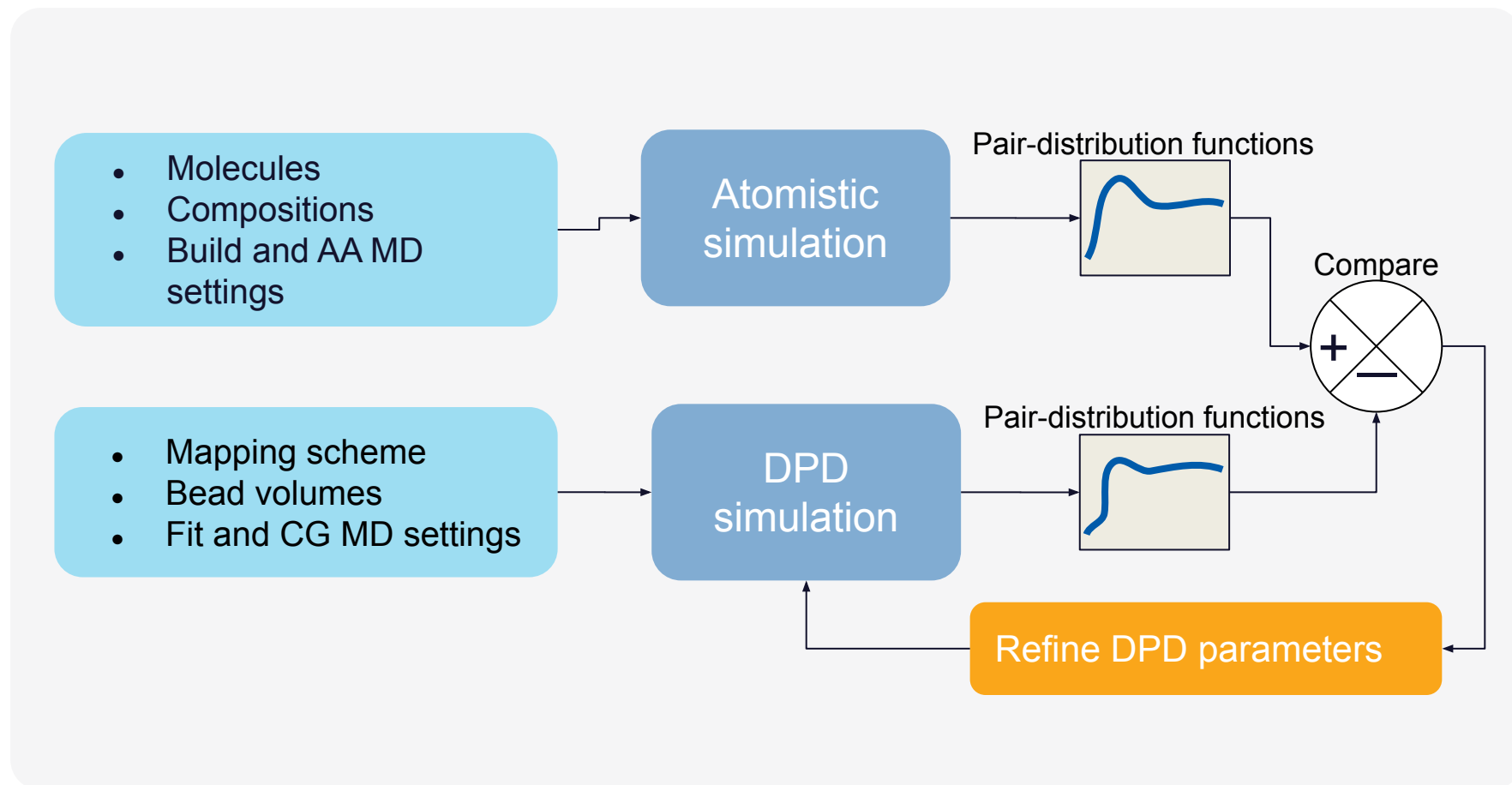
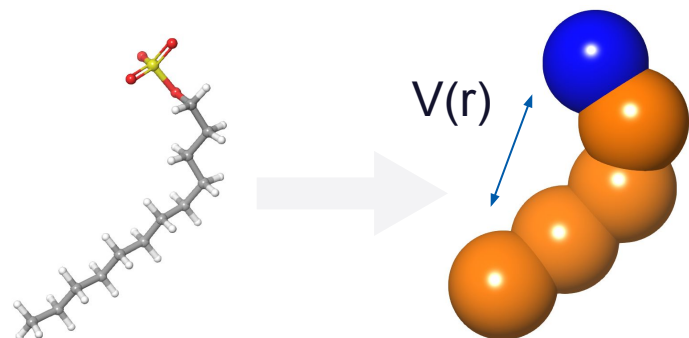


Potential for a soft particle



e.g., Dissipative Particle Dynamics (DPD)

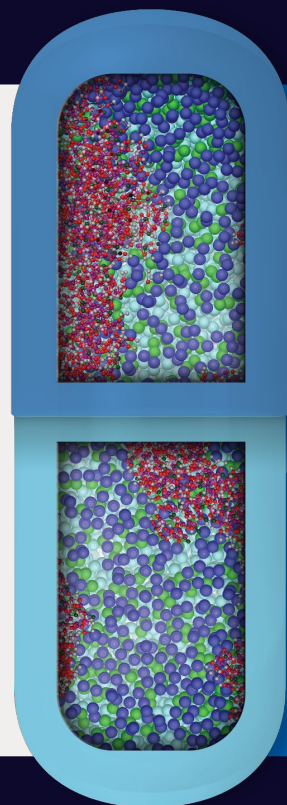
Automated DPD procedure



Optimize drug formulations with digital simulation

CHALLENGE

Developing efficient drug formulations requires a clear understanding of mechanism behind the rate and extent of dissolution of amorphous solid dispersions (ASDs). However, such molecular-level understanding is difficult to obtain through experiments alone.



SOLUTION

AbbVie and Schrödinger used coarse-grained MD to understand dissolution profiles of ASDs. Doing so enabled evaluation of drug/polymer combination dissolution rates, identification of interactions responsible for delayed release, and enabled rational design of new excipients for drug formulations.

Result: Identified the **root cause of unusual release profiles** and enabled rational design of new polymers for desired drug release.



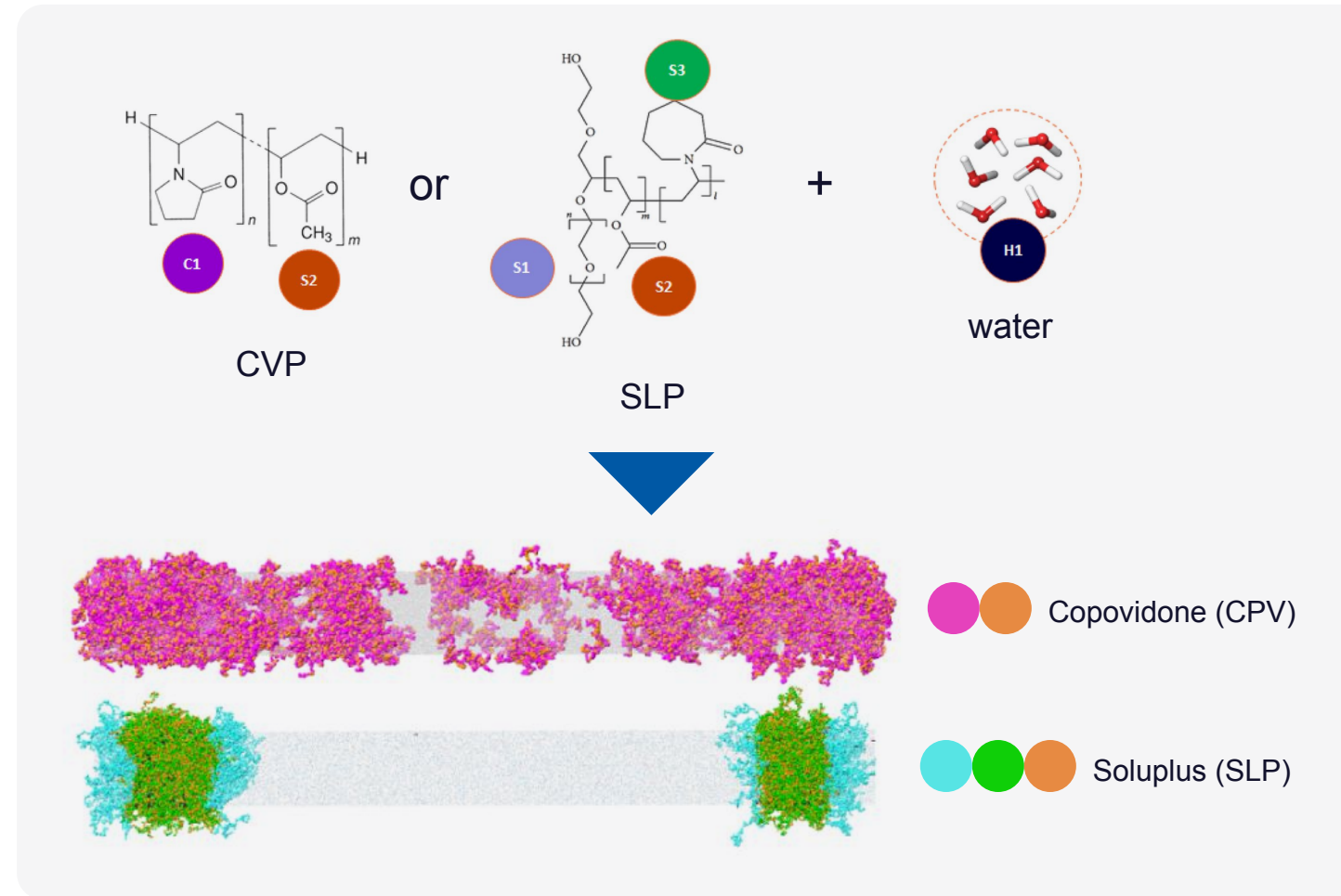
ASD dissolution: API release

Changing polymer in amorphous solid dispersions (ASD) has major impact on dissolution and release profile

Experiments show **water permeates copovidone (CPV) ASD quickly** while **Soluplus (SLP) ASD dissolves slowly**

GOAL

Can we understand the polymer impact on dissolution? If so, formulation scientists can guide ASD selection and screening to best candidates



AbbVie/Schrodinger collaboration involving: Mohammad Atif Faiz Afzal, Kristin Lehmkeper, Ekaterina Sobich, Thomas F. Hughes, David J. Giesen, Teng Zhang, Caroline M. Krauter, Paul Winget, Matthias Degenhardt, Samuel O. Kyeremateng, Andrea R. Browning, John C. Shelley

ASD dissolution: dissolution of pure polymers at 37°C

Experimental observations:

