From bench assays to bedside: a context-embedding transformer predicts monoclonal antibody viscosity, clearance, and regulatory success

Sabitoj Singh Virk ¶* and Akashdeep Singh Virk¶
Rensselaer Polytechnic Institute, 110 8th St, Troy, New York 12180

¶Both authors contributed equally to this work

Late-stage failures of monoclonal antibody (mAb) programs often reflect developability liabilities, including high-concentration viscosity and rapid clearance, that standard early screens only partially capture. We introduce ACeT, an attention-based context-embedding transformer that fuses routine early-stage assay readouts to predict three endpoints: high-concentration viscosity, mouse intravenous clearance, and Phase I-to-approval outcomes. On held-out mAbs, the model achieved $R^2 \approx 0.75$ for viscosity and $R^2 \approx 0.80$ for clearance, and for clinical progression reached ~78% balanced accuracy on a 112-IgG1 cohort with outcomes locked by October 2022, then 0.83 on a temporally independent set of 14 clinical-stage mAbs whose outcomes matured by May 2025. The approach outperformed single-assay heuristics and established machine-learning baselines, and model attributions recover known biophysical drivers. Retrospective portfolio simulations, under typical industry cost and success-rate assumptions, suggest that prioritising antibodies with favourable transformer scores could reduce late-stage attrition and potentially redirect up to US$ 17 billion in capital, while also lowering animal use. By unifying heterogeneous assays in a single encoder, this framework improves the fidelity of early-stage developability decisions and offers a practical route to triage viscosity and clearance risks while prioritizing mAbs with a higher likelihood of clinical progression.

Keywords: Monoclonal antibodies; antibody developability; integrated developability screening; antibody viscosity prediction; antibody clearance prediction; regulatory approval forecasting; machine learning; transformer model; multimodal data integration; biophysical interpretability

*Author to whom correspondence should be addressed; electronic mail: sabitoj@umich.edu

Introduction

Monoclonal antibodies now anchor standard-of-care across oncology, immunology, and neurology, yet the attrition between first in human dosing and licensure remains stubbornly high. Across modalities, the likelihood of approval (LOA) from Phase I hovers around one in ten; analyses of 7,455 development programs from 2006-2015 placed the LOA at 9.6% overall (11.9% non-oncology), with wide variation by disease area and biomarker use.[1-4] Antibodies fare better than small molecules but still face material risk: recent modality specific estimates place Phase I to approval success for mAbs at ~15-30% depending on indication and cohort.[5] Because late-stage failures dominate the fully capitalized cost per approval (median ~$1.1 billion when failures are included), improving early triage is both a scientific and economic imperative.[6,7]

A growing body of evidence points to "developability" liabilities: manufacturability, formulation, delivery, and pharmacokinetic (PK) risks, as major contributors to that attrition.[8,9] Two liabilities are especially consequential for antibodies. First, high solution viscosity at clinical concentrations can render otherwise potent molecules infeasible for subcutaneous administration: patient acceptable syringeability typically degrades above ≈15-20 cP for standard devices and injection times, a limit frequently encountered near ≥100 mg mL^{-1}.[10-12] Mechanistically, viscosity emerges from a delicate balance of weak protein-protein interactions; simple heuristics based on net charge or a single dilute-solution metric (e.g., second virial or diffusion interaction coefficients, B22 or kD) often fail because short-range attractions and long-range repulsions jointly sculpt the microstructure.[12-17] Second, unexpectedly rapid clearance, via nonspecific interactions, target mediated disposition, or neonatal Fc receptor (FcRn) pathway perturbations, reduces exposure and forces impractical dosing. Quantitative links between in vitro surrogates and PK are increasingly well documented: for example, heparin chromatography can report on non-specific cell surface interactions predictive of pinocytotic uptake and faster clearance, while human-FcRn affinity and chromatography correlate with recycling and half-life.[18-24]

To surface these liabilities before the clinic, discovery teams now run multiplex, partially orthogonal biophysical batteries. Landmark analyses of clinical stage mAbs profiled >10 assays spanning self-interaction, hydrophobicity, charge, stability, and aggregation, and showed that liabilities cluster across semi orthogonal axes rather than along any single readout.[8] High-throughput variants (e.g., kD, Affinity-Capture Self-Interaction Nanoparticle Spectroscopy (AC-SINS), and Hydrophobic Interaction Chromatography (HIC), etc.) are now routine and reproducible across organizations; improved implementations closely track legacy methods, and anonymized assay-level readouts can be shared and aggregated across sites without disclosing proprietary sequences. [8,25-27]Heuristic "red flag" thresholds remain common (e.g., Polyspecificity Reagent (PSR) > 0.27 to denote polyspecificity; heparin column thresholds to mark nonspecific binding), but performance varies by panel

and context, and single assay cutoffs rarely achieve the sensitivity-specificity balance needed for confident go/kill decisions. [8,24,28]

Predicting clinical progression (Phase I→approval) requires integrating across assays and contexts. Retrospective studies consistently show that no single developability metric explains success; instead, multi-dimensional patterns spanning colloidal behavior, stability, non-specificity, and PK surrogates, better stratify risk.[8,29] More broadly, when prospective enrichment signals are used (for example, biomarker based patient selection), LOA can shift dramatically (e.g., 25.9% with selection biomarkers versus 8.4% without in 2006-2015 cohorts), underscoring that integrated evidence moves the needle on outcomes.[2] Yet most computational tools in antibody discovery still emphasize sequence only features or structural surrogates; while powerful for affinity maturation and design, such models typically cannot ingest buffer composition, formulation variables, or assay condition dependencies that dominate manufacturability, clearance, and clinical success.[30-36] At the same time, purely sequence/structure based machine-learning models remain hamstrung by sparse public corpora: variable-domain sequences are disclosed piecemeal via WHO INN lists or patents, omitting many private assets, while the critical formulation context and screening outcomes are almost never published.[37-39] This gap motivates learning frameworks that are assay aware and context aware, fusing heterogeneous early stage measurements into clinically meaningful forecasts.

Here we introduce the Agnostic Context-embedding Transformer (ACeT), a modeling framework that unifies diverse early-stage readouts into a single latent representation and attaches modular prediction heads for distinct endpoints. The encoder is agnostic to data modality (continuous, categorical, binary), enabling principled fusion of routinely collected assays; task specific heads then learn mappings to high concentration viscosity, in vivo clearance, and Phase I→approval outcomes. The architecture emphasizes data efficiency and interpretability, combining augmentation strategies for low n regimes with feature attribution analyses that recover mechanistically sensible drivers (e.g., kD and size-exclusion chromatography (SEC) features for viscosity; heparin/PSR like signals for clearance/progression). In comprehensive evaluations, the integrated model outperforms single assay rules and conventional baselines, generalizes to held out antibodies and an external clinical cohort, and translates predictive gains into portfolio level economic value by reducing false positive continuation and false negative kills (see Results).[6,7]

Collectively, our results argue that assay aware representation learning can close the "developability gap" that separates discovery from clinical success. By moving beyond single metric screens and sequence only proxies to an integrated, context sensitive predictor, ACeT provides a practical path to triage high viscosity and fast clearance liabilities and to forecast clinical progression with decision grade fidelity, improving the odds that the right antibodies reach patients, faster and at lower cost.

Results

Overview of the Agnostic Context-embedding Transformer

We engineered the agnostic context-embedding transformer (ACeT) (Fig. 1a-c) to fuse heterogeneous early-stage read-outs into a unified latent representation. Each raw measurement x_i, whether a continuous titre, a categorical binding score, a binary flag, or any other data type, is first linearly projected into a low-dimensional value embedding and summed with a learned feature-identity embedding, yielding one token per input and rendering the model intrinsically agnostic to data modality (Fig. 1a). The resulting token sequence is passed through a multi-layer, pre-norm transformer encoder equipped with multiple attention heads and non-linear feed-forward blocks; global average pooling then collapses the sequence into an order-invariant latent vector (Fig. 1b). Order invariance assumes that assay columns are supplied in the canonical order used for training; permuting columns at inference would mis-align the feature-identity embeddings.

