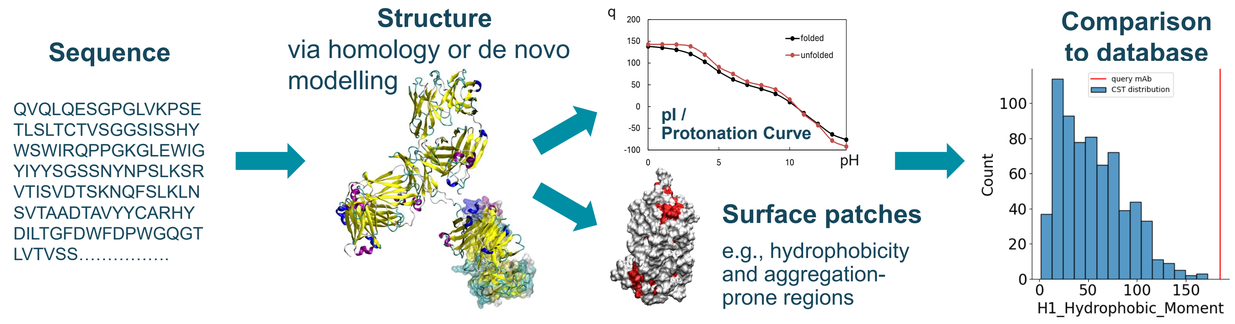
A novel in silico platform combining data-driven and physics-based models for protein formulation developability assessment

Andrea Arsiccio, Kristian Le Vay, Andre Sao Pedro, Thomas Pabstmann, Constanze Helbig, Dominik Brandstetter, Eleonora Cerasoli, Andrea Hawe, Tim Menzen

Coriolis Pharma Research GmbH, Fraunhoferstr. 18b, 82152 Martinsried, Germany

Coriolis Pharma

Combined data-driven and physics-based modeling platform to derisk protein formulation



- Formulation corridor → **Test fewer formulations**
- Susceptibility to degradation → **Focus selection of analytics**
- Formulation platform readiness → **Potentially skip pre-formulation screening**

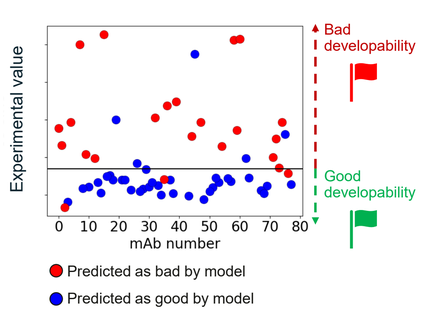
Speed

Risk

The development of novel therapeutic proteins is accompanied by huge financial and time investments. Advancing molecules through drug product development without proper scrutiny often leads to costly failures. In some cases, such failures are related to the inherent instability of the candidate molecule, and the difficulty of minimizing these instabilities by selecting a suitable formulation. Early characterization of a protein drug candidate's formulation developability is therefore crucial to reduce the risk. A comprehensive in vitro assessment is often constrained by the scarcity of drug substance material available in early stages of development, raising the interest in simulations and computational models. Currently, only very few in silico methods focusing on formulation development and assessing a protein drug candidate's formulation developability are available. We present a novel in silico platform based on data driven and physics-based models that combines computational techniques spanning different areas, including structure prediction, bioinformatics, machine learning and molecular dynamics calculations. The platform only requires the primary sequence of a therapeutic protein as input, thus eliminating the need for physical material.

Case study 1: Monoclonal antibody (mAbs 1-3)

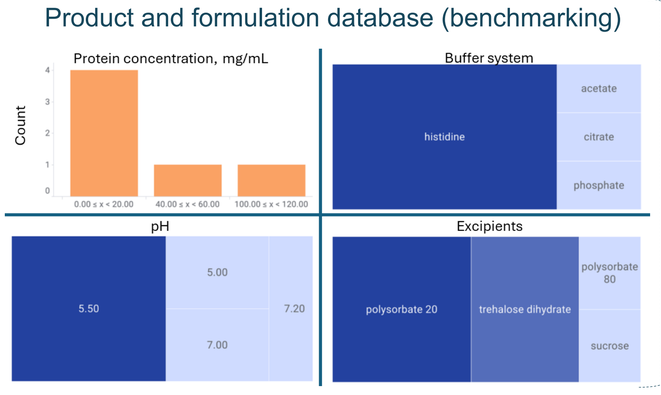
Machine learning (ML) models built from descriptors