CPV

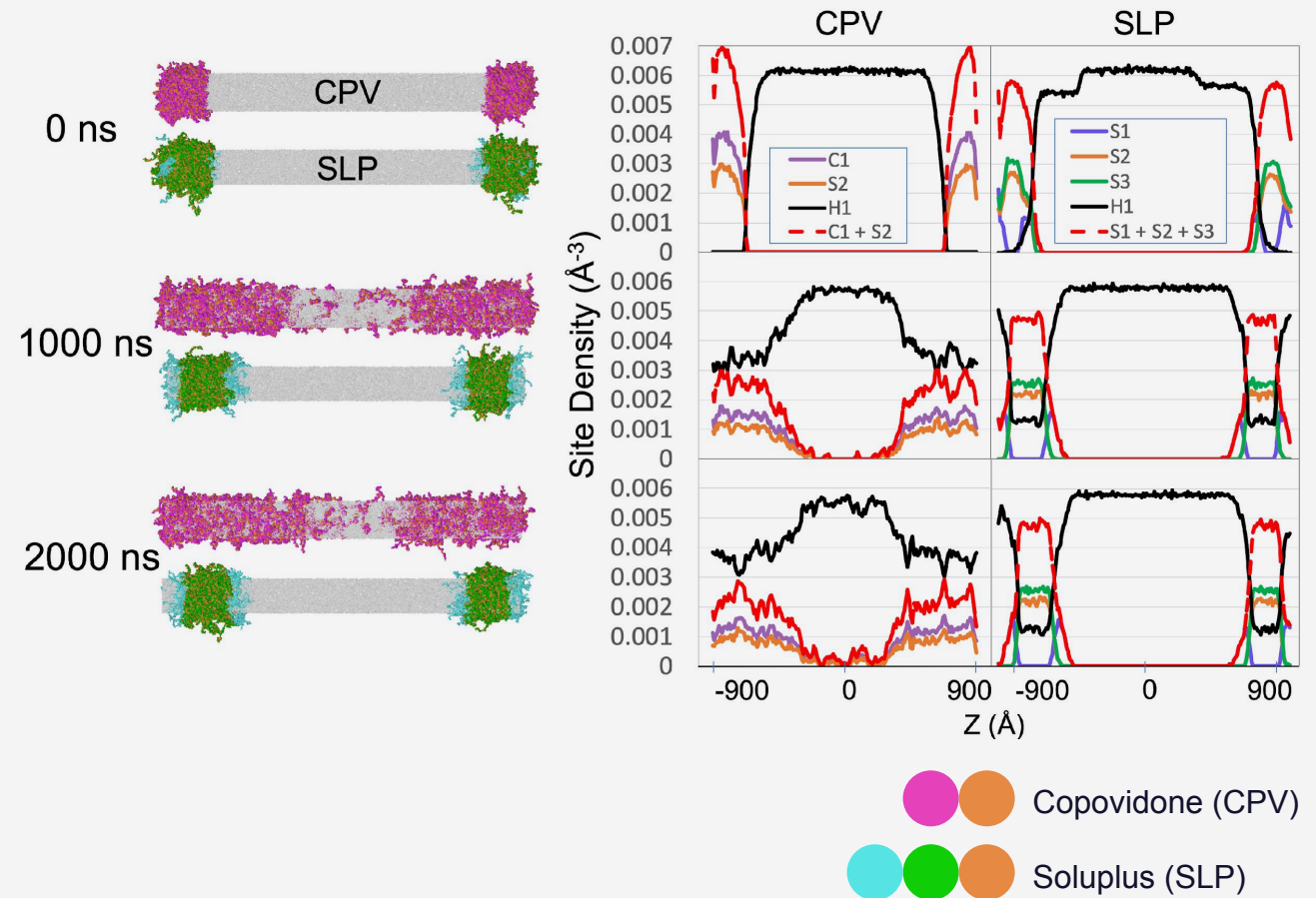
- Water permeates CPV very rapidly
- CPV largely dissolved within 30-60 min

SLP

- SLP dissolves more slowly and with a noticeable hydration front
- Some SLP remains undissolved at 60 min

SLP interfacial structures slow down water penetration

DPD simulation results



ASD dissolution: protonated acid API at low pH

Experimental observations:

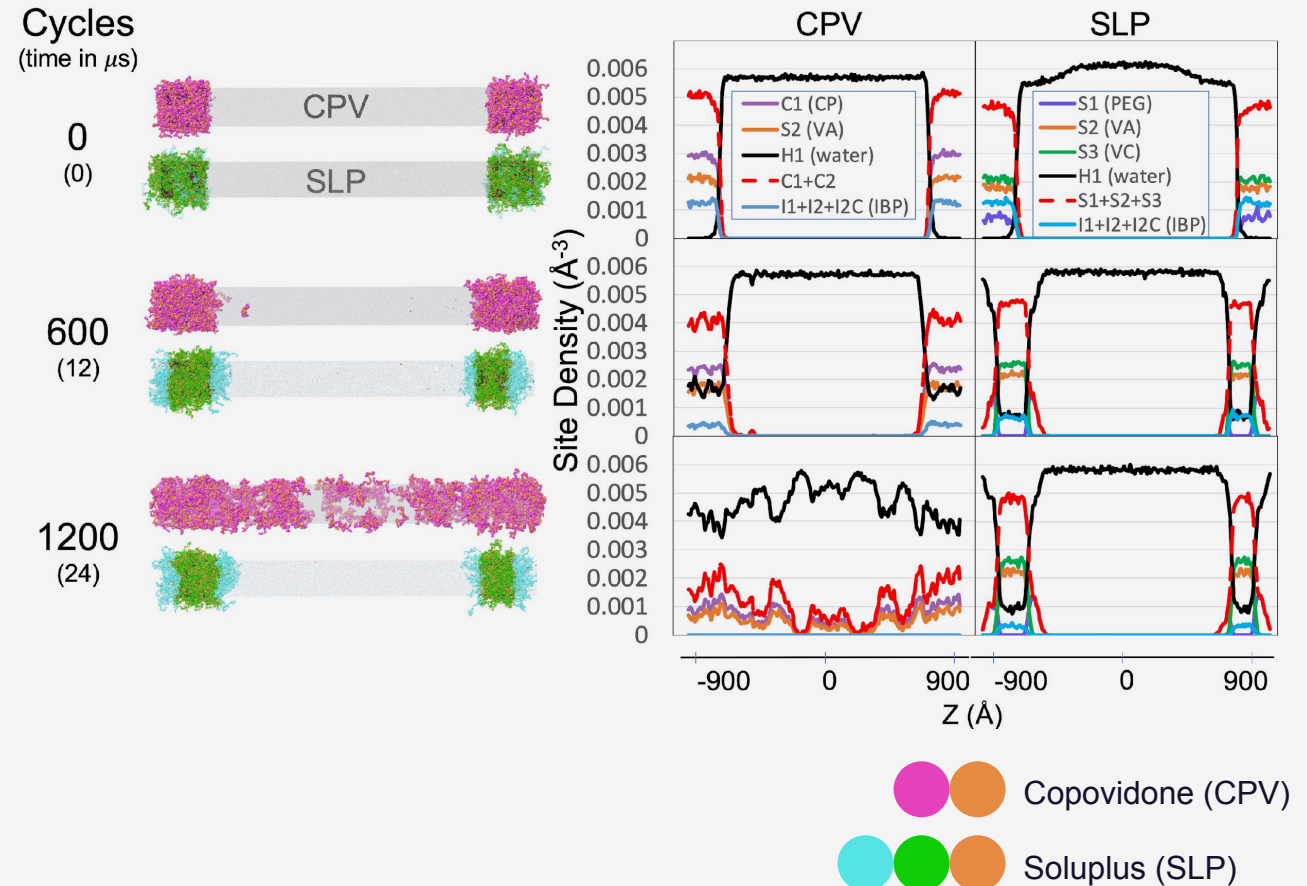
CPV

- Complete water penetration after 20 minutes
- Potential, mostly transient, precipitation of drug after 5 minutes
- Drug and polymer still detectable at 30 minutes

SLP

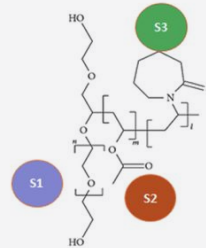
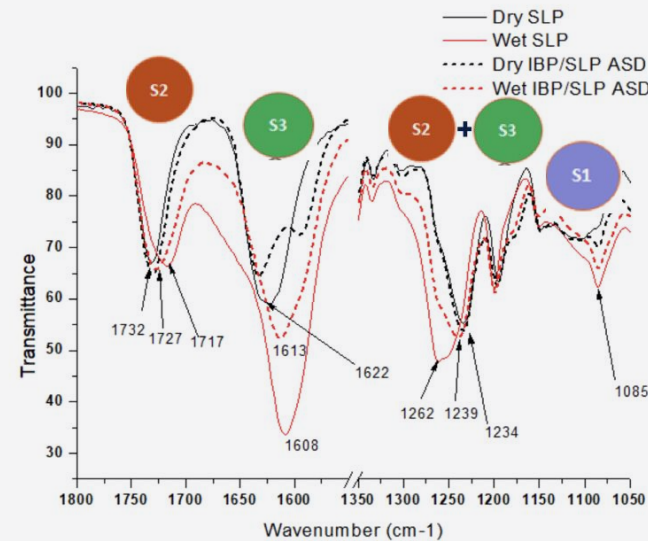
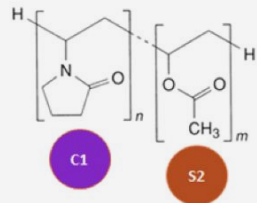
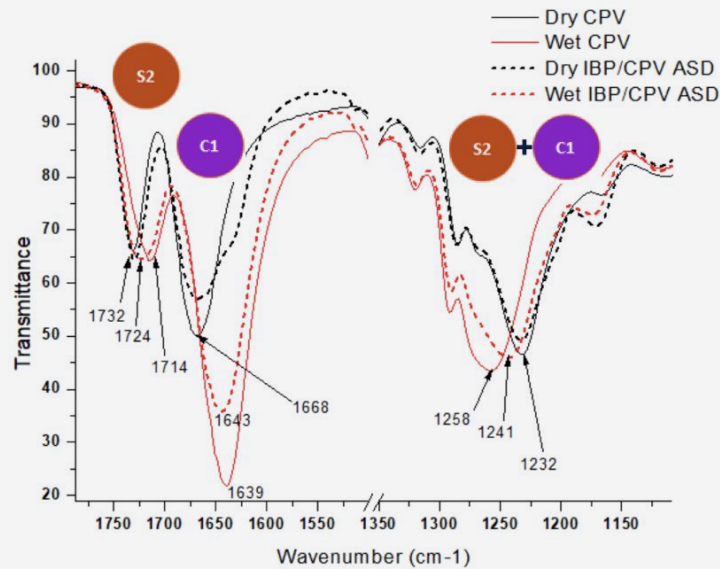
- Essentially no water ingress
- Essentially no dissolution
- Drug and polymer signals largely unchanged over 30 minutes

DPD simulation results



ASD dissolution: hydrogen bonding in the solid state

IR spectra



Hydrogen bonding

CPV
(C=O, C1)

SLP
(C=O, S3)

Acid-vinylpyrrolidone < Acid-caprolactam
Water-vinylpyrrolidone > Water-caprolactam



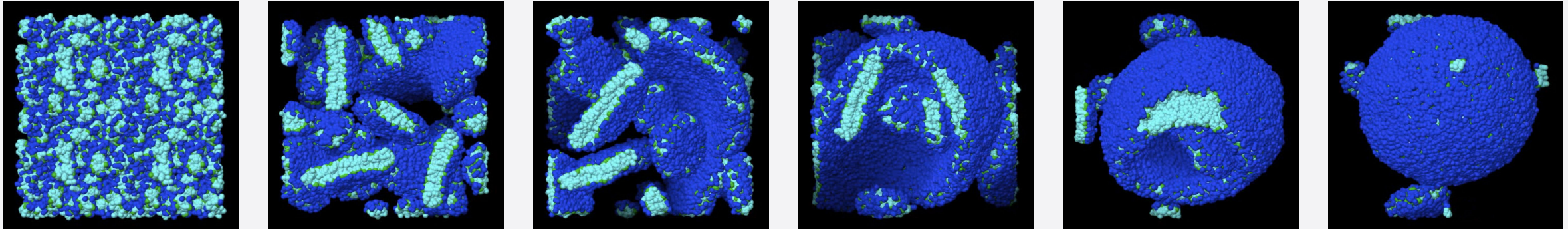
DPD simulations in solution show:

- Greater number of Acid-S3 vs Acid-C1 interactions
- Greater hydration of C1 vs S3

Structuring at in SLP may also be a factor

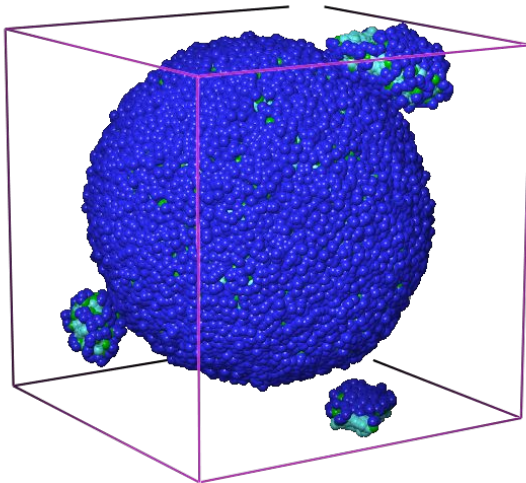
Self assembly of pure DPPC into a liposome