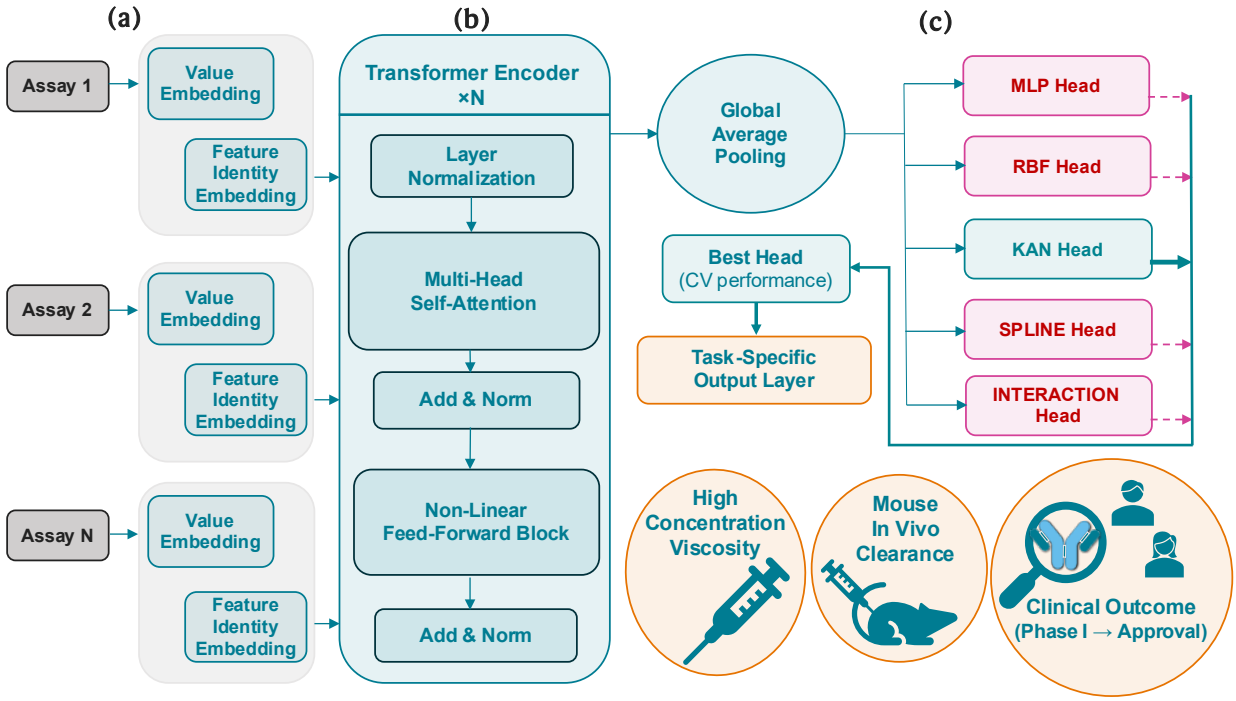

Fig. 1 | Architecture of the agnostic context-embedding transformer (ACeT) for developability prediction. a, Each assay read-out is encoded as a value embedding plus a learned feature-identity embedding, generating one token per measurement. **b,** A stack of pre-norm self-attention layers followed by global average pooling converts the unordered token set into an order-invariant latent vector. **c,** Parallel prediction heads, each instantiating a distinct inductive bias, are evaluated on identical data splits; the head with the highest cross-validation performance is routed to a task-specific output layer, enabling rapid adaptation to endpoints such as high-concentration viscosity, in-vivo clearance and clinical outcome. Together, (a-c) implement a modality-agnostic encoder with order-invariant pooling and task-tailored heads that perform well even with few tokens.

At the encoder, ACeT is data- and modality-agnostic: it ingests numeric features from experimental or computational pipelines and, once mapped to numeric descriptors or embeddings, can also fuse sequences, structures, images, or text. Beyond the encoder, ACeT features a modular prediction stage in which several interchangeable heads, each

embodying a distinct inductive bias, operate in parallel on the shared latent vector (Fig. 1c). Because these heads attach at the representation level, they can be added, removed, or replaced without modifying the encoder. A lightweight evaluation routine benchmarks their performance on identical cross-validation splits and automatically selects the best-performing option for the task at hand. This plug-and-play design streamlines rapid experimentation, preserves the core architecture, and lets the data decide which functional form is most appropriate for each prediction objective.

Taken together, these design choices: feature specific token embeddings, order-invariance via global-average pooling over token outputs, and task-tailored prediction heads, inject domain specific inductive biases, enabling strong performance even on very short input sequences. A similar benefit from inductive bias has been observed in compact Vision-Transformer designs, where introducing locality improves data efficiency[40]. ACeT underpins all subsequent results, beginning with high-concentration viscosity prediction and extending to in-vivo clearance and clinical outcome forecasting.

Viscosity prediction from biophysical assays

Using the ACeT architecture, we achieved accurate predictions of high-concentration solution viscosity from low-concentration biophysical assay data. In a parity analysis of predicted vs. measured viscosities for held-out antibodies (n = 23), the model's predictions closely tracked the experimental values across a ~5-45 cP range (Fig. 2a).

Most points fell near the unity line, yielding an overall coefficient of determination $R^2 \approx 0.75$ and a root-mean-square error ~4.8 cP on the test set. The model maintained strong performance even at the extremes, showing minimal error in the low-viscosity (~5-15 cP) and mid-viscosity (~15-30 cP) ranges, and only a slight underestimation bias for the highest-viscosity samples (>30 cP), all within experimental error. Because viscosities were measured at varying protein concentrations (110-200 mg mL^{-1}) without normalization[41], we restrict interpretation to this range; despite potential variability introduced by mixed concentrations, performance remained robust, and future work may normalize to a reference concentration.

Stratifying predictions by viscosity regime confirmed this robustness (Fig. 2b). In the low, medium and high bins, the transformer's estimates remained tightly clustered around the unity line, and the surrounding bootstrap cloud showed no systematic deviation, indicating that the model does not rely on interpolation within any single range. By capturing nonlinear assay-viscosity relationships without systematic deviation across the viscosity spectrum, the transformer outperforms conventional approaches.

The model markedly exceeded baseline machine-learning models in both cross-validation and independent tests (Fig. 2c). For example, on the test set the transformer achieved R^2 >0.7, substantially higher than the best alternative (ridge regression, R^2 ~0.5) and well above support-vector regression (~0.47) or a random forest (~0.29).