Exemplary descriptors:

- pI
- Net charge
- Formal charge
- Patch energy
- SASA
- …

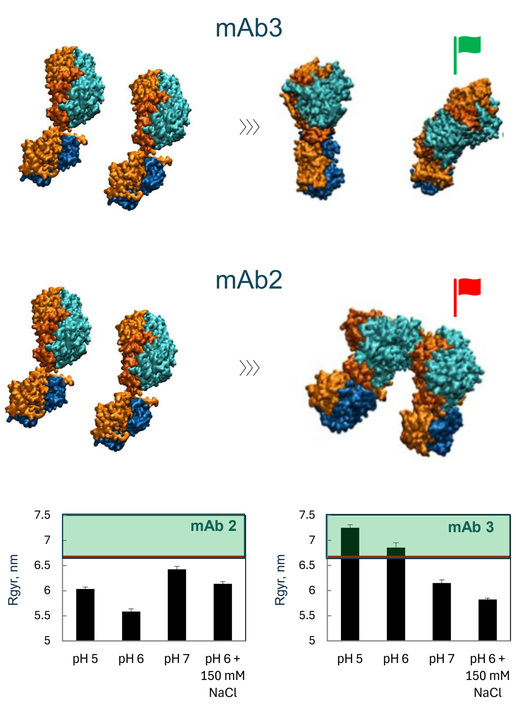
Product and formulation database (benchmarking)

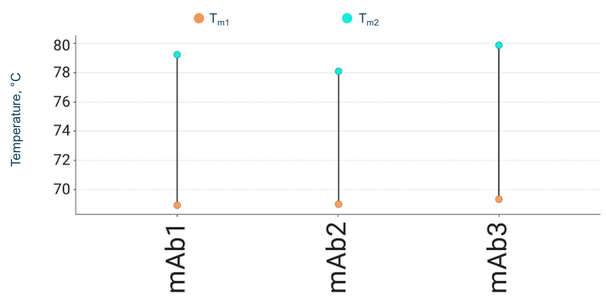


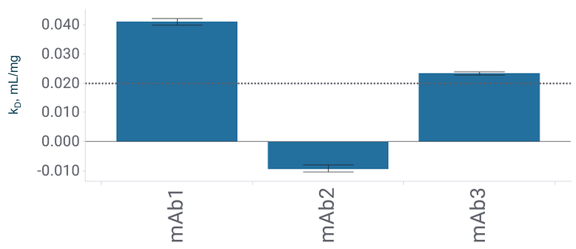
	mAb 1	mAb 2	mAb 3
Predicted pI (folded; unf.)	8.86; 8.95	7.52; 7.22	8.77; 8.80
	Data-driven ML models for liability identification		
Self-association	✓	✗	✗
Viscosity	✓	✗	✗
Deamidation	✗	✓	✓
Isomerization	✓	✓	✓
Oxidation	✓	✗	✓
	Physics-based MD simulations for self-association propensity		
Self-association	n.a.	✗	✓
	Formulation corridor suggestions from benchmarking		
mAb concentration	Up to 180 mg/mL	Mostly below 10 mg/mL	Mostly up to 25 mg/mL
Buffer	Histidine, pH 6.0	Acetate, pH 5.0-6.0	Phosphate, pH 6.0
Excipients	Polysorbate, NaCl	Polysorbate, sucrose	Polysorbate, trehalose