Self assembly from a random distribution of DPPC in water



Simulation time

self-assembled
liposome



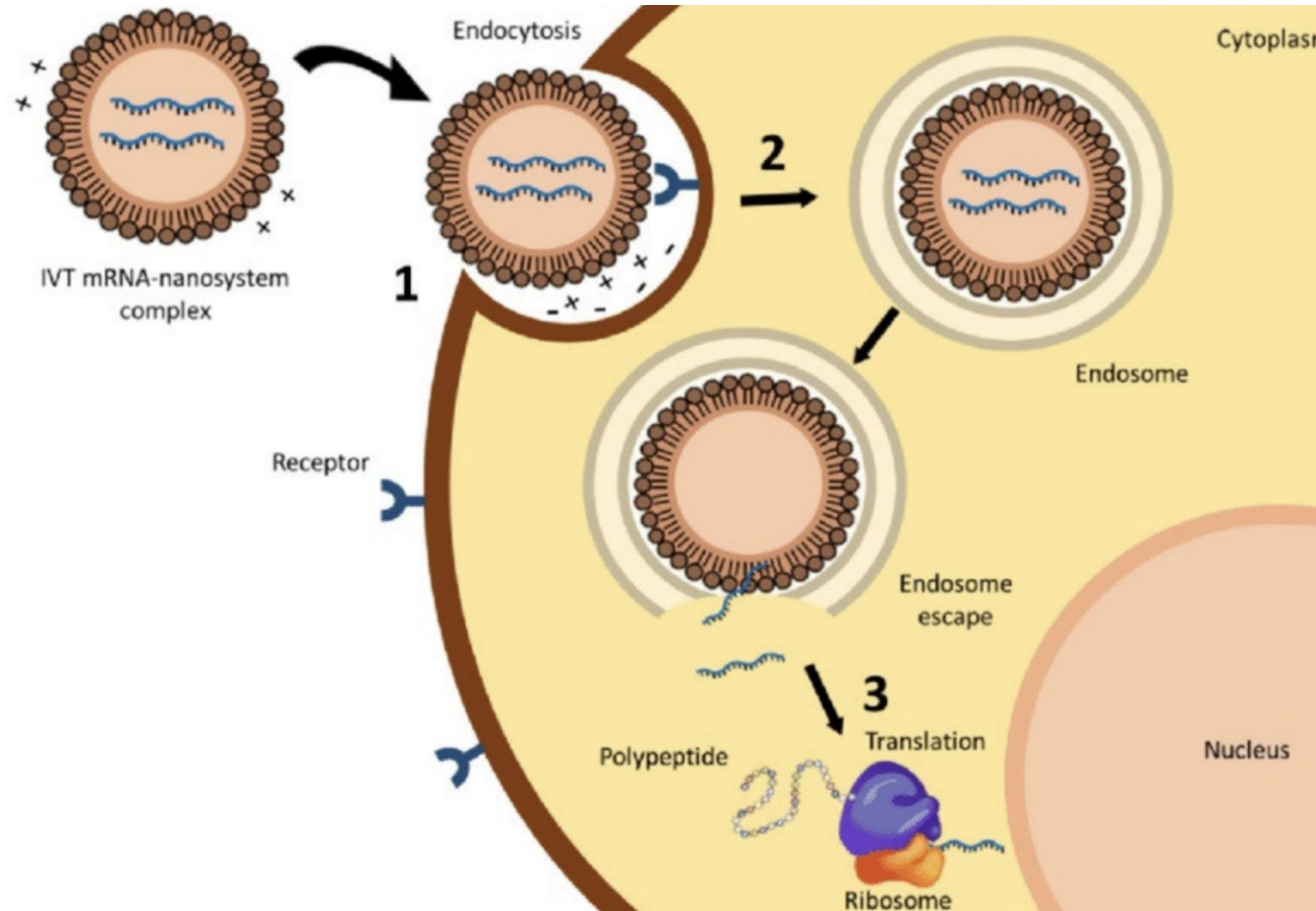
System contains 546,832 coarse-grained particles with 10,496 DPPC molecules (representing >5,000,000 atoms; 400 Å simulation box size)

■ lipid head groups
■ lipid tails
water not shown

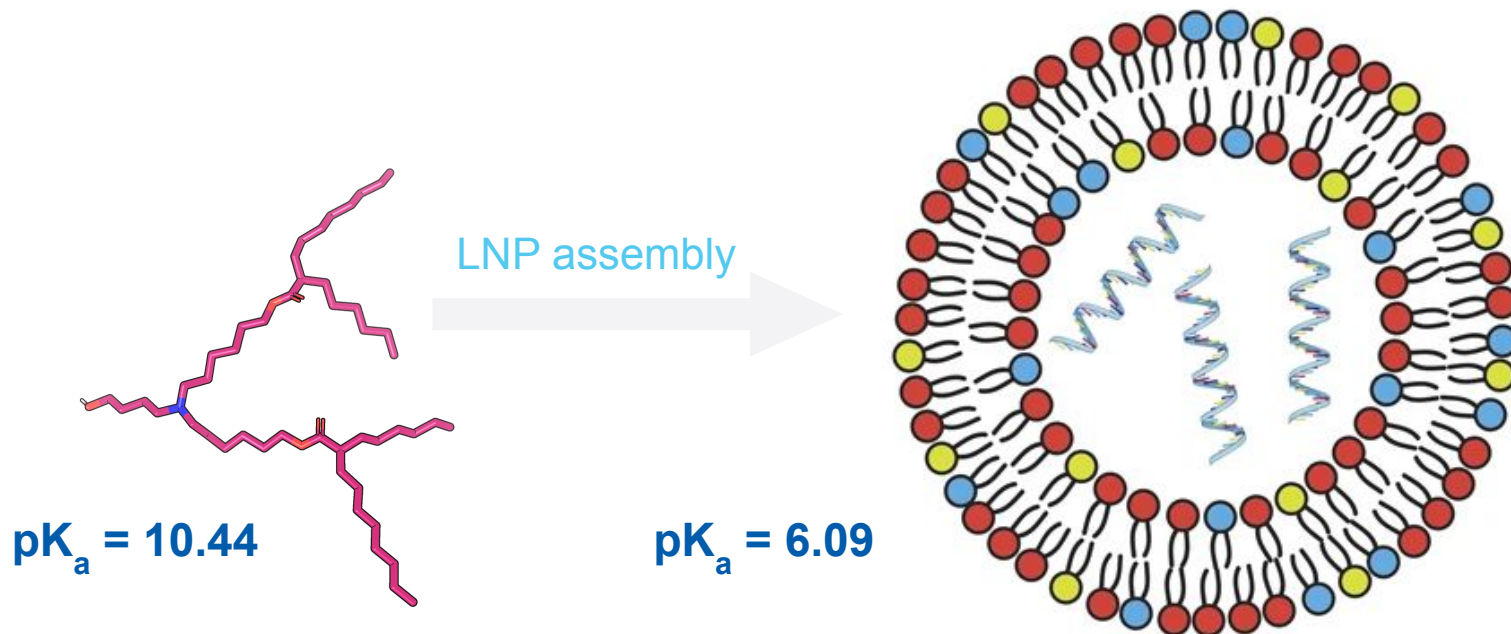
Biologics formulation

RNA Technology

- LNP protects cargo from cleavage/biochemical mutation
- After transfection, pH shift within the endosome will cause protonation of more ionizable lipids
- More charged lipids associate with negative lipids from endosome and degrade endosome
- mRNA cargo must escape endosome and begin translation



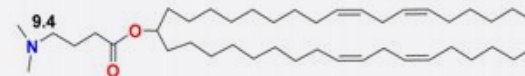
Apparent pK_a



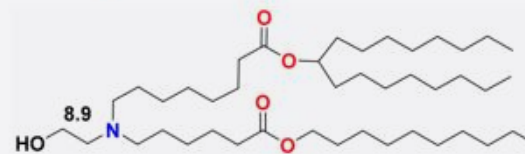
- Defined shift from solution single-molecule pK_a values
- molar ratio of components within bilayer influences pK_a
- It represents the pH where ionized = neutral species

All ionizable lipids in use have apparent pK_a values within one unit

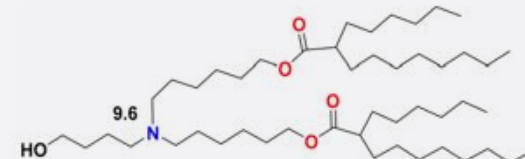
A. Dlin-MC3-DMA:
 pK_a : 6.44



B. SM-102:
 pK_a : 6.75

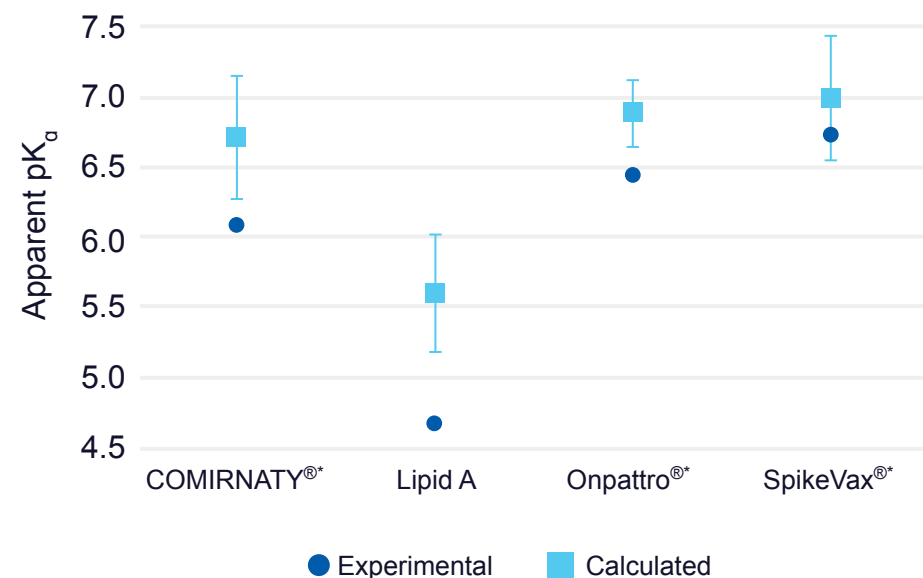
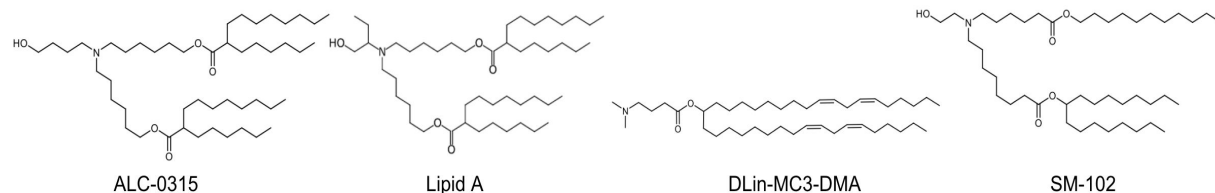


C. Alc-0315:
 pK_a : 6.09



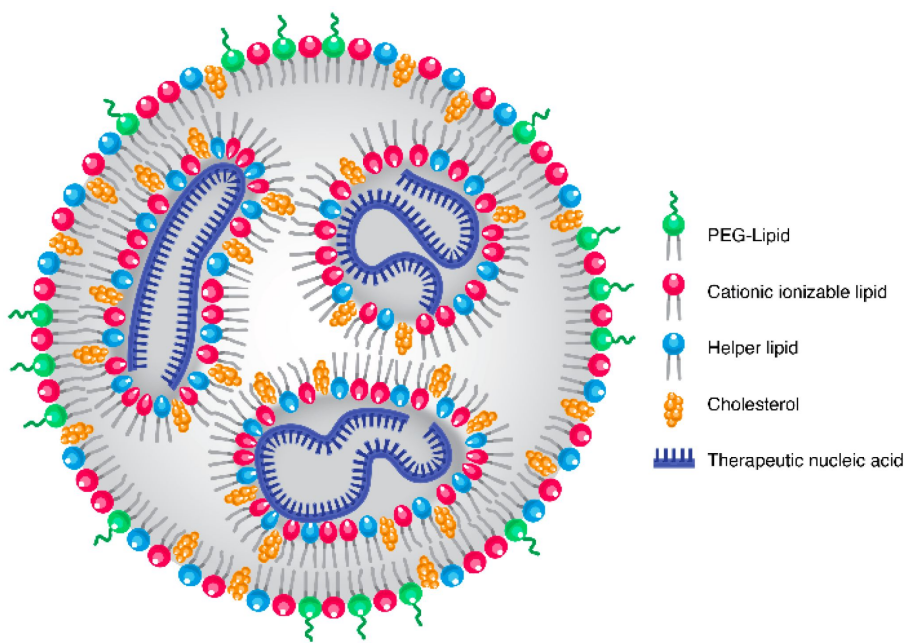
Apparent pK_a values of ionizable lipids in LNPs