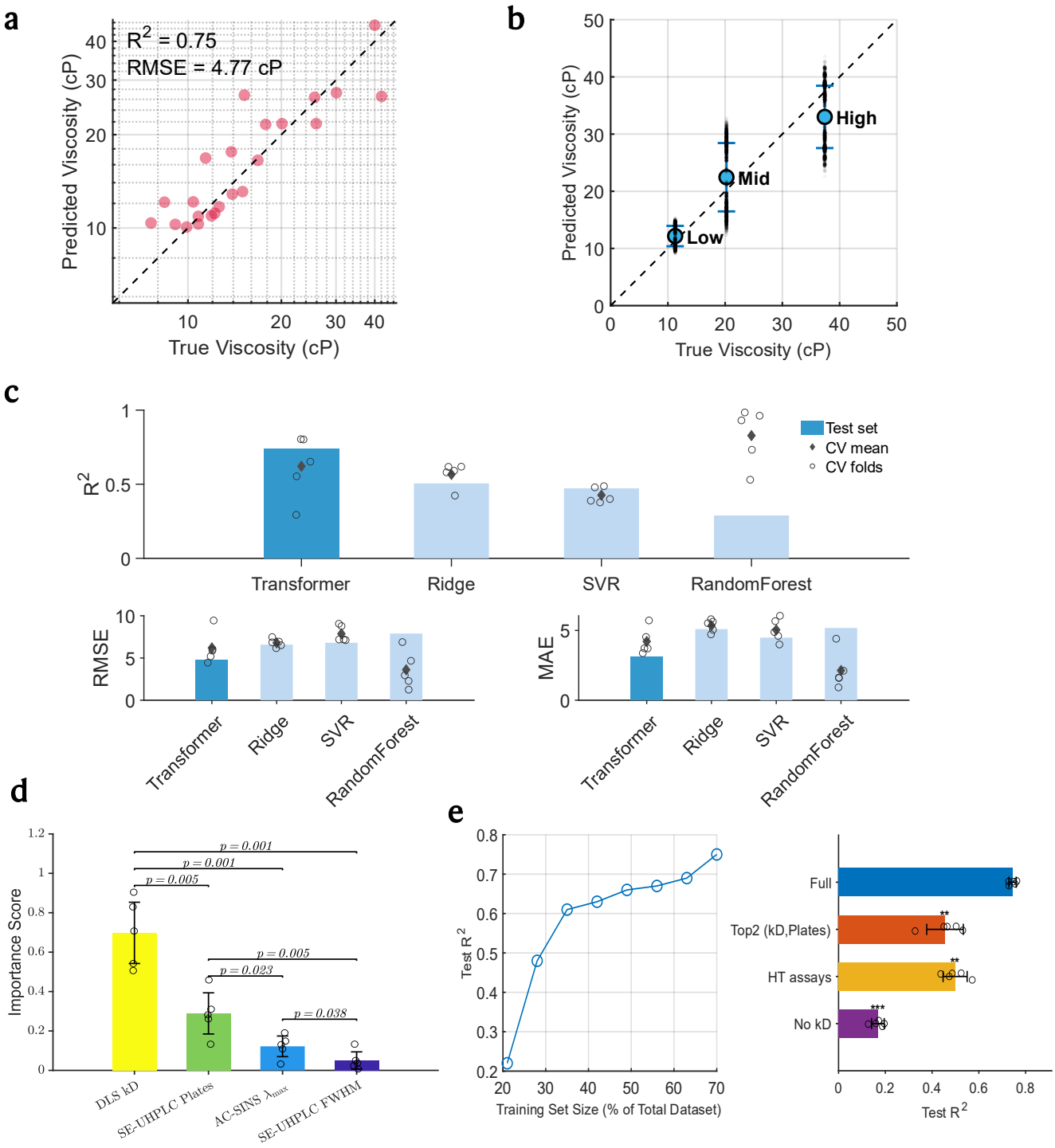

Fig. 2 | ACeT infers high-concentration viscosity from low-concentration assays. a, Parity plot for 23 test antibodies (unity line shown; R^2=0.75, RMSE ≈ 4.8 cP). **b,** Calibration by viscosity tertile: circles, mean true vs. predicted; vertical bars, bootstrap RMSE (1000 draws); faint dots, resamples; dashed line, unity. **c,** Model comparison (transformer, ridge, SVR, random forest): bars, test metrics; diamonds, five-fold CV means; open circles, CV folds. **d,** Permutation importance of the top four assays: bars, means ± s.d.; open circles, fold values; brackets, pairwise P values (Holm–Šidák). **e,** Left, learning curve of test R^2 versus training-set size. Right, feature-ablation impact on test R^2: horizontal bars, mean ± s.d. across five runs; open circles, individual runs; asterisks, significance versus full model (* < 0.05, ** < 0.01, *** < 0.001; paired two-tailed *t*-test, Holm-corrected).

This translated to markedly lower prediction errors: the transformer's test RMSE was ~5 cP, versus ~7-8 cP for the baselines, and its mean absolute error (~3 cP) was about 2 cP lower than those of conventional models. Notably, the transformer's performance generalized consistently (mean R^2 ~0.62 across folds, nearly matching the independent test R^2) with no signs of overfitting, whereas the random forest exhibited high variance and severe overfitting (some cross-validation folds showed spuriously high R^2 up to ~0.9, but overall test R^2 dropped to ~0.3; Fig. 2c). These results highlight the ACeT's ability to learn generalizable viscosity predictors from limited data, unlike simpler linear models that under-fit or complex ensembles that overfit. We included ridge, SVR and random-forest baselines as domain-standard comparators spanning linear, kernel and tree-ensemble families that remain competitive in low-feature biotherapeutics settings[42-44].

To understand the model's decision process, we analysed feature importance and found that the predictions were grounded in physically meaningful assay features. The dynamic light scattering interaction parameter kD emerged as the single most influential feature (Fig. 2d), with a considerably higher importance score (~0.7 on a 0-1 scale) than any other input. This parameter, measured at dilute concentration, reflects protein-protein interaction propensity: more negative kD values indicate attractive interactions that often portend high viscosity[15,45]. Consistent with domain knowledge, antibodies with unfavourable (negative or low) kD indeed showed elevated viscosity[15,45,46], and the model appropriately up-weighted this feature. The second most important predictor was the SEC peak sharpness (column plate count), an indicator of sample purity/aggregation[47]. A high SEC plate count (sharp, monomer-dominated peak) correlated with lower viscosity[41], and accordingly this feature received a moderate importance score (0.30). By contrast, other assays contributed only modestly, for instance, the AC-SINS self-interaction nanoplasmonic assay[48] and the SEC peak width (polydispersity) had importance scores around 0.1 or below, suggesting that while these features provide some complementary information (e.g. colloidal stability under orthogonal conditions), they are not primary drivers. In the dataset[41], SEC plate count and peak width are inversely related measures of peak broadening and were highly correlated; accordingly, the model relied on plate count as the representative descriptor, rendering peak width largely redundant under permutation. This explains the low apparent importance of peak width despite its conceptual linkage to plate count. The dominance of kD and SEC metrics in the learned importance ordering aligned with biophysical expectations, reinforcing the interpretability and credibility of the model's predictions.

Finally, learning-curve and ablation studies underscored data-efficiency and the critical nature of kD (Fig. 2e). Model performance improved monotonically with training set size and showed no sign of plateau within the available data, indicating that additional training examples would likely further enhance accuracy. In parallel, removing the DLS kD feature, identified above as critical, caused the test-set R^2 to plummet from ~0.75 to ~0.17 ("No kD"), a highly significant loss of predictive power ($p < 0.001$). Using the two most important

features (kD and SEC plate count) only partially recovered performance ($R^2 \approx 0.45$, "Top2"), and relying solely on the high-throughput assays (omitting SEC) improved only slightly ($R^2 \approx 0.49$, "HT assays"). Together, these tests confirm that while all assay inputs contribute, the DLS kD metric is indispensable for capturing the bulk of the viscosity signal. Moreover, the steadily climbing learning curve suggests the ACeT model is not data-saturated and would benefit from larger training datasets, implying that its current performance represents a balance limited by available data rather than an intrinsic ceiling.

Mouse-IV clearance prediction from PK assays

Extending the ACeT framework to pharmacokinetics, we trained the architecture to translate a suite of in-vitro PK assay read-outs directly into quantitative predictions of mouse clearance, reported as the area under the plasma concentration-time curve from time 0 to the last quantifiable sample ($AUC_{0\text{-}t}$, hereafter AUCt), the standard measure of drug exposure captured before extrapolation to infinity.

Using paired in-vitro / in-vivo data, ACeT was optimised by cross-validation and then evaluated on a held-out test set. It achieved a high predictive fidelity, with predicted AUCt values closely matching experimental measurements across a ~20-fold range of exposures (Fig. 3a). The test-set predictions showed a strong correlation with observed AUCt ($R^2 = 0.80$) and low error (normalised root-mean-square error, nRMSE = 0.15). Most compounds were predicted with high accuracy: the median percentage error was only ~5.9%, and 75% of test compounds had predicted AUCt within 15% of the in vivo value (inset, Fig. 3a). Only a few outliers exhibited larger deviations (90th percentile error ~41.5%), indicating that the model maintained robust performance for the vast majority of compounds. This level of agreement is remarkable given that the model's inputs are only in vitro measurements, underscoring its potential to surrogate in vivo clearance studies.

Because the original dataset was small, we generated realistic, noise-level synthetic variants of the in-vitro assay profiles to enlarge and regularise the training distribution, and this augmentation dramatically enhanced ACeT's test-set accuracy. Without augmentation, ACeT over-fit the limited data and failed to generalise, yielding poor performance ($R^2 \approx 0.5$) with black square markers in Fig. 3b scattered far from the parity line. In contrast, the model trained with augmented data achieved $R^2 = 0.80$ on the same compounds (red circles in Fig. 3b), effectively restoring the high fidelity noted above. The overlay in Fig. 3b shows that, for every compound, the augmented prediction (red) lies substantially closer to the experimental AUC (unity line) than the unaugmented one (black); for example, a molecule with an actual AUC of ~12,000 ng·h mL^{-1} was under-predicted by >30 % without augmentation, yet predicted within 5 % after augmentation. Overall, data augmentation cut test set error substantially, raising R^2 from 0.5 to 0.8 and lowering nRMSE from ~0.25 to 0.15 (a roughly 1.6-fold increase in R^2 and 40% reduction in nRMSE).

