ΛPOHΛ

Case study

From properties to behaviour: What antibody developability actually depends on

At a glance

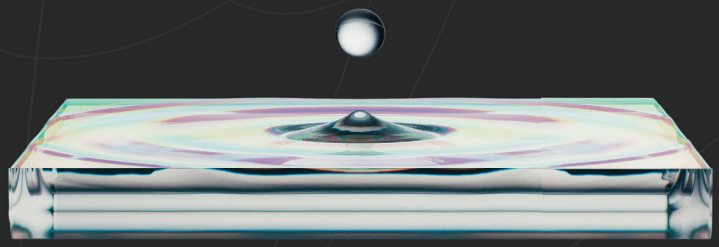
Sequences told us what a protein is. Structures told us what it looks like. But molecular behaviour, which determines whether a drug candidate survives formulation, manufacturing, and delivery, has remained a blind spot.

Apoha's Liquid State Intelligence resolves this missing layer directly, capturing how the molecule behaves in one high-dimensional readout. We call this the **state data**: a new fundamental data class of molecular science, alongside sequence and structure.

So what does state data measure?

In this study, we cofolded 231 clinical-stage antibodies with Chai-1, ran all-atom molecular dynamics on each, and asked which of 3,061 structural and dynamic descriptors explain Apoha's developability signal (VIBE1).

The answer: state data integrates across several molecular properties that drive developability **— surface-exposed charge distribution, CDR hydrophobic exposure, spatial aggregation propensity, and CDR flexibility —** in a single measurement, before enough material exists to run a conventional assay panel.



A new data layer for early screening — before critical decisions are locked in.

1) Which molecular properties determine Apoha's developability signal?

To link the physical measurement to molecular structure, we computed 892 static descriptors for 231 antibodies from cofolded VH/VL complexes (Chai-1, 5 models per antibody) and supplemented these with 2,169 dynamic descriptors from 10 ns all-atom MD simulations (3 replicates per antibody; 693 trajectories total) — a combined descriptor space of 3,061 features.

A 20-feature stability-selected logistic regression on this space classifies high vs. low VIBE1 antibodies at ROC AUC 0.925 with 81% recall — from a model that is essentially a weighted sum of biophysically named inputs.

The signal integrates across four biophysical axes simultaneously

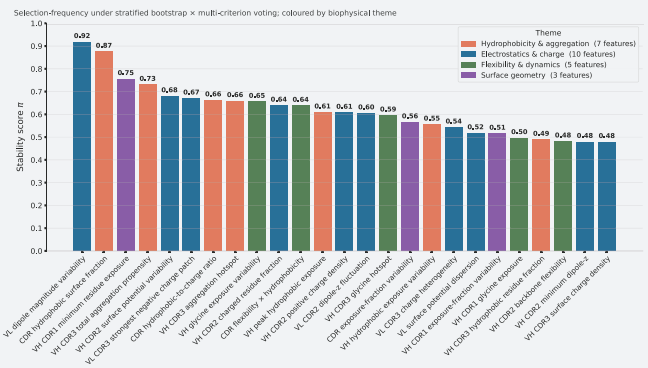


Figure 1: The 20 features selected by stability selection across the 3,061-feature descriptor space, ranked by stability score π. Bars are coloured by biophysical theme: charge and electrostatics (8 features), CDR hydrophobic exposure (6), surface area (4), and composite flexibility–hydrophobicity terms (2). The highest-ranked feature is VL dipole moment variability (π = 0.917).

- **Surface charge distribution and variability** — VH CDR3 positive SASA variability, VL dipole magnitude, charge patch size variability
- **CDR hydrophobic exposure** — spatial aggregation propensity, CDR GRAVY, hydrophobic SASA variability
- **Conformational variability** — disagreement across Chai-1 models; Cα RMSD/RMSF across MD trajectories
- **Composite flexibility × hydrophobicity** — engineered products of per-residue flexibility and hydrophobic SASA (the strongest linear correlates of VIBE1, Spearman ρ = 0.34–0.40)

VIBE1 correlates weakly with any single conventional readout but identifies multi-liability molecules with high precision — because surface activity at a liquid interface is itself a multi-property phenomenon.

2) From signal to fix: understanding why a molecule is a poor candidate for developability

VIBE1 flags risky candidates. The interpretable model goes one step further — for every flagged antibody, it decomposes the prediction into contributions from hydrophobicity, electrostatics, and flexibility.

This means customers don't just see that a candidate is at risk; they see which biophysical class is driving the risk, and where to focus engineering effort.