COMIRNATY®*		Lipid A		Onpattro®*		SpikeVax®*	
Lipid	Mol %	Lipid	Mol %	Lipid	Mol %	Lipid	Mol %
ALC-0315	48	Lipid A	47.1	Din-MC3-DMA	49	SM-102	49
CHOL	42	CHOL	43.1	CHOL	40.8	CHOL	40.8
DSPC	10	DSPC	9.8	DSPC	10.2	DSPC	10.2



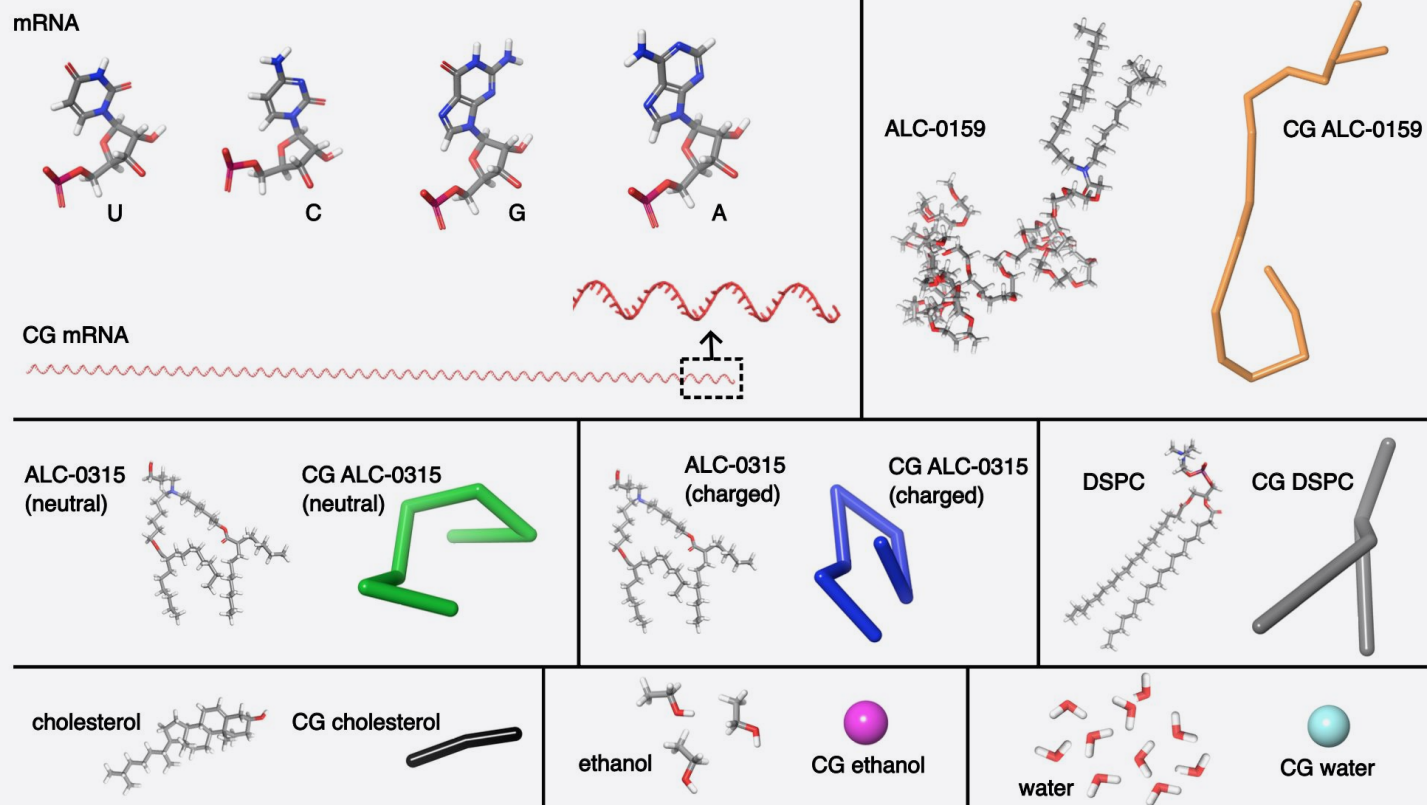
- Apparent pK_a values are 1.5 to 3.5 units lower than intrinsic pK_a values
- Calculated apparent pK_a values are highly correlated with experimental values

Modelling the assembly of RNA-loaded nanoparticles



Kularatne et al., *Pharmaceutics*, 15, 897 (2022)

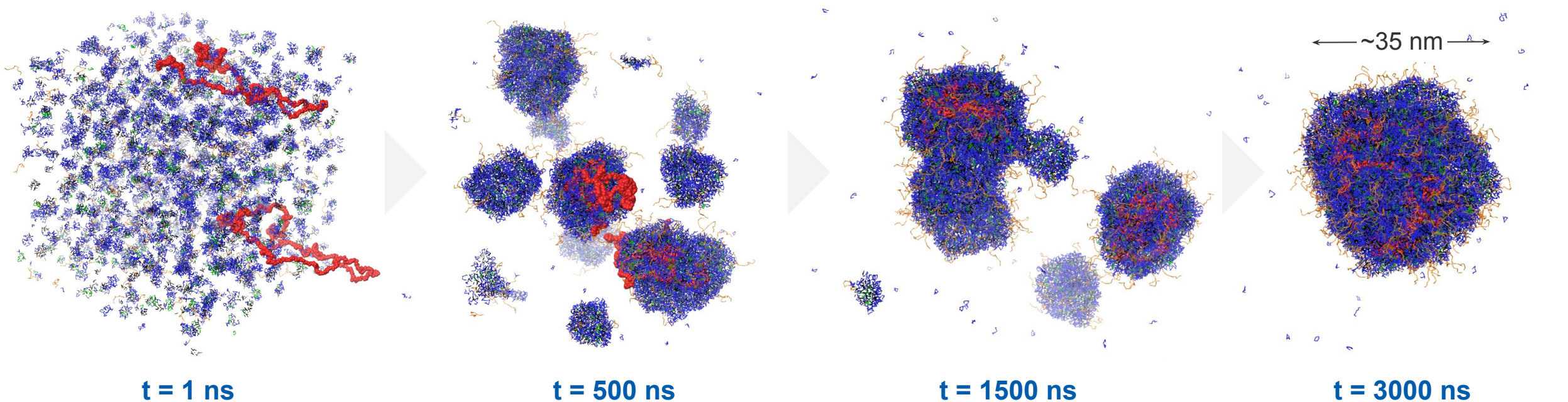
Model System for Pfizer-BioNTech COVID-19 Vaccine



All atom system (~72,000 atoms, simulated for 1.5 μ s) mapped into CG
 1 particle = 10 heavy atoms

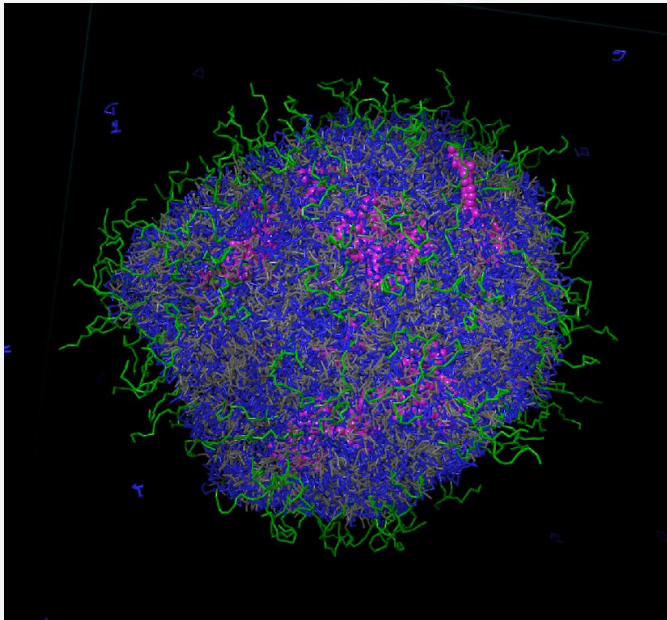
RNA-loaded LNP self-assembly (low pH)

- Ionizable lipid is mostly (90%) protonated
- LNP self-assembles from homogeneous mixture in $\sim 3 \mu\text{s}$
- LNP has the correct characteristics:
 - Roughly 35 nm in diameter
 - mRNA encapsulated in aqueous channels lined with ionizable lipid
 - PEGylated lipid localized on the LNP surface



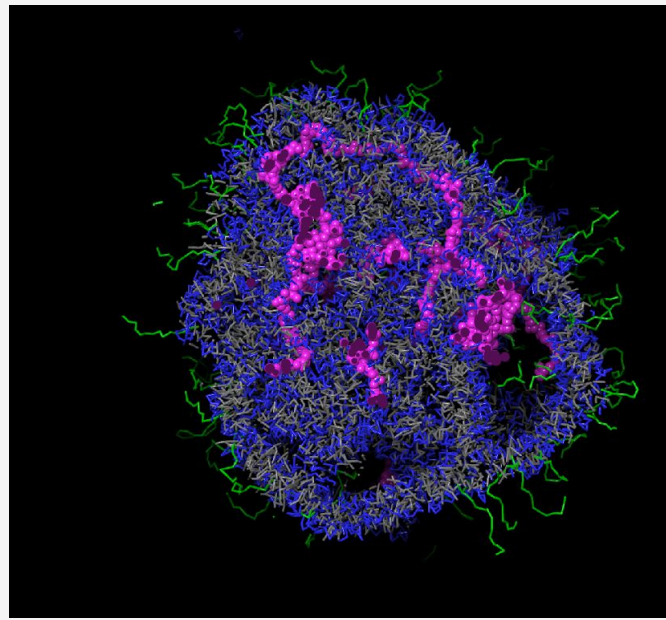
RNA-loaded LNP self-assembly - effect of pH