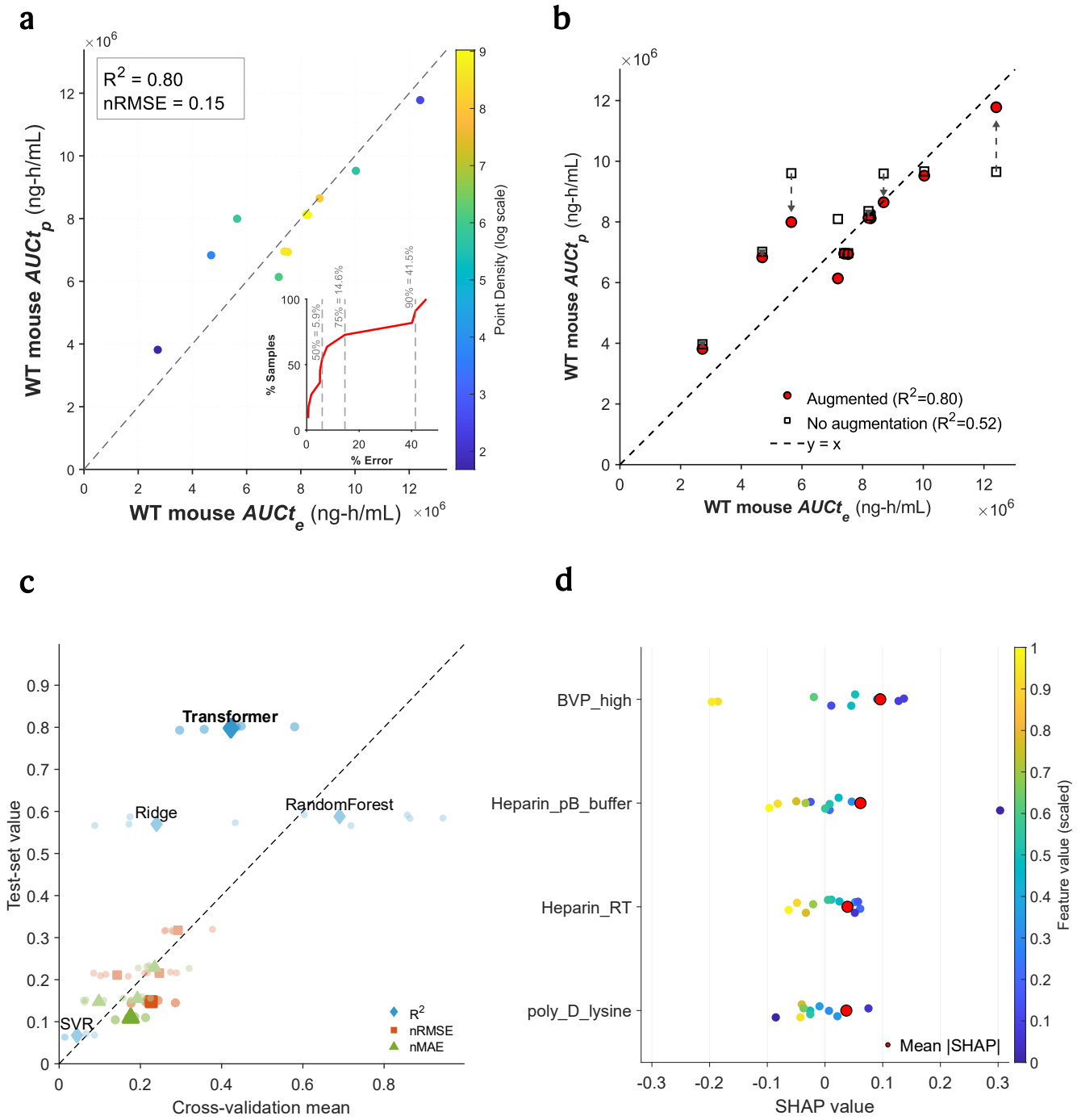

Fig. 3 | ACeT converts in-vitro assay data into quantitative mouse clearance. a, Density-coloured parity plot for the 11-antibody test set (dashed line, unity; R^2=0.80, nRMSE ≈ 0.15); inset, cumulative distribution of absolute-percentage error. **b,** Effect of synthetic augmentation: red circles, augmented model (R^2=0.80); open black squares, unaugmented (R^2=0.52); dashed arrows link paired predictions for each antibody, highlighting the systematic improvement. **c,** Generalization comparison of transformer, ridge, SVR and random-forest regressors. Diamonds, R^2; squares, nRMSE; triangles, nMAE; each marker plots cross-validation mean (x-axis) against test-set value (y-axis); dashed diagonal, y = x. **d,** KernelSHAP beeswarm for the trained transformer; dots, individual SHAP values colored by scaled feature level; red circles, mean |SHAP|, identifying the four dominant assays (BVP_high, Heparin_pB_buffer, Heparin_RT, poly_D_lysine).

By broadening the training distribution with plausible variations of the real profiles (while keeping the test set 100% real), augmentation prevented the model from latching onto noise and allowed it to capture the underlying in vitro-in vivo correlations more faithfully. This highlights the importance of synthetic regularization for reliable performance in this low-*n* regime.

ACeT generalised far better than baseline models (Fig. 3c). Although its cross-validation R^2 was modest (0.42), training on the full dataset and testing on unseen compounds yielded R^2=0.80 with the lowest errors (nRMSE = 0.15; nMAE = 0.10). The held-out test set covers a comparable AUCt range to the training data (272-1241 vs 182-1318; SD 259 vs 244), so the higher test R^2 is consistent with training the final ensemble on the full augmented data rather than an unduly easy split. Because each cross-validation model was fit on ~80% of a small, high-dispersion dataset, the out-of-fold R^2 represents a conservative estimate; ensembling models trained on the full augmented set recovers additional signal, yielding R^2 =0.80 on unseen compounds. Importantly, the scale-normalised errors (nRMSE, nMAE) were lowest on the test set, indicating genuine error reduction rather than variance-driven inflation of R^2. In contrast, a Random Forest regressor overfit (R^2 =0.58), Ridge regression underfit (R^2 =0.56), and SVR underperformed in both settings (R^2 <0.1). Thus, the ACeT framework again provides the most reliable in vitro-to-in vivo mapping.

To clarify ACeT's decision-making and verify alignment with established pharmacokinetic drivers, we applied SHAP (Shapley additive explanations)[49] to the trained model. High values of BVP_high (Baculovirus Particle, high concentration, binding assay) or Heparin_pB_buffer (Heparin percent B buffer elution assay) consistently yielded negative SHAP contributions (yellow points left of centre), indicating that strong signals in these assays drive lower predicted AUCt (i.e. faster clearance). Conversely, low feature values (purple points) shifted SHAP contributions to positive, reflecting higher exposure for compounds with minimal binding or clearance in those assays. Heparin_RT (Heparin chromatography retention time) followed the same trend, strong heparin interactions were associated with reduced AUCt, consistent with accelerated clearance of highly bound molecules. This pattern accords with pharmacological intuition: both assays capture a propensity for nonspecific binding or rapid metabolism, traits that inversely correlate with systemic exposure[24,50]. Finally, poly_D_lysine, an assay for nonspecific binding to a cationic surface, exhibited the same relationship: compounds that bind strongly (high feature value) were predicted to have lower AUCts, whereas weak binders retained higher exposure[41,51].

Collectively, these interpretability results show that ACeT has learned mechanistically relevant relationships: it leverages in-vitro indicators of high clearance to reduce predicted AUCt appropriately. Crucially, no spurious correlations dominate, ACeT's behaviour remains explainable through known PK features, bolstering confidence in its translational utility. Such high-fidelity AUCt estimation offers a practical surrogate for live-animal PK

studies, advancing the 3Rs by prioritising compounds with favourable clearance profiles before in-vivo testing.

Clinical-outcome prediction (Phase I → approval) and portfolio impact

The ACeT-based classifier achieved strong performance for Phase I-to-approval outcomes. Cross-validated training showed no over-fitting, and the held-out internal test set of 23 antibodies yielded a sensitivity of 76.9 % and specificity of 80.0 % (10 / 13 true approvals identified, 8 / 10 terminations correctly flagged), corresponding to a balanced accuracy of ~78 % (Fig. 4a). In practical terms, ACeT captured roughly three-quarters of the eventual approvals while filtering out four-fifths of the likely failures. The classifier's operating point, chosen by maximising the Youden index (sensitivity + specificity − 1), lies near the default 0.5 probability threshold; the bubble confusion plots show only three false-negative approvals and two false-positive terminations.

We benchmarked ACeT's decision boundary against literature-derived developability assays to contextualise its performance (Fig. 4b). Prior studies[8,29] have defined "red-flag" thresholds for individual biophysical read-outs, for example, a poly-specificity reagent (PSR) signal > 0.27 or an AC-SINS self-interaction shift > 11.8 nm, to warn that an antibody may be problematic. Yet no single assay threshold achieved a sensitivity-specificity balance comparable to ACeT's.

Using the previously published rule set[8], where the presence of any single developability flag mandates termination, indeed captures most eventual failures (high sensitivity) but simultaneously discards many antibodies that ultimately progressed to approval (poor specificity). Tightening the criteria, requiring multiple assays to exceed their limits, boosts specificity at the expense of sensitivity, as seen in the drop in recall when moving from one-flag to two-flag triggers (Fig. 4b). By contrast, the trained ACeT model implicitly optimizes a multi-assay threshold: at its operating point it simultaneously reaches ~77 % sensitivity and 80 % specificity on the test set, a balance unattainable by any single-assay rule or fixed cut-off. This underscores that an integrated, learned predictor distinguishes likely "winners" from "losers" far more effectively than individual developability metrics.