Here, three high-VIBE1 illustrate the three failure modes the model resolves:

- **Solrikitug** — hydrophobicity-driven
- **Lodapolimab** — electrostatics-driven
- **Urelumab** — flexibility-driven

Per-antibody attribution localises the failure mode to a single biophysical class

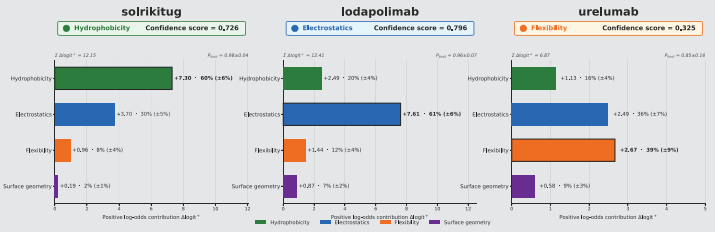


Figure 2: Descriptor-class attribution for three model-flagged antibodies. Top: positive log-odds contributions of the highest-ranked features for solrikitug, lodapolimab, and urelumab, coloured by descriptor class (hydrophobicity, electrostatics, flexibility, chemical liability). Bottom: category enrichment, defined as the mass fraction of positive log-odds in each class divided by the base rate of that class across the full feature set. Values above 1 indicate the dominant biophysical driver. P(c = class) gives the probability that the indicated class is correctly identified as the dominant attribution category.

3) Where the signal is most robust, and what comes next.

Across 6 model configurations (3 model classes × 2 feature counts), most high-VIBE1 antibodies are recovered consistently. Three are strictly irreducible — atoltivimab, dectrekumab, and vulinacimab — misclassified by every configuration. A further 8 are recurrently missed by 5 or 6 of the 6 configurations.

These cases appear to fail through mechanisms not captured by the current Fv-level descriptor set: candidate sources include immunogenicity, post-translational modifications, full-length antibody features beyond the variable fragment, and formulation-specific instabilities. Closing this gap is an active area of work — including protein language model embeddings as a new descriptor class.

Most flagged candidates are recovered consistently across model configurations

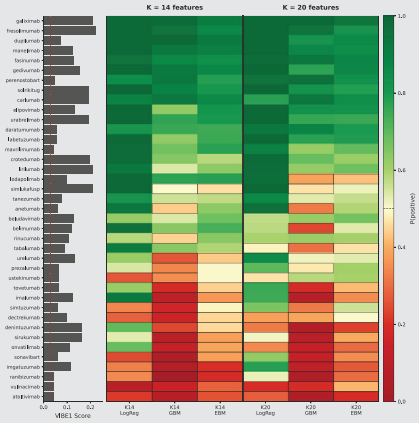
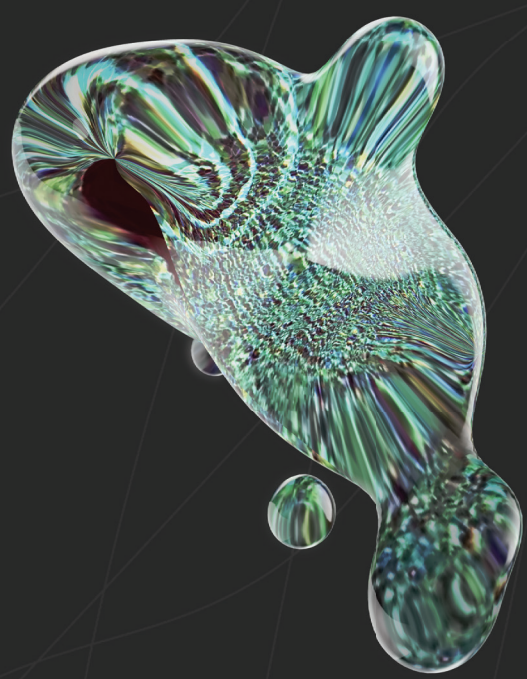


Figure 3: Per-antibody out-of-fold probability for the 39 high-VIBE1 antibodies across 6 configurations (K=14 and K=20 feature sets × logistic regression, gradient boosting, and EBM). Rows are antibodies, sorted by mean predicted probability. The strictly irreducible group (bottom) is consistently misclassified by every configuration.

Using state data in early screening.

- A single state measurement using as little as 8 µg of material delivers two things no panel of conventional tools delivers together:
- A precision flag for high-risk candidates, VIBE1 flags multi-liability molecules with zero false positives, before enough material exists for a full developability panel.
- A mechanistic readout of why each flagged candidate fails — hydrophobicity, electrostatics, surface geometry or flexibility — so engineering effort is directed where it matters.

For discovery teams, this means moving developability calls from late-stage characterisation up to hit ID, eliminating doomed candidates earlier in the pipeline, with a single measurement that integrates across the biophysical axes that decide whether a drug reaches patients.



About Apoha

Apoha has invented the world's first technology platform to resolve States - a new fundamental data class of molecular science, alongside Sequences and Structures.

Sequences told us what a molecule is. Structures showed us what it looks like. But molecular behaviour - the layer that determines whether a drug works, a formulation holds, or a material performs - has remained elusive. Until now.

We call this new paradigm Liquid State Intelligence: unlocking complex molecular behaviour under real-world conditions, generating data no simulation or model can reproduce. As this data layer grows, it gives humanity a new ability - not just to analyse the physical world, but to redesign it.