Full LNP

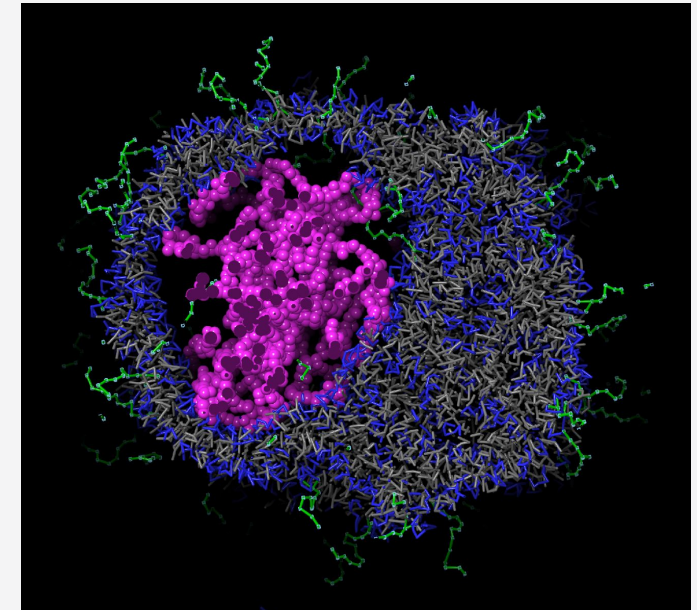


50 nm

Cross section



Bleb formation after
shift to higher pH

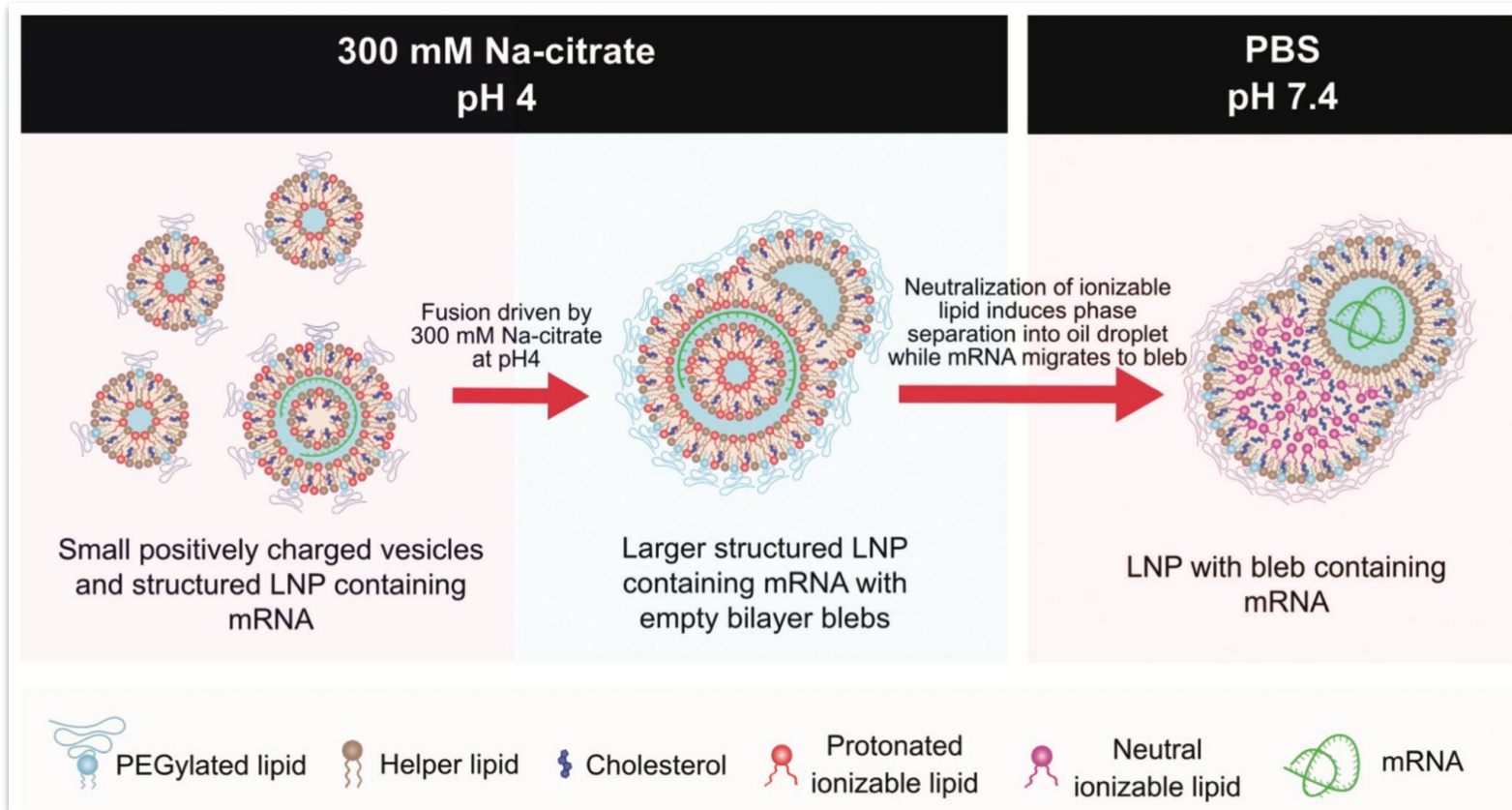


Self assembly from a random aqueous solution in about 3 μ s

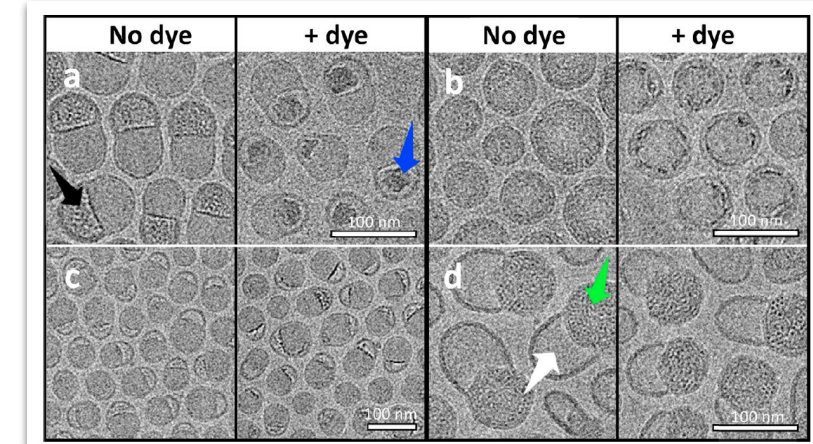
- | | |
|--------------------|---------------------|
| ■ Ionizable lipids | ■ PEG lipids |
| ■ mRNA | ■ Cholesterol, DSPC |

RNA-loaded LNP self-assembly - effect of pH

Circulation occurs at physiological pH (ionizable lipid largely deprotonates). **Bleb formation at physiological pH has been observed!**



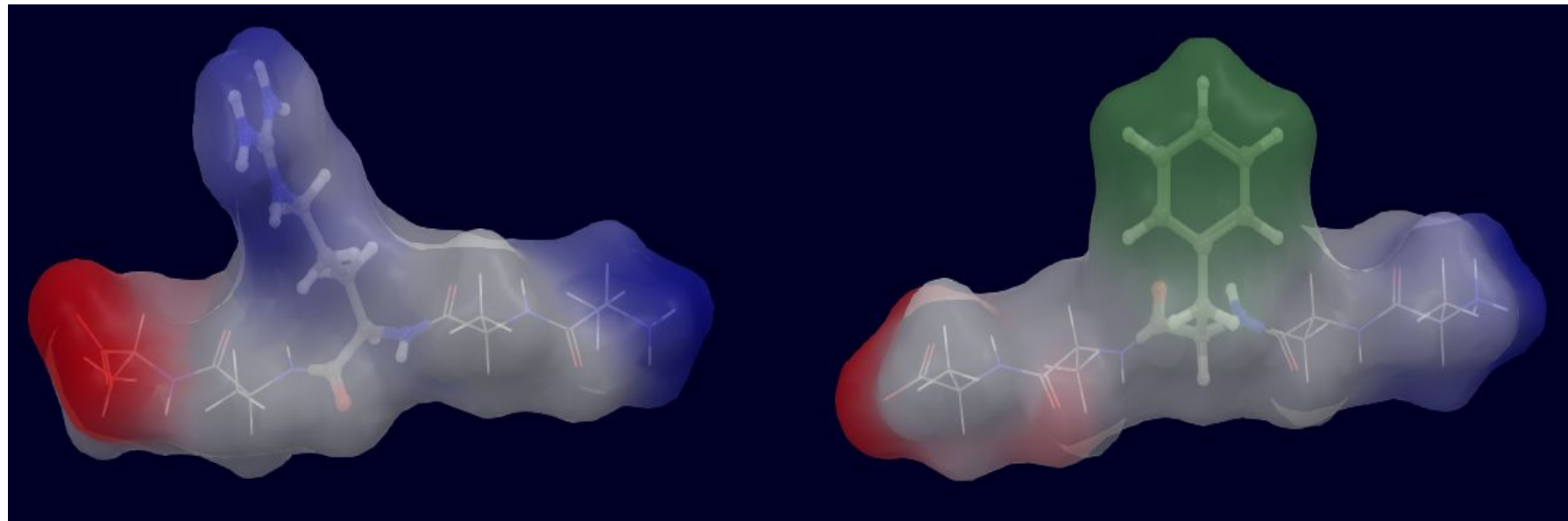
Cheng *et al.*, *Adv. Mater.*, 3, 2303370 (2023)



Brader *et al.*, *Biophysical Journal* 120, 2766 (2021)

Aggregation prediction: AggScore/Surface patch analyzer

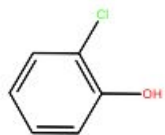
- Aggregation is a complex and not well understood process
 - Irreversible self-association
- Hydrophobic and electrostatic surface properties may all contribute to aggregation
- We can project the contribution of hydrophobicity and electrostatics to the protein surface



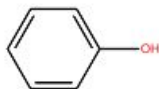
- Positive patch
- Negative patch
- Hydrophobic patch

Exploratory study: aggregation of recombinant human growth hormone (hGH) induced by phenolic compounds

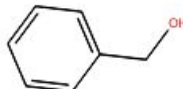
- Maa *et al.* *International Journal of Pharmceutics* 1996, 140:155-168
- The authors looked at insoluble and soluble aggregation of a number of phenolic compounds
- Idea: Use MXMD and AggScore to estimate the aggregation propensity induced by the excipients