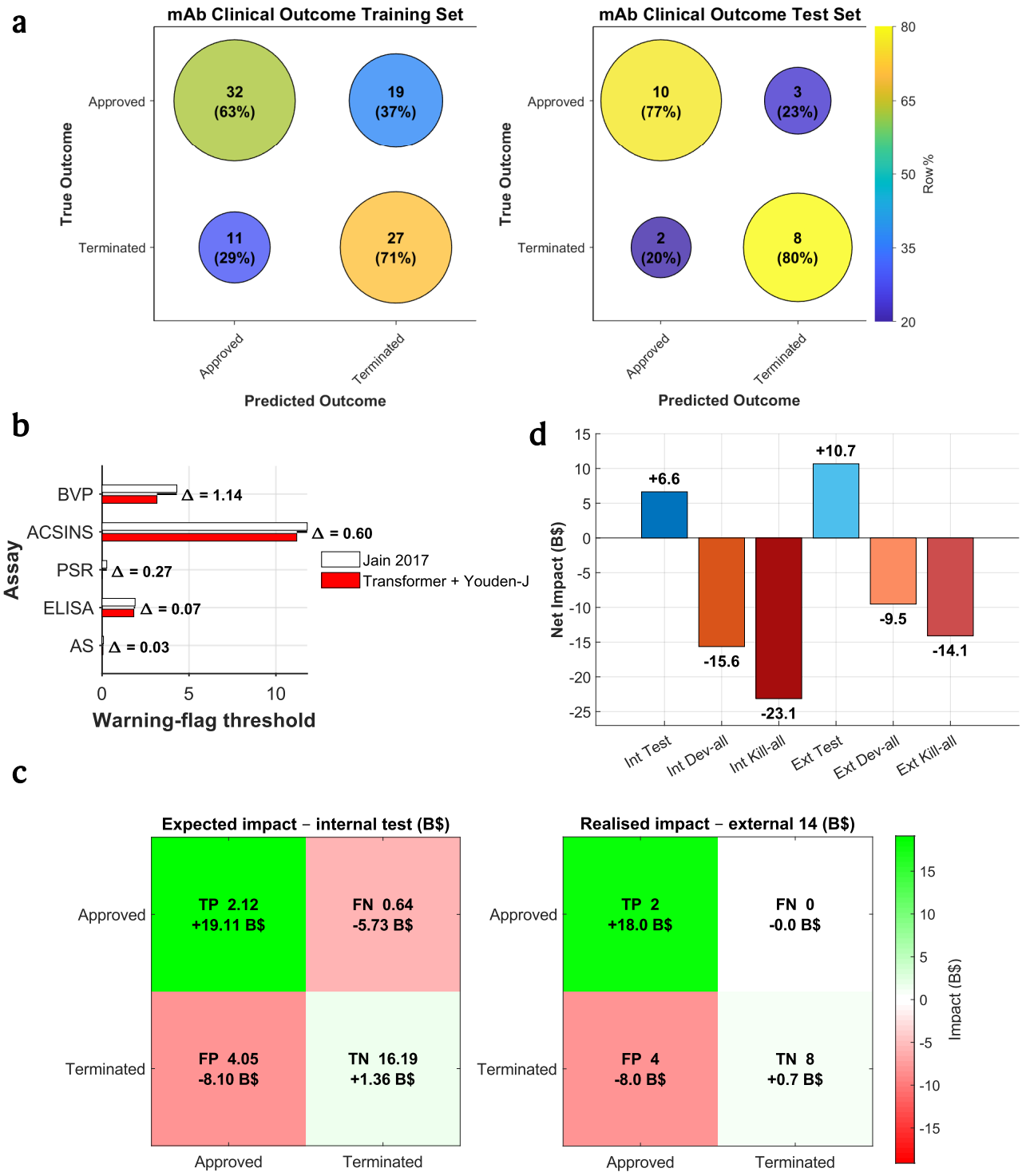

Fig. 4 | ACeT triages Phase-I antibodies and re-allocates development capital. a, Bubble confusion matrices: left, training set; right, held-out internal test set (n = 23). Bubble area scales with sample count and color with row percentage. **b,** Single-assay 'red-flag' thresholds from the 2017 rule set (white) versus ACeT's Youden-optimized cut-offs (red); horizontal brackets annotate the offset Δ. **c,** Economic heat maps converting confusion cells to NPV (green = gain, red = loss); expected impact for the internal test set (left) and realised impact for the 14-mAb external cohort (right). **d,** Portfolio-level NPV for three development policies, ACeT triage, develop-all and kill-all, applied to the internal cohort (left trio) and the external cohort (right trio); bars are labelled with net billions (US$).

Starting from ACeT's internal confusion matrix (13 historical approvals, 10 historical failures), we mapped each cell to a financial consequence using accepted biopharma benchmarks:

Table 1 | Per-asset economic assumptions for monetizing ACeT's Phase I predictions. The Phase I probability-of-success (PoS = 12 %)[1,4] is applied to the 23-mAb internal cohort and then multiplied by ACeT's sensitivity/specificity to give fractional counts (e.g. TP = 23 × 0.12 × 0.769 ≈ 2.12). Each outcome is assigned a net-present-value: +$9 B for a launch (TP)[52], −$9 B for a missed launch (FN)[7,52], −$2 B for a continued failure (FP)[7] and +$0.084 B for an early kill (TN)[7]. These figures feed directly into the heat-map shown in Fig. 4c.

Outcome	Abbrev.	Assumed phase-I prior	Per-asset value	Rationale
Launch success	TP (true positive)	12 %→fractional 2.12	+$9 B	$11 B fully capitalized mean net-present value (NPV) minus $2 B R&D cost
Missed success	FN (false negative)	0.64	−$9 B	Opportunity cost of wrongly killing a future launch
Continued failure	FP (false positive)	4.05	−$2 B	Sunk Phase II/III spend with no return
Early kill	TN (true negative)	16.19	+$0.084 B	Median Phase II/III savings ($25 M + $59 M)

Multiplying each fractional count by its per-asset NPV yielded the colour-scaled matrix (green = gain, red = loss) for the internal test subset of 23 mAbs whose clinical outcomes were already resolved (approved or terminated) by October 2022[8,29] (Fig. 4c, left): +$19.11 B (TP), −$5.73 B (FN), −$8.10 B (FP) and +$1.36 B (TN). Summed, ACeT delivers an expected net of +$6.6 B for this 23-mAb portfolio, despite the low 12 % base rate of success. For context, a develop-all approach would spend $46 B and recoup only $30 B, producing −$15.6 B, while a kill-all strategy would save $1.9 B but forfeit $27 B in upside (−$23.1 B) as shown in Fig. 4d. Thus, ACeT captures ~90 % of the potential value while avoiding ~80 % of wasted expenditure. Even modest improvements in false-positive control translate into billions: each avoided FP saves the full late-stage burn; each rescued TP secures nine billion dollars in lifetime value. ACeT's operating point (sensitivity ≈ 0.77, specificity ≈ 0.80) is therefore near-optimal for maximizing expected net present value under realistic priors.

The external validation set comprised 14 mAbs that were still progressing through Phase I-III at the 2022 cut-off[8,29]; by May 2025 their fates had become clear, two ultimately gained approval, whereas twelve were discontinued. Without any retraining, ACeT produced 2 TP, 0 FN, 4 FP, 8 TN (balanced accuracy = 0.83) (Fig. 4c, right). Applying the same cost analysis (Fig. 4c, right), but now to actual rather than expected counts, yielded a net realized gain (Fig. 4d) of +$10.7 B, a swing of $20.2 B versus develop-all (−$9.5 B) and $24.8 B versus kill-all (−$14.1 B). The bar plot in Fig. 4d juxtaposes the net values for both cohorts, underscoring that ACeT's economic advantage persists across the roughly two-and-a-half-year interval between the 2022 outcome freeze and the May 2025 external read-out, and within a set of molecules the model never encountered during development.

Varying the clinical-success prior between 8 % and 20 % and scaling the fully capitalized launch NPV by ±30 % changes the absolute dollar values but never the hierarchy: ACeT stays at least ≈ $6 B ahead of a develop-all policy and ≈ $9 B ahead of a kill-all policy across the entire sensitivity grid (data not shown). By feeding micro-scale biophysical features through ACeT's learned decision boundary, a sponsor can shift from a baseline expectation of multi-billion-dollar losses to multi-billion-dollar gains, first in silico (expected) and then in reality (realised). Panels 4c and 4d therefore extend ACeT's statistical superiority into the language that matters to executives and investors: net present value.

Discussion

The agnostic context-embedding transformer (ACeT) shows that a single, schema-free attention encoder coupled to a plug-and-play prediction stage can learn developability rules that elude specialised architectures restricted to fixed input schemas. While tabular deep-learning models such as TabNet[53], FT-Transformer[54] and TabPFN[55] advance column-wise attention or foundation-model pre-training, they require retraining or redesign when assay panels change and cannot readily switch inductive biases across tasks. ACeT instead accepts arbitrary feature sets without manual redesign and selects the most suitable prediction head on demand, providing a complementary solution rather than a direct replacement. In the antibody arena, pretrained language models[34,36,56] (e.g. AntiBERTy, IgBERT, IgT5) excel at sequence representation but still require handcrafted pooling or downstream regressors to combine heterogeneous biophysical read-outs; ACeT fills this methodological gap by natively fusing assay data. Crucially, its tokenisation scheme remains compatible with Shapley additive explanations, allowing every prediction to be decomposed into assay-level attributions that meet emerging regulatory expectations for transparency[49].