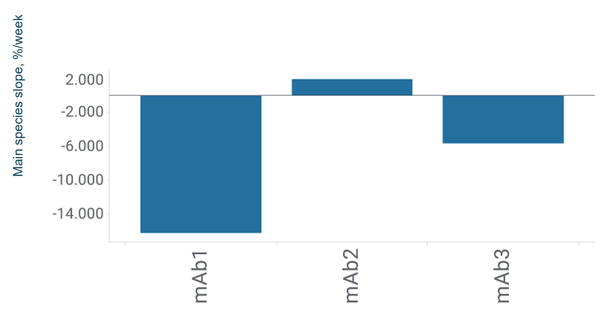
- ✓ mAb 1: sensitive to deamidation, employ an acidic pH
- ✓ mAb 2: self-association problems, not tunable by adapting formulation (expect problems at high conc.)
- ✓ mAb 3: self-association problems, can be addressed by reducing pH/ minimizing ionic strength

Self-interaction molecular dynamics (SMOD)









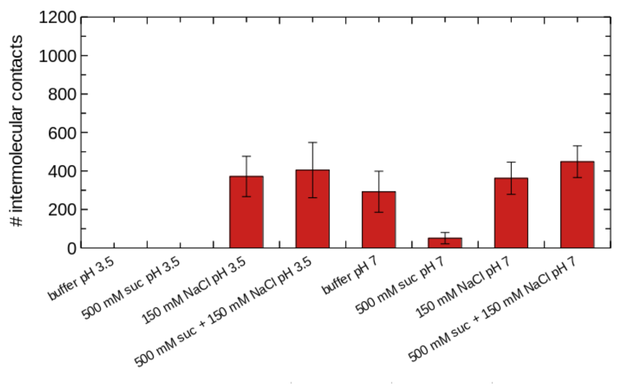
- ✓ Good conformational stability for all 3 mAbs
- ✓ Cation exchange chromatography confirms that mAb1 is at risk for chemical degradation
- ✓ Diffusion interaction parameter (k_D) measurements confirm that mAb 1 shows repulsive self-interactions, while mAb 2 is prone to self-association. mAb 3 has an intermediate behavior

Case study 2: Therapeutic protein (Granulocyte colony-stimulating factor, G-CSF)

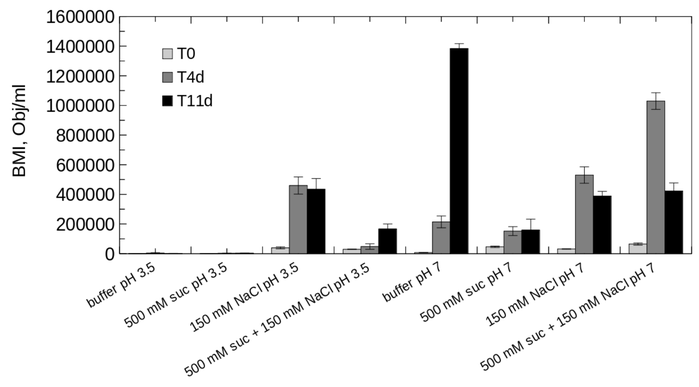
Physico-chemical outlier detection based on sequence- and structure-based descriptors (red flag warnings)

- ✓ Low propensity for deamidation or oxidation based on ASN and MET exposure, respectively
- ✓ Higher propensity for aggregation

MD to predict protein aggregation in different formulations



Subvisible particles by BMI after storage up to 11 days at 37 °C



- ✓ Elevated levels of subvisible particles were observed when G-CSF was stored at 37 °C at pH 7 and pH 3.5 in the presence of salt
- ✓ Molecular dynamics (MD) simulations of G-CSF in the same formulations correctly predicted G-CSF to be more stable at pH 3.5 at low ionic strength

Take home messages

- ✓ Avoiding costly late-stage drug development failures through early risk mitigation by performing formulatability assessment
- ✓ Physics-based and data-driven simulations (*in silico*) can save valuable time and material during formulatability assessment
- ✓ Although *in silico* models are becoming more advanced, also for non-mAb proteins, a combination with comprehensive analytical characterization is recommended

Arsiccio et al. J Pharm Sci. 2026 Apr; 115(4): 104196. doi: 10.1016/j.xphs.2026.104196.

Presenter contact: tim.menzen@coriolis-pharma.com

www.coriolis-pharma.com