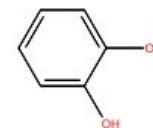
2-Cholophenol



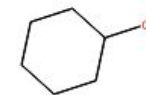
Phenol



Benzyl Alcohol



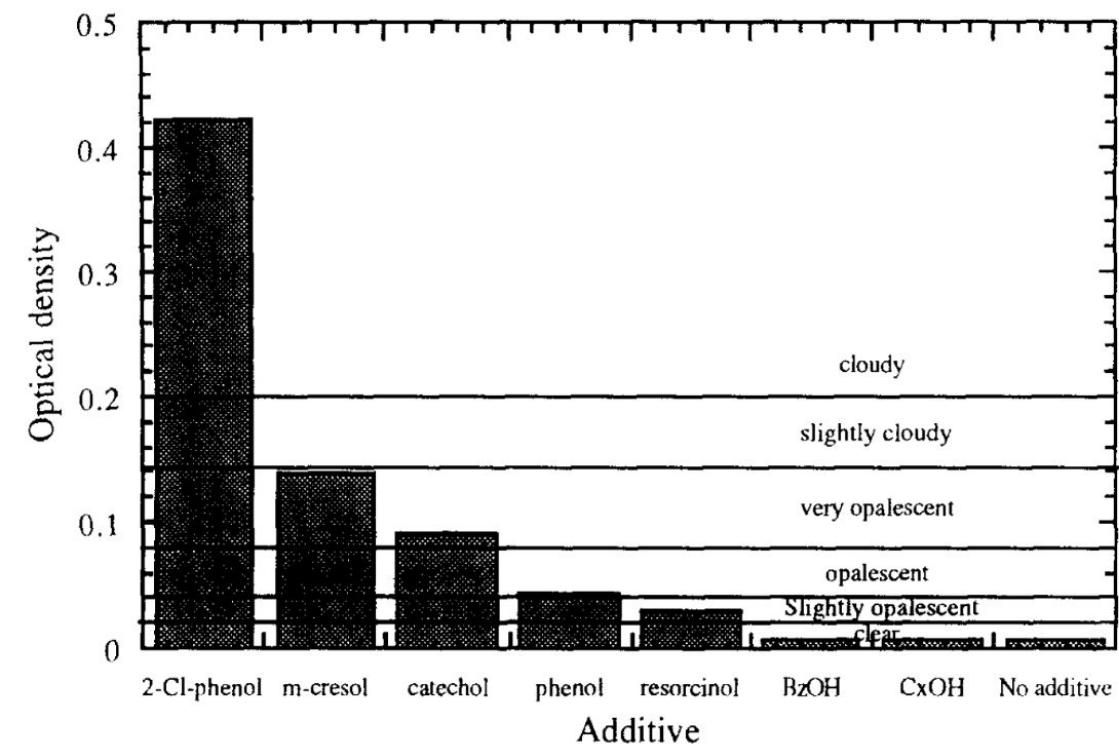
Catechol



Cyclohexanol

Series of compounds are quite similar

Experimental aggregation data

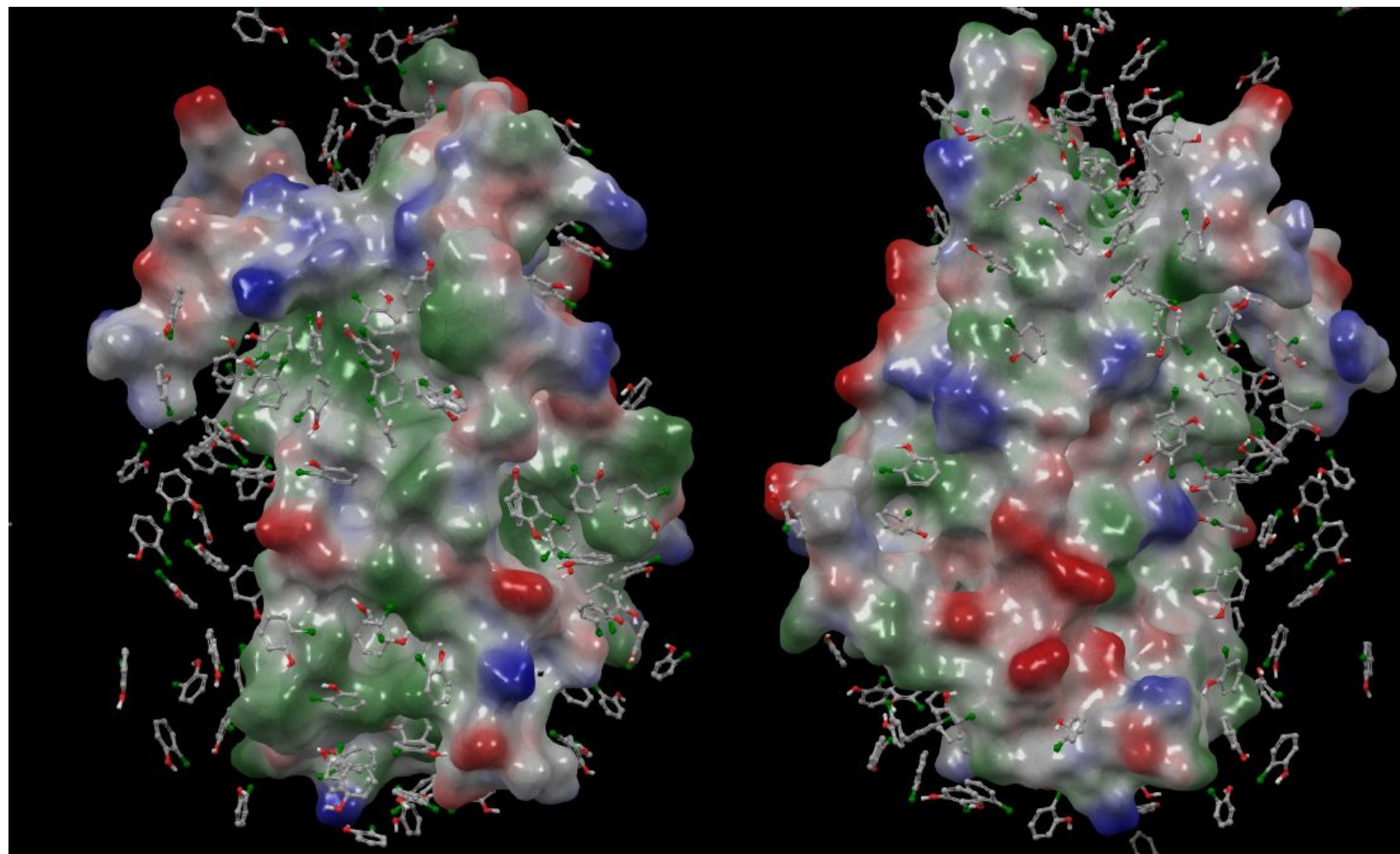


Compound	Size (nm)	% soluble aggregate
2-Chlorophenol	32	40
Catechol	15	17.6
Phenol	7	7.2
Benzyl alcohol	5	2.8
c-hexanol	5	2.2
None/Water	5	1.5

The solution is filtered and soluble aggregates were measured by quasi-elastic light scattering (size) and SEC(% aggregate)

Conclusion: 2-Chlorophenol > Catechol > Phenol > Benzyl alcohol ~ c-Hexanol > Water

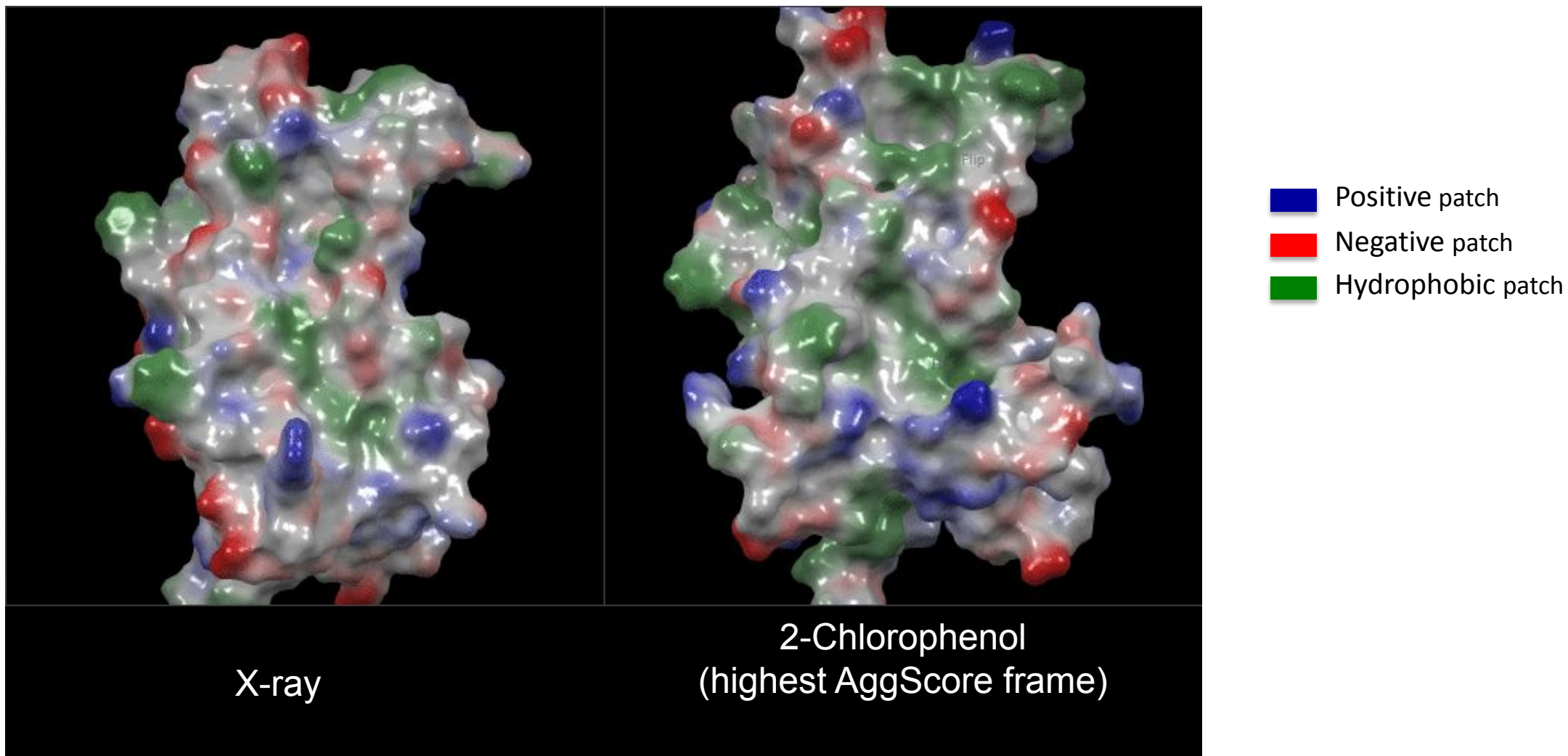
Cosolvent tends to bind at hydrophobic hotspots



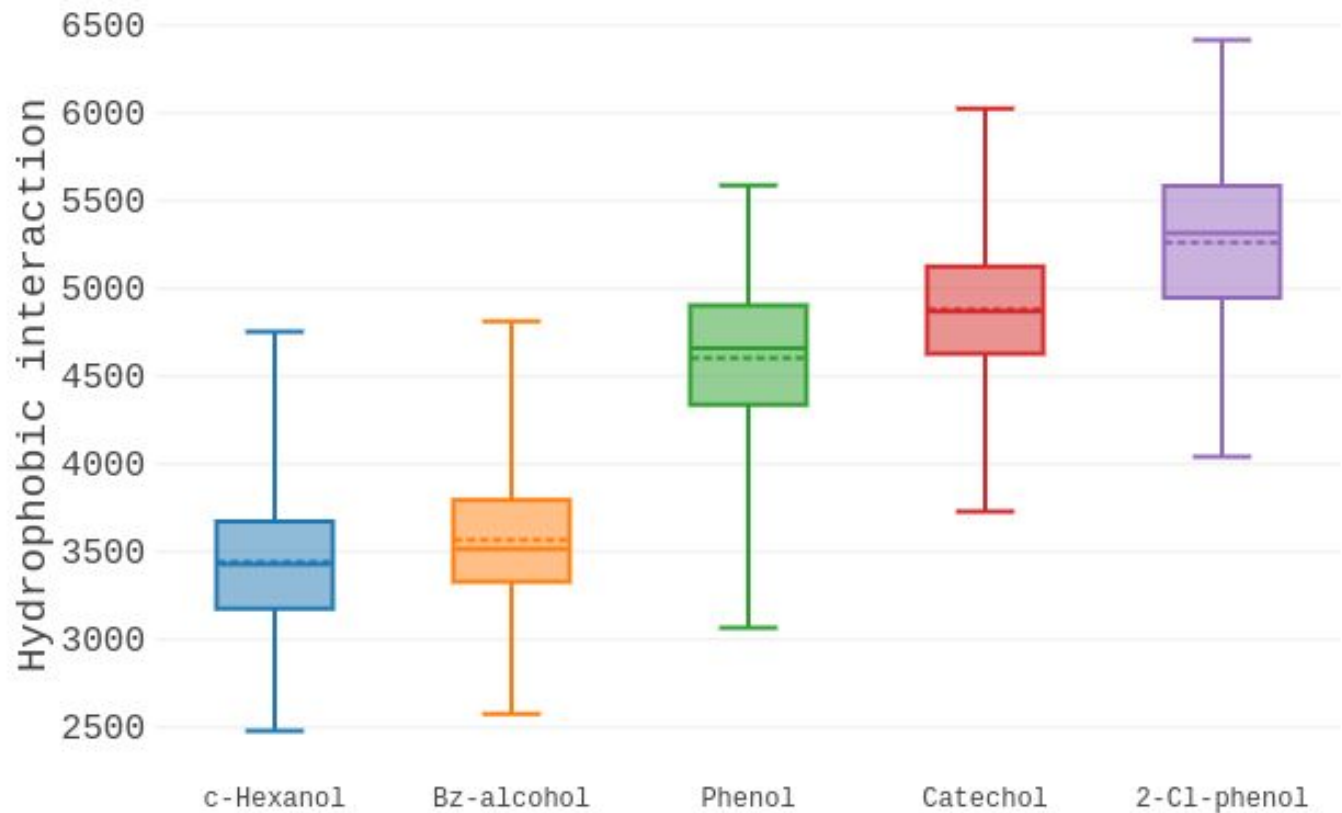
- Positive patch
- Negative patch
- Hydrophobic patch

2-Chlorophenol

Excipients induce exposure of hydrophobic pockets



Hydrophobic interactions with the excipients



Compound	% soluble aggregate
2-Cl-phenol	40
Catechol	17.6
Phenol	7.2
Bz-alcohol	2.8
c-hexanol	2.2
None	1.5
X-ray	NA