This interpretability reveals physically consistent drivers of high-concentration viscosity. Dynamic light-scattering interaction parameter kD dominates the model's attention, echoing canonical reports[15,45,46] that increasingly negative kD values trigger attractive self-association and exponential viscosity rises in the >100 mg mL^{-1} regime. ACeT

therefore reproduces, from dilute-solution data alone, the 15-20 cP soft ceiling above which subcutaneous administration becomes impractical, a heuristic first quantified in large-scale colloidal studies and later formalized in guidelines for formulation scientists[13,57]. On a contemporary 75-mAb panel, the transformer achieves a test-set R^2 in the mid-0.7 range on a hold-out cohort representing roughly one-third of the data. It outperforms ridge, random-forest and support-vector baselines by more than 25 % while retaining accuracy at both viscosity extremes, thus surpassing earlier machine-learning studies that were trained on only a few dozen antibodies and reported categorical pass-fail outcomes without absolute error metrics[58]. Ablation of kD collapses performance, supporting molecular-dynamics evidence[59] that short-range intermolecular interactions become dominant at >100 mg mL^{-1}, and the unbeaten learning curve indicates that further gains are likely if sponsors federate viscosity data rather than silo them.

The same encoder delivers strong performance in the pharmacokinetic domain: when trained on a quartet of standard in-vitro assays, ACeT provides highly reliable estimates of mouse intravenous clearance, surpassing contemporary physiologically based PK approaches that must be laboriously recalibrated for every new antibody[51]. SHAP analysis confirms that the predictions hinge on assay read-outs classically associated with nonspecific, charge-mediated interactions, lending mechanistic credibility and echoing recent reports that electrostatic surface properties drive FcRn-independent clearance differences among otherwise isotype-matched antibodies[60,61] .Data augmentation via noise-level perturbations more than halved prediction error, mirroring the benefits reported for transformer pretraining on small tabular datasets[55] and emphasizing that synthetic regularization can rescue generalization when $n \ll 200$. Looking forward, cross-pharma federated learning initiatives such as MELLODDY illustrate how privacy-preserving gradient aggregation could unlock the thousands of proprietary PK datapoints needed to smooth residual heteroscedasticity[62,63].

Perhaps the most commercially consequential finding is ACeT's capacity to forecast which Phase-I antibodies will progress to approval. By integrating a focused, experimentally tractable panel of developability assays, the model achieves balanced accuracy in the high-seventies on historical candidates, substantially outperforming single-threshold heuristics such as polyspecificity-reagent or self-interaction cut-offs, which either cull promising leads or allow liabilities to advance[8,29]. Translating its confusion-matrix counts into net-present value under widely accepted R&D economics yields a portfolio swing of roughly US$ 17 billion for a 37-mAb programme compared with a "develop-all" policy; per-asset costs are conservatively bracketed between the Congressional Budget Office's fully capitalized US$ 1-2 billion range[64] and the sector-wide median of US$ 1 billion reported by cross-sector analyses[6]. These gains illustrate how even moderate improvements in statistical precision compound into outsized financial value: each avoided late-stage failure frees nine-figure capital, while every rescued launch secures multi-billion-dollar lifetime revenue.

More broadly, ACeT reshapes capital-allocation logic in early-stage biotherapeutics. Traditional go/no-go decisions rely on fixed cut-offs for isolated assays, treating Phase-I programmes as speculative gambles; ACeT replaces these static rules with a learned decision boundary that converts high-uncertainty trials into transparent, risk-adjusted bets. By identifying a handful of critical failures for early termination and prioritizing the most promising successes, the model moves investment from low-yield to high-yield assets, turning statistical gains into board-room currency. Its robustness on antibodies whose outcomes became known only in 2025 further attests that the learned boundary captures fundamental determinants of clinical success, ensuring strategic relevance as the development landscape evolves, even if future transparency eventually compresses absolute development costs.

Yet the framework is not without boundaries. The present study focuses on IgG scaffolds; extending to bispecifics or Fc-engineered variants may require fine-tuning or domain adaptation, although ACeT's agnostic tokenisation readily accepts categorical identifiers for new modalities. Because the model was trained on assays generated under specific experimental protocols, deployment in other laboratories, or on emerging assay formats, may first require recalibration or transfer learning to absorb systematic measurement differences. Causal extrapolation from murine pharmacokinetics or 150 mg mL^{-1} rheology to human pharmacodynamics and formulation at 200 mg mL^{-1} likewise demands careful calibration, potentially by fusing ACeT's latent vectors with residue-level embeddings from frontier antibody language models that capture subtle paratope chemistries[36]. Finally, while SHAP provides local interpretability, global causal inference will benefit from counterfactual sensitivity analyses, and multi-site federated training must still satisfy emerging ethical frameworks for distributed medical AI[65].

Despite these caveats, ACeT demonstrates that early-stage assay data can now be leveraged more effectively than previously shown[8,15]: it shortens experimental cycles, reduces animal use and may redirect substantial resources toward molecules with a higher probability of clinical and commercial success. By bridging the methodological divide between tabular deep learning and antibody-specific language models, and by pairing predictive accuracy with interpretability and economic analysis, ACeT represents a substantial advance in computational developability assessment and offers a proof of concept for next-generation platforms.

Methods

Dataset preparation and train-test splitting

Viscosity dataset preparation

The high-concentration viscosity corpus comprised 82 IgG1 monoclonal antibodies[41] assayed across 110-200 mg mL^{-1} (most at ~150 mg mL^{-1}). Quality control removed seven entries, three statistical outliers exceeding 100 cP (practical syringeability and statistical

stability[10,66]) and four profiles with > 20 % missing assay values, yielding 75 complete descriptor-response pairs for subsequent composite-stratified splitting and model development. Assay features included, among others, the diffusion interaction parameter *k*D (measured by DLS at ~1 mg mL^{-1}) as a proxy for colloidal stability, as well as chromatographic and biophysical readouts (e.g., SEC peak metrics and AC-SINS). The > 20 % missingness threshold was a data-quality safeguard and does not imply every assay must be measured at inference; in practice, the final viscosity model uses only a small subset of informative assays (the four features with the highest absolute Spearman correlation with viscosity), so only those inputs are required for prediction.

Mouse-clearance dataset preparation

The clearance corpus comprised 55 IgG1 antibodies extracted from a 2023 murine pharmacokinetic study[51]; two extreme AUCt outliers (1340.8 and 157.18 ng·h mL^{-1}) were eliminated, leaving 53 profiles. The four in-vitro assays exhibiting the highest absolute Spearman correlation with AUCt were retained, after which all features were rank-normalized to N(0,1); values lying beyond ±5 standard deviations were multiply imputed through a ten-iteration MICE protocol, preserving the target column. The resulting table was input directly into the composite-stratified splitter.

Composite-stratified algorithm for viscosity and mouse clearance

To derive representative yet non-overlapping partitions we developed a composite-stratified splitting procedure in which the descriptor matrix was first reduced by principal-component analysis (≤ 5 components) and tessellated with k-means ($3 \leq k \leq 8$). The continuous target, neat viscosity (cP) or plasma AUCt, was independently divided into b equiprobable quantile bins ($1 \leq b \leq 5$). Concatenating the cluster label with the target-bin label yielded a composite tag that encodes joint membership in feature and response space; a single stratified shuffle split on these tags produced the final train and test subsets. All (k, b) pairs were scored by (i) the fraction of test samples lacking a near neighbour in the training set (out-of-distribution rate) and (ii) the mean train-to-test nearest-neighbour distance. The configuration with the lexicographically minimal (OOD-rate, distance) was retained, resulting in a 70: 30 split for viscosity and an 80: 20 split for clearance and giving markedly lower variance in R^2 and Spearman ρ than conventional continuous stratification.

Clinical-outcome cohorts

From the 137-antibody biophysical landscape[8,29], we analysed two cohorts distinguished by outcome timing rather than provenance. The internal cohort (outcome-locked 2022) comprised 112 IgG1s with definitive regulatory outcomes recorded on or before October 2022; a single 80: 20 train_test_split with stratify=y was performed, and the test partition served as the development hold-out. The external cohort (temporal validation) comprised 25 antibodies still in Phase I-III at that cutoff; 14 of these had reached an outcome by May 2025 and were evaluated as an intact, out-of-time set, while 11 remain under observation. The classifier was fit on the training partition of the outcome-locked

2022 cohort and applied unchanged to both the development hold-out and the temporal validation cohort to assess prospective-style performance.