Preferential Interaction Coefficient

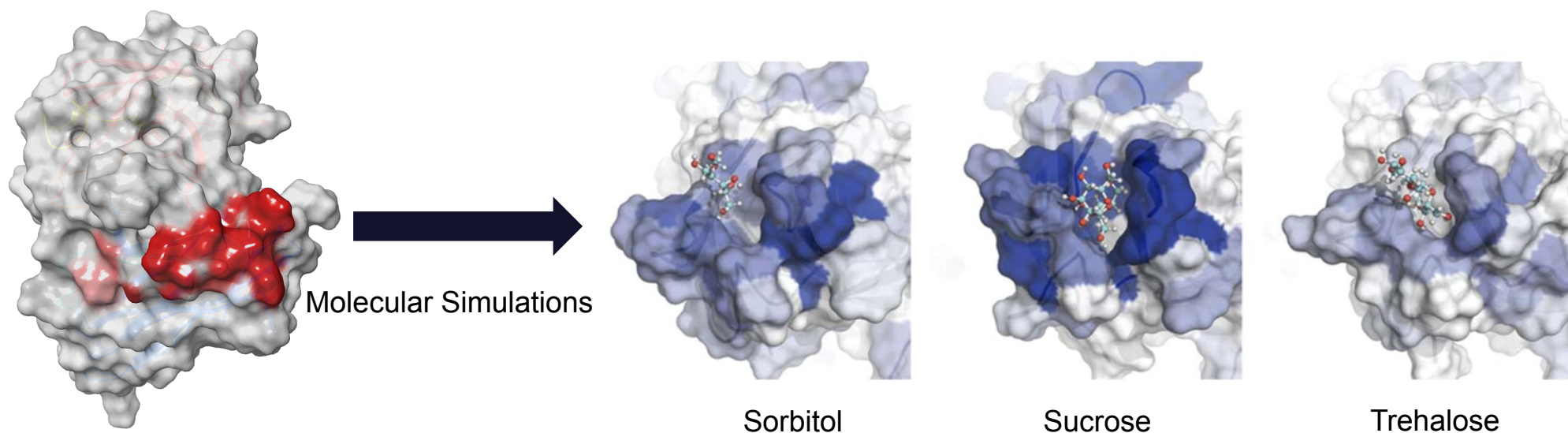
- Many biotherapeutics reach clinical development whilst carrying aggregation liabilities
 - At this stage the therapeutic protein is sequence-locked and cannot be modified
- Formulation approaches often aim to negate aggregation by masking residues with a higher aggregation propensity
 - Pairing the correct excipient(s) with the right protein is challenging
 - Molecular simulations can be used to observe protein-excipient interactions
- Preferential interaction co-efficient (PIC) can be used to identify which excipients interact with certain residues better than others to impact aggregation
 - *Does a particular excipient interact with a particular residue preferentially over water?*

$$\Gamma_{23}(r) = \left\langle n_3(r) - n_1(r) \left(\frac{n_3^{\text{total}} - n_3(r)}{n_1^{\text{total}} - n_1(r)} \right) \right\rangle$$

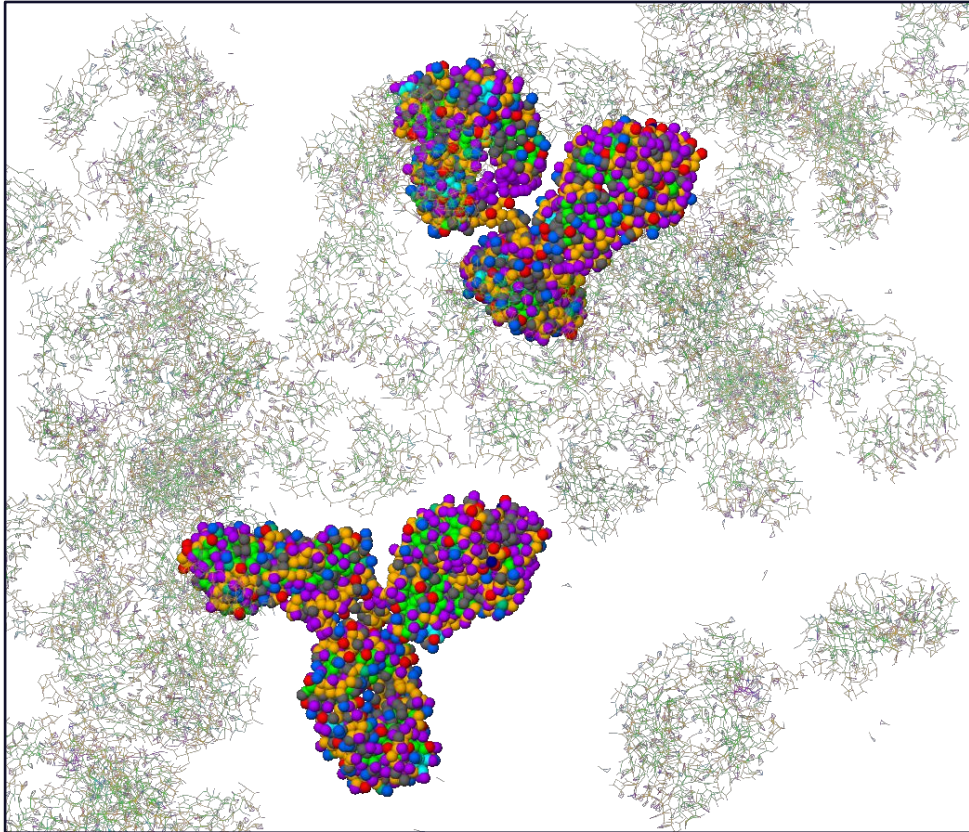


Preferential Interaction Coefficient: workflow

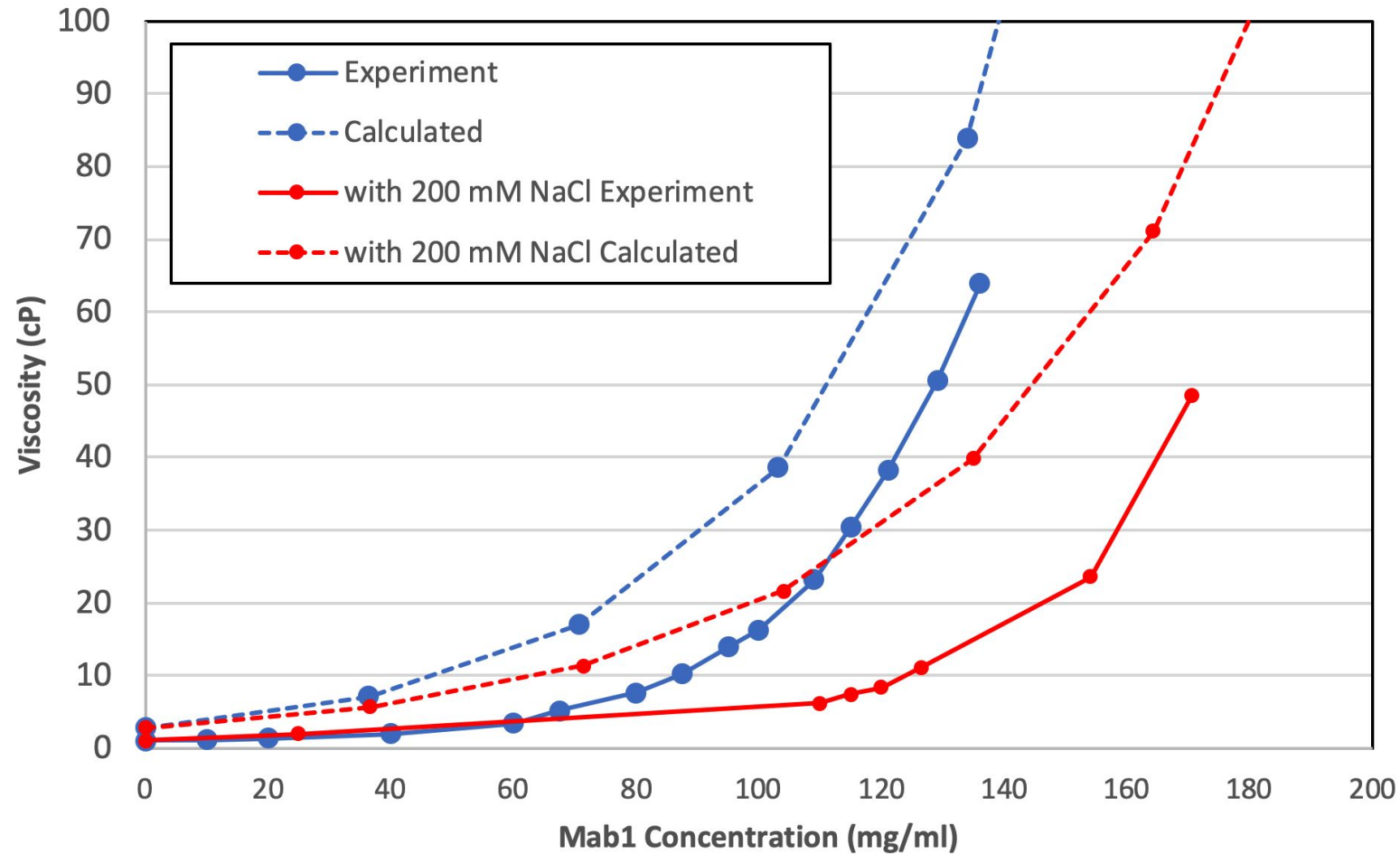
- Determine aggregation prone protein residues using AggScore
- Run molecular simulations of protein in presence of excipient(s) of interest
- Calculate PIC of aggregation prone residues to determine which excipient pairs best with which aggregation prone residue