Synthetic augmentation

Synthetic augmentation for the two regression endpoints was performed with the Gaussian Copula Synthesizer (SDV 1.11.0), instantiated with default distribution 'norm' so that each empirical marginal was first rank-mapped onto N (0,1) before maximum-likelihood estimation of the correlation matrix; higher-order regularization was disabled to preserve tail dependence among assays. For the viscosity corpus the synthesizer drew one artificial observation per two real samples (1: 2), whereas a ratio of 2: 1 was adopted for the clearance corpus. These synthetic rows were concatenated with the native training frames prior to any scaling or imbalance remediation, thereby allowing the subsequent scalers to learn from the augmented marginal distributions and eliminating the possibility of leakage from test data. No synthetic augmentation was applied to the clinical-outcome classifier.

Feature transformation, imbalance control and model-selection workflow

After composite-stratified splits had been fixed, the regression and classification pipelines shared a uniform, post-split processing strategy designed to respect the minute sample sizes yet minimize information leakage.

Regression pipelines (viscosity and mouse clearance)

All numerical predictors, native plus Gaussian-copula synthetics, were cast to 32-bit floats and transformed with a Quantile Transformer (1000 quantiles, uniform output) fitted solely on the augmented training matrix; the fitted map was applied verbatim to the held-out test set, ensuring strictly out-of-sample scaling. Within each of the five cross-validation folds the so-scaled training subset was passed to the Imbalanced-Learning-Regression (IBLR 0.2) routine, which automatically chooses between an Edited-Nearest-Neighbours (ENN) under-sampler and a SMOTE over-sampler, or their cascade in either order, according to fold-specific neighbourhood densities. This order-agnostic design guarantees that the minority tail of the continuous endpoint is adequately represented while avoiding artefactual duplication when local sparsity is high. Target vectors remained on their original physical scale; a per-fold Quantile Transformer was used only to stabilize the loss surface during training and was inverted once after prediction.

Clinical classification pipeline

For the 112-antibody internal cohort, continuous assay read-outs were variance-stabilized with a Yeo-Johnson Power Transformer fitted on the training partition; no synthetic rows were introduced. Each fold then executed an order-free ENN⇄SMOTE block identical in spirit to the IBLR procedure: ENN first attempted to excise borderline points, the class distribution was re-counted, and SMOTE was applied only if the cleaned minority class still contained at least two examples. The SMOTE k-neighbour parameter was adaptively set to

min (5, n_{min}−1) to respect the updated class count. Detailed class tallies before and after each operation were printed to the console and logged for audit. Yeo-Johnson PowerTransformer was reused for both the internal test set and the 14-antibody external validation cohort; categorical outcome labels were integer-encoded before one-hot expansion for soft-max training.

Prediction-head benchmarking and ensemble derivation

Five modular heads, multilayer perceptron (MLP), radial-basis function (RBF), Kolmogorov-Arnold network (KAN), cubic spline and pairwise interaction, were benchmarked under an identical five-fold outer cross-validation loop. The head maximizing mean R^2 (regression) or balanced accuracy (classification) was declared the best head. The five best-head checkpoints (one per fold) constituted an ensemble predictor; test-set estimates were obtained by arithmetic averaging of scalar outputs (regression) or class-probability vectors (classification). This ensembling preserves the variance-reduction benefits of cross-validation while retaining a single, reproducible hyper-parameter configuration for all reported metrics.

ACeT training and evaluation

Each task-specific ACeT model was trained under five-fold stratified cross-validation (folds shuffled once). Architectural constants were fixed: 16-dimensional token embeddings, two self-attention heads, a 32-unit feed-forward sub-layer, one transformer block, 0.3 dropout, $L_2=10^{-3}$ weight decay, and a 64-unit hidden layer in the read-out multilayer perceptron. Optimization employed Adam (initial learning rate 1×10^{-3}, mini-batch 32) for ≤1000 epochs; early stopping halted training after 20 stagnant epochs, while ReduceLROnPlateau halved the rate after ten stagnant epochs, with a floor of 1×10^{-6}. Regression heads minimized log-cosh loss; the clinical classifier minimized binary cross-entropy. Python, NumPy and TensorFlow seeds were aligned to guarantee bit-wise determinism. In each fold the epoch that maximized the validation metric, R^2 for regression or balanced accuracy for classification, was checkpointed, and the five checkpoints were averaged to yield the ensemble predictor evaluated once on the held-out test set. Internal hyper-parameters specific to the benchmarked heads (e.g. RBF length-scale) were retained at the empirically optimal defaults established during cross-validation; a future release can substitute this manual tuning with Bayesian optimization via Optuna.

Statistical analysis

Hold-out performance for viscosity and clearance was quantified with R^2, root-mean-square error, mean absolute error and Spearman's ρ. Pair-wise method comparisons used a two-tailed exact permutation test on absolute errors, as recommended for small biomedical datasets.

Model interpretability

Feature relevance for the viscosity model was quantified with permutation-feature importance: the five-checkpoint ACeT ensemble, exposed through a scikit-learn wrapper

that averages predictions and inversely scales them, was evaluated ten times while permuting each input column, and the mean drop in R^2 served as the importance score. For the mouse-clearance model, KernelSHAP was applied to every checkpoint using 100 representative background points; the resulting tensors (fold × sample × feature) were stacked and averaged, and features were ranked by the grand mean of the absolute Shapley values. All raw importance vectors, SHAP matrices and accompanying figures were archived to enable independent verification.

Software and hardware

All code was executed on a Windows 11 Pro (23H2) workstation equipped with a single NVIDIA A100 GPU (40 GB), CUDA 12.1 and cuDNN 8.9. The Python stack comprised TensorFlow 2.15.0 with Keras 2.15.0, tf-kan 0.1.0, scikit-learn 1.4.1 post 1, imbalanced-learn 0.12.2, SDV 1.17.4 (Copulas 0.12.1, RDT 1.14.0, SDMetrics 0.18.0) and SHAP 0.45.0. Core numerics relied on NumPy 1.23.5, SciPy 1.15.2 and pandas 2.2.2, while visualizations were drafted in Matplotlib 3.9.2 and finalized in MATLAB R2022b.

Data and code availability

All raw and processed datasets analysed in this study were obtained from published sources (Refs. 8,29,41,51) and will be deposited in the corresponding author's public GitHub repository upon publication. This repository will also host all Python scripts required to reproduce the analyses. No material-transfer agreements are necessary.

Disclosure statement

The authors report there are no competing interests to declare.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ORCID

Sabitoj Singh Virk http://orcid.org/0000-0002-5801-5630
Akashdeep Singh Virk http://orcid.org/0000-0002-9157-4637

References

1. Hay M., Thomas D.W., Craighead J.L., Economides C., Rosenthal J. Clinical development success rates for investigational drugs. *Nat. Biotechnol.* **32**, 40-51 (2014).
2. Thomas D.W. Clinical development success rates 2006-2015. *BIO Ind. Anal.* **1**, 16 (2016).
3. Wong C.H., Siah K.W. & Lo A.W. Estimation of clinical trial success rates and related parameters. *Biostatistics* **20**, 273-286 (2019).
4. BIO | QLS Advisors | Informa UK Ltd. Clinical development success rates and contributing factors 2011-2020. *BIO / QLS / Informa* (2021).
5. Crescioli S. et al. Antibodies to Watch in 2024. *mAbs* **16**, 2297450 (2024).
6. Wouters O.J., McKee M., Luyten J. Estimated Research and Development Investment