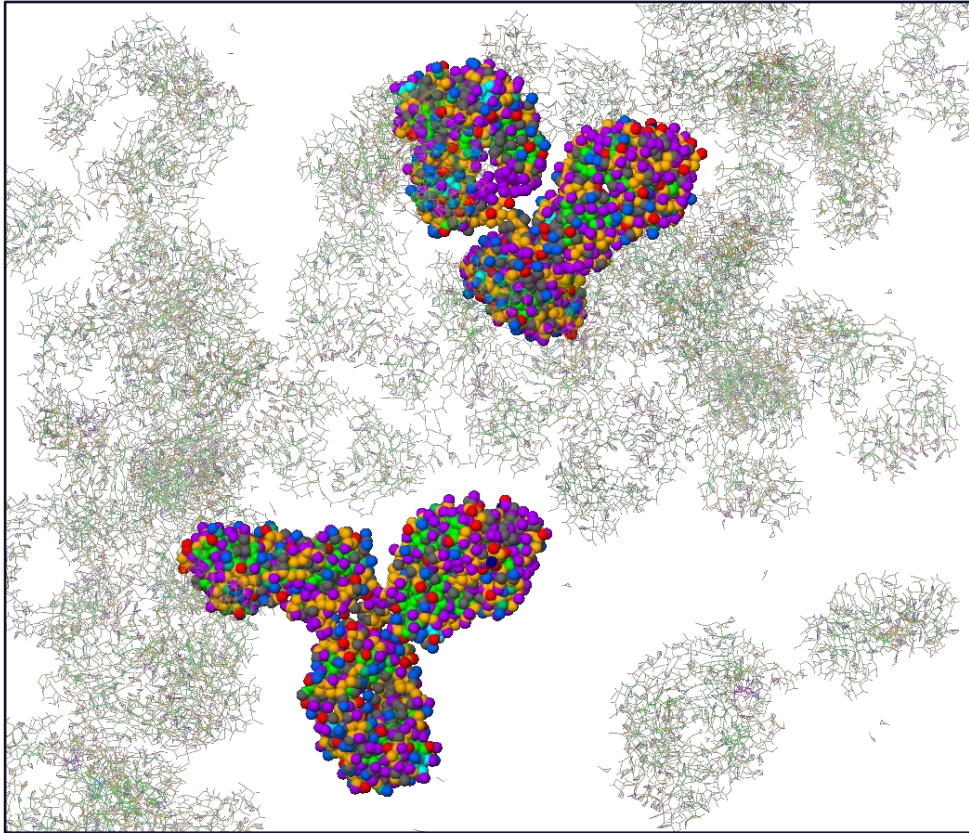
Salt effect on antibody viscosity



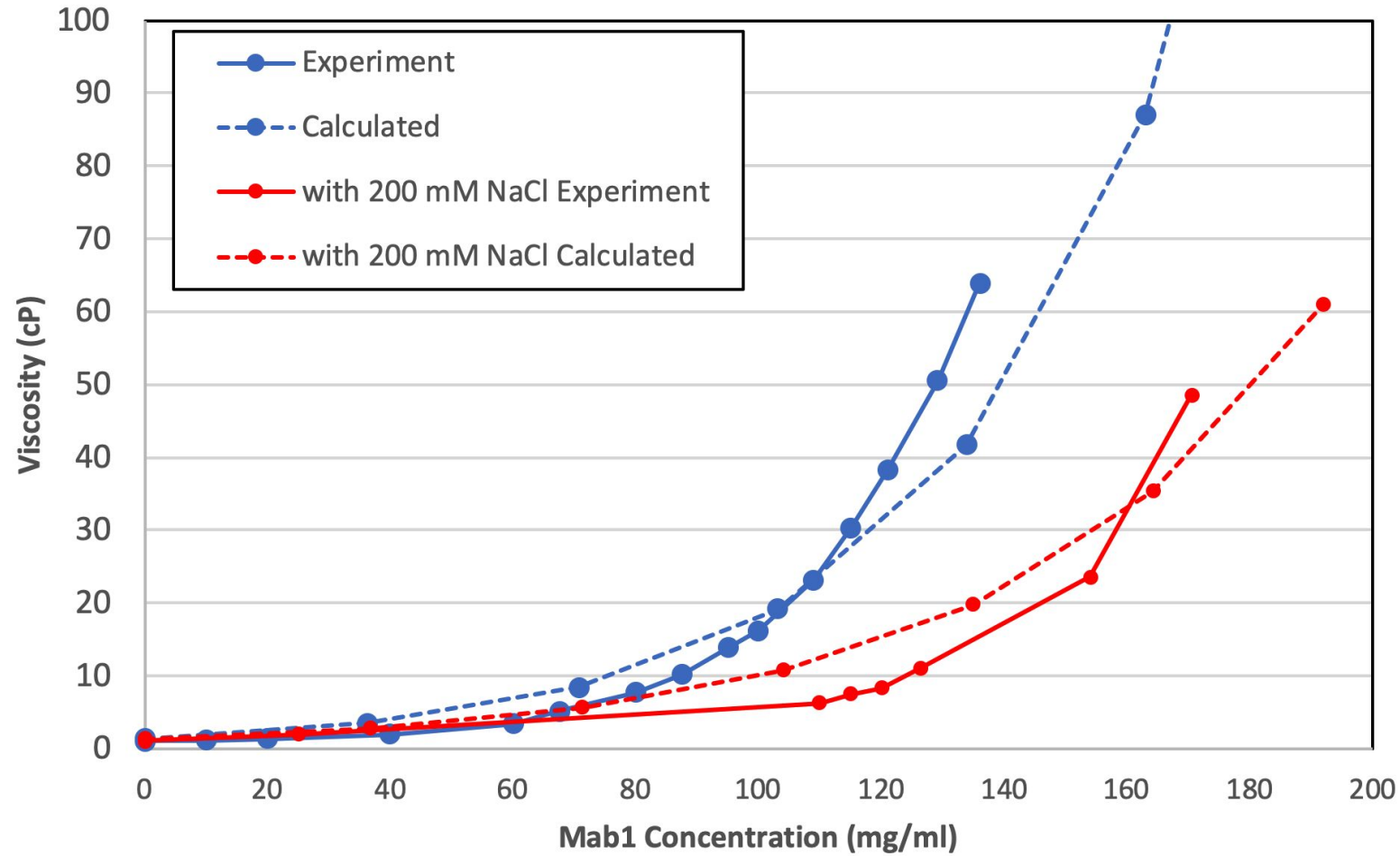
Viscosity: Effect of 200 mM NaCl



Salt effect on antibody viscosity

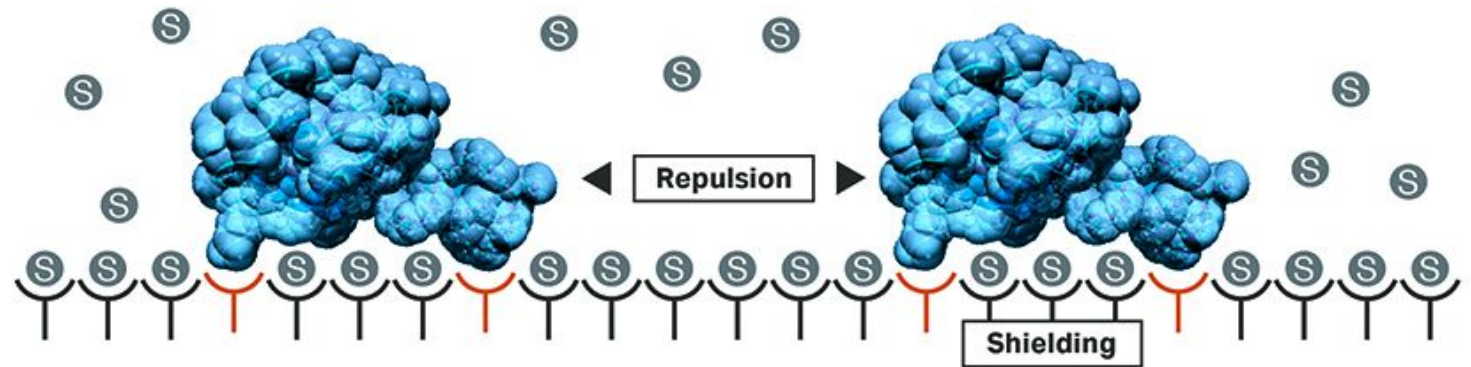


Viscosity: Effect of 200 mM NaCl, Scaled to match at 20 cP



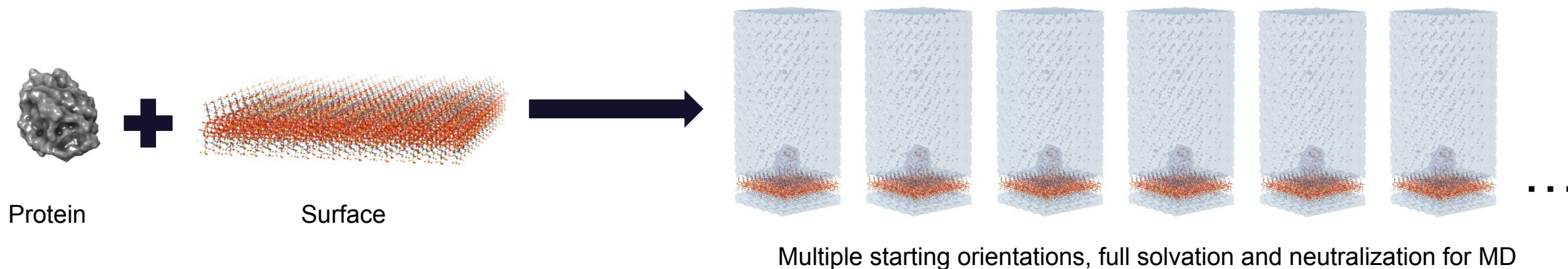
Extraction of steric mass action parameters

- In ion exchange chromatography (IEC), the steric mass action (SMA) isotherm model is often used to describe the interaction between proteins and IEC resins
- The model assumes multi-point binding and the exchange of counter ions
- Key SMA model parameters:
 - Characteristic charge (*number of sites of the protein with charged ligands on surface*)
 - Equilibrium coefficient (for the exchange reaction between the protein and salt counterions)
 - Shielding factor (*sites shielded by the protein*)
- Parameters can be derived from molecular simulations of proteins with appropriate surfaces



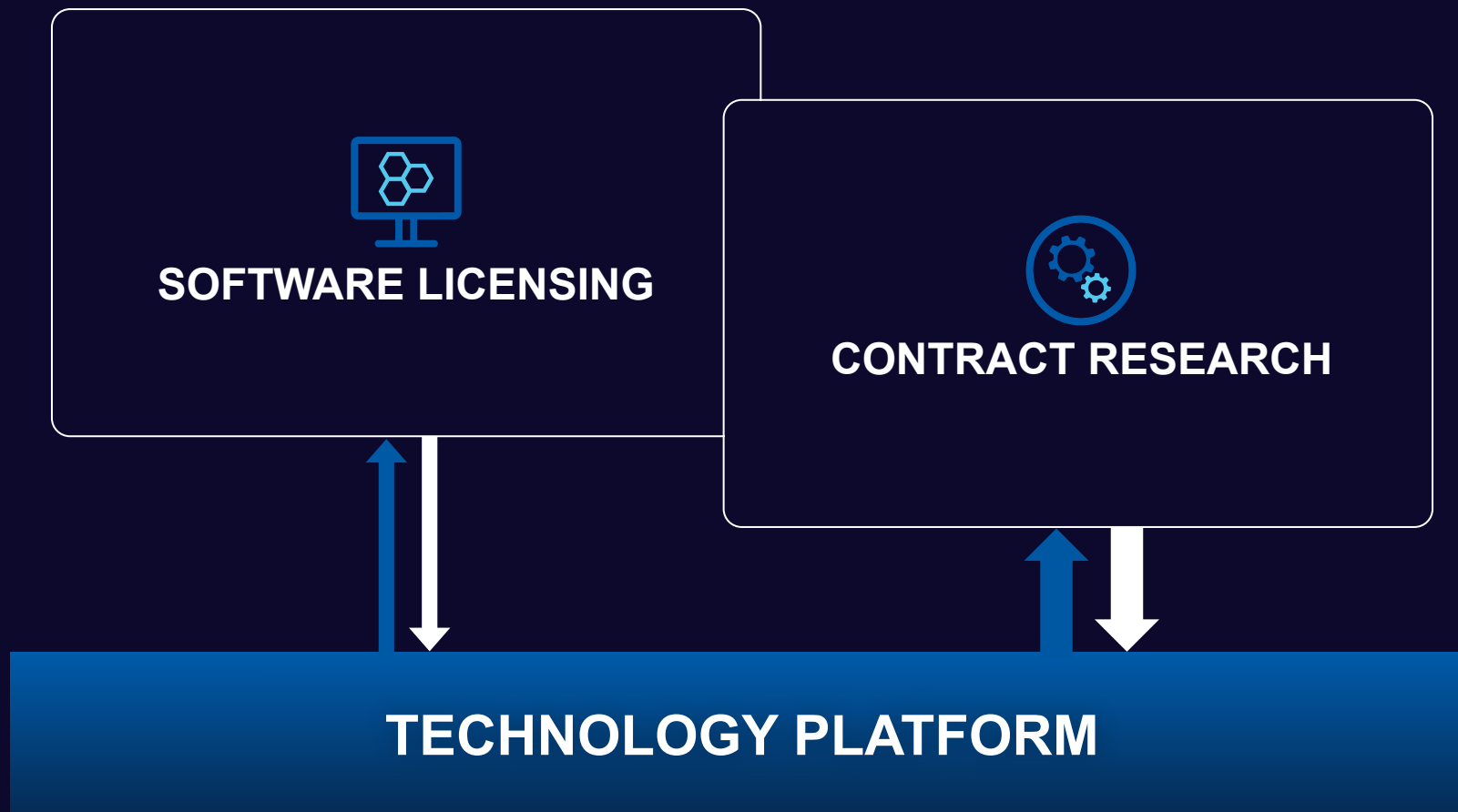
Protein-surface simulations

- Orient the protein above the surface with multiple unique orientations to avoid initial bias and to improve sampling
- Monitor orientations by comparing planar angles between surface and protein
- Calculate SMA parameters throughout simulation and observe convergence



- Use free energy simulations to estimate protein-surface binding energy
 - Equilibrium coefficient

Engagement models



How to work with Schrödinger

We can help with software,
technology evaluation, cloud access, workflows, cluster
setup, paid research, collaborations

Let's start a discussion
We are at booth 54

Adrian Komainda	adrian.komainda@schrodinger.com
Irene Bechis	irene.bechis@schrodinger.com
Dan Cannon	dan.cannon@schrodinger.com



INTEGRATING
Delivery Science
ACROSS DISCIPLINES





Schrödinger

Thank you