Needed to Bring a New Medicine to Market, 2009-2018. *JAMA* **323**, 844-853 (2020).
7. DiMasi J.A., Grabowski H.G., Hansen R.W. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J. Health Econ.* **47**, 20-33 (2016).
8. Jain T. et al. Biophysical properties of the clinical-stage antibody landscape. *Proc. Natl. Acad. Sci. USA* **114**, 944-949 (2017).
9. Xu Y. et al. Structure, heterogeneity and developability assessment of therapeutic antibodies. *mAbs* **11**, 239-264 (2019).
10. Berteau C. et al. Evaluation of the impact of viscosity, injection volume, and injection flow rate on subcutaneous injection tolerance. *Med. Devices Evid. Res.* **8**, 473-484 (2015).
11. Garidel P., Kuhn A.B., Schäfer L.V., Karow-Zwick A.R. & Blech M. High-concentration protein formulations: how high is high? *Eur. J. Pharm. Biopharm.* **119**, 353-360 (2017).
12. Yadav S., Laue T.M., Kalonia D.S., Singh S.N. & Shire S.J. The influence of charge distribution on self-association and viscosity behavior of monoclonal antibody solutions. *Mol. Pharm.* **9**, 791-802 (2012).
13. Tomar D. S., Kumar S., Singh S. K., Goswami S. & Li L. Molecular basis of high viscosity in concentrated antibody solutions: Strategies for high concentration drug product development. *mAbs* **8**, 216-228 (2016).
14. Raybould M.I.J. et al. Five computational developability guidelines for therapeutic antibody profiling. *Proc. Natl. Acad. Sci. USA* **116**, 4025-4030 (2019).
15. Kingsbury, J.S. et al. A single molecular descriptor to predict solution behavior of therapeutic antibodies. *Science Advances* **6**, eabb0372 (2020).
16. Virk S.S., Underhill P.T. Application of a simple short-range attraction and long-range repulsion colloidal model toward predicting the viscosity of protein solutions. *Mol. Pharm.* **19**, 4233-4240 (2022).
17. Tang E.M., Virk S.S., Underhill P.T. Coupling between long ranged repulsions and short ranged attractions in a colloidal model of zero shear rate viscosity. *J. Rheol.* **66**, 491–504 (2022).
18. Roopenian D.C. & Akilesh S. FcRn: the neonatal Fc receptor comes of age. *Nat. Rev. Immunol.* **7**, 715-725 (2007).
19. Zalevsky J. et al. Enhanced antibody half-life improves in vivo activity. *Nat. Biotechnol.* **28**, 157-159 (2010).
20. Acqua W.F.D. et al. Increasing the affinity of a human IgG1 for the neonatal Fc receptor: biological consequences. *J. Immunol.* **169**, 5171-5180 (2002).
21. Boswell C.A. et al. Effects of charge on antibody tissue distribution and pharmacokinetics. *Bioconjug. Chem.* **21**, 2153-2163 (2010).
22. Schoch A. et al. Charge-mediated influence of the antibody variable domain on FcRn-dependent pharmacokinetics. *Proc. Natl. Acad. Sci. USA* **112**, 5997-6002 (2015).
23. Tabrizi M.A., Tseng C.M.L. & Roskos L.K. Elimination mechanisms of therapeutic monoclonal antibodies. *Drug Discov. Today* **11**, 81-88 (2006).
24. Kraft T. E. et al. Heparin chromatography as an in vitro predictor for antibody clearance rate through pinocytosis. *mAbs* **12**, 1683432 (2020).
25. Liu Y. et al. High-throughput screening for developability during early-stage antibody discovery using self-interaction nanoparticle spectroscopy. *mAbs* **6**, 483-492 (2014).

26. Wolf Pérez A.M. et al. In vitro and in silico assessment of the developability of a designed monoclonal antibody library. *mAbs* **11**, 388-400 (2019).
27. Arsiwala A. et al. A high-throughput platform for biophysical antibody developability assessment to enable AI/ML model training. *bioRxiv* **2025**, 2025-05 (2025).
28. Jain T. et al. Assessment and incorporation of in vitro correlates to pharmacokinetic outcomes in antibody developability workflows. *mAbs* **16**, 2384104 (2024).
29. Jain T., Boland T. & Vásquez, M. Identifying developability risks for clinical progression of antibodies using high-throughput in vitro and in silico approaches. *mAbs* **15**, 2200540 (2023).
30. Rao R. et al. Evaluating protein transfer learning with TAPE. *Adv. Neural Inf. Process. Syst.* **32** (2019).
31. Raybould M.I. et al. Thera-SAbDab: the therapeutic structural antibody database. *Nucleic Acids Res.* **48**, D383-D388 (2020).
32. Jumper J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
33. Rives A. et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc. Natl. Acad. Sci. USA* **118**, e2016239118 (2021).
34. Ruffolo J. A., Chu L. S., Mahajan S. P. & Gray, J. J. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. *Nat. Commun.* **14**, 2389 (2023).
35. Hie B. L. et al. Efficient evolution of human antibodies from general protein language models. *Nat. Biotechnol.* **42**, 275-283 (2024).
36. Kenlay H. et al. Large scale paired antibody language models. *PLoS Comput. Biol.* **20**, e1012646 (2024).
37. WHO. Proposed International Nonproprietary Names (INN) List 120. *WHO Drug Inf.* **32**, 559-689 (2018).
38. WHO. Recommended International Nonproprietary Names (INN) List 81. *WHO Drug Inf.* **33**, 59-134 (2019).
39. Krawczyk K., Buchanan A. & Marcatili P. Data mining patented antibody sequences. *mAbs* **13**, 1892366 (2021).
40. Lee S.H., Lee S. & Song B.C. Vision transformer for small-size datasets. *arXiv preprint* arXiv:2112.13492 (2021).
41. Mock M. et al. Development of in silico models to predict viscosity and mouse clearance using a comprehensive analytical data set collected on 83 scaffold-consistent monoclonal antibodies. *mAbs* **15**, 2256745 (2023).
42. Makowski E. et al. Optimization of therapeutic antibodies for reduced self-association and non-specific binding via interpretable machine learning. *Nat. Biomed. Eng.* **8**, 45-56 (2024).
43. Grinshpun B. et al. Identifying biophysical assays and in silico properties that enrich for slow clearance in clinical-stage therapeutic antibodies. *mAbs* **13**, 1932230 (2021).
44. Snyder S.H. et al. The Goldilocks paradigm: comparing classical machine learning, large language models, and few-shot learning for drug discovery applications. *Commun. Chem.* **7**, 134 (2024).

45. Dippel A. et al. Developability profiling of a panel of Fc engineered SARS-CoV-2 neutralizing antibodies. *mAbs* **15**, 2152526 (2023).
46. Eskens O. N. et al. Temperature and excipient mediated modulation of monoclonal antibody interactions revealed by k D, rheology, and Raman spectroscopy. *mAbs* 17, 2521511 (2025).
47. Hong P., Koza S. & Bouvier E. S. A review size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates. *J. Liq. Chromatogr. Relat. Technol.* **35**, 2923-2950 (2012).
48. Sule S. V., Dickinson C. D., Lu J., Chow C. K. & Tessier P. M. Rapid analysis of antibody self-association in complex mixtures using immunogold conjugates. *Mol. Pharm.* **10**, 1322-1331 (2013).
49. Lundberg S. M. & Lee, S. I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **30** (2017).
50. Hötzel I. et al. A strategy for risk mitigation of antibodies with fast clearance. *mAbs* **4**, 753–760 (2012).
51. Liu S. et al. Utility of physiologically based pharmacokinetic modeling to predict inter-antibody variability in monoclonal antibody pharmacokinetics in mice. *mAbs* **15**, 2263926 (2023).
52. Evaluate Ltd. *World Preview 2024, Outlook to 2030: Biopharma's Answer to the IRA* (2024).
53. Arik S.Ö. & Pfister T. TabNet: Attentive interpretable tabular learning. *Proc. AAAI Conf. Artif. Intell.* **35**, 6679-6687 (2021).
54. Gorishniy Y., Rubachev I., Khrulkov V. & Babenko A. Revisiting deep learning models for tabular data. *Adv. Neural Inf. Process. Syst.* **34**, 18932–18943 (2021).
55. Hollmann N. et al. Accurate predictions on small data with a tabular foundation model. *Nature* **637**, 319-326 (2025).
56. Ruffolo J. A., Gray J. J. & Sulam J. Deciphering antibody affinity maturation with language models and weakly supervised learning. *arXiv* preprint arXiv:2112. 07782 (2021).
57. Sharma V.K. et al. In silico selection of therapeutic antibodies for development: viscosity, clearance, and chemical stability. *Proc. Natl. Acad. Sci. USA* **111**, 18601-18606 (2014).
58. Lai P. K., Gallegos A., Mody N., Sathish H. A. & Trout, B. L. Machine learning prediction of antibody aggregation and viscosity for high concentration formulation development of protein therapeutics. *mAbs* **14**, 2026208 (2022).
59. Prass T. M., Garidel P., Blech M. & Schafer L. V. Viscosity prediction of high-concentration antibody solutions with atomistic simulations. *J. Chem. Inf. Model* **63**, 6129-6140 (2023).
60. Makowski E. K. et al. Co-optimization of therapeutic antibody affinity and specificity using machine learning models that generalize to novel mutational space. *Nat. Commun.* **13**, 3788 (2022).
61. Ausserwöger H. et al. Surface patches induce nonspecific binding and phase separation of antibodies. *Proc. Natl. Acad. Sci. USA* **120**, e2210332120 (2023).
62. Heyndrickx W. et al. Melloddy: Cross-pharma federated learning at unprecedented scale unlocks benefits in qsar without compromising proprietary information. *J. Chem. Inf. Model* **64**, 2331-2344 (2023).

63. Khanna A. et al. Accelerating Parkinson's Disease drug development with federated learning approaches. *npj Parkinsons Dis.* **10**, 225 (2024).
64. Congressional Budget Office. *Research and Development in the Pharmaceutical Industry*. Report 57025 (2021).
65. Hähnel M. Ethical challenges and solutions in AI-driven medical data management: a focus on distributed machine learning. *Discov Artif Intell* **5**, 53 (2025).
66. Desai M., Kundu A., Hageman M., Lou H. & Boisvert D. Monoclonal antibody and protein therapeutic formulations for subcutaneous delivery: high-concentration, low-volume vs. low-concentration, high-volume. *mAbs* **15**, 2285277 (2023).