

REPORT



Predictive modeling of concentration-dependent viscosity behavior of monoclonal antibody solutions using artificial neural networks

Jonathan Schmitt, Abbas Razvi, and Christoph Grapentin

Lonza AG/Ltd, Drug Product Services, Basel, Switzerland

ABSTRACT

Solutions of monoclonal antibodies (mAbs) can show increased viscosity at high concentration, which can be a disadvantage during protein purification, filling, and administration. The viscosity is determined by protein-protein-interactions, which are influenced by the antibody's sequence as well as solution conditions, like pH, buffer type, or the presence of salts and other excipients. To predict viscosity, experimental parameters, like the diffusion interaction parameter (kD), or computational tools harnessing information derived from primary sequence, are often used, but a reliable predictive tool is still missing. We present a modeling approach employing artificial neural networks (ANNs) using experimental factors combined with simulation-derived parameters plus viscosity data from 27 highly concentrated (180 mg/mL) mAbs. These ANNs can be used to predict if mAbs exhibit problematic viscosity at distinct concentrations or to model viscosity-concentration-curves.

ARTICLE HISTORY

Received 21 October 2022
Revised 11 January 2023
Accepted 12 January 2023

KEYWORDS

Monoclonal antibody; high concentration; viscosity; artificial neural network; predictive modeling

Introduction


Therapeutic monoclonal antibodies (mAbs) are now commonly used as a treatment for a broad variety of diseases, including cancer, immune-mediated disorders, or infectious diseases.¹ They are typically administered via intravenous infusion, which requires the drug product (DP) to be administered by a healthcare professional in a clinical setting. For patients with chronic diseases, the need for repeated drug infusions is inconvenient and time-consuming, which puts the success of the intended therapy at risk. Subcutaneous (s.c.) injection allows patients to self-administer mAb DPs by use of pre-filled syringes, auto-injectors or other delivery devices, which often increases quality of life and compliance for patients with chronic conditions. There are, however, certain limitations with s.c. administration. By general consensus, a single injection volume should be limited to < 2 mL, determined by the available subcutaneous space and sensation of tolerable pain by the patient,² although investigation of injection volumes >2.5 mL has been suggested.³

Although mAbs typically have a high specificity, they also require considerable therapeutic dosages. This consequently often results in high concentrations exceeding 100 mg/mL protein in solutions for s.c. administration. With increasing protein concentration, inter-molecular distances reduce and protein-protein-interactions (PPIs) do not increase linearly but exponentially, influencing or determining mAbs' solubility, aggregation, and also viscosity.^{4,5} PPIs are affected by a mAb's sequence and resulting three-dimensional structure with charged or hydrophobic patches. Solution conditions can influence PPIs by modulating the size of charged patches via pH, shielding of charged patches via short-ranged electrostatic interaction using salts, buffer substances, amino acids, or

other charged excipients.^{6–8} Arginine is a common excipient tested for viscosity reduction, its dual mode of action being both the shielding of charged as well as of hydrophobic patches.⁹ Of 34 US Food and Drug Administration (FDA)-approved DPs with high mAb concentration, 17 use salts or amino acids as excipients, likely with the aim to reduce PPIs and thus lower mAb solution viscosity.¹⁰ More explorative excipients with demonstrated potential for viscosity reduction are poly-L-glutamic acid,¹¹ caffeine,¹² hydrophobic salts,¹³ or amino acid derivatives,¹⁴ but no application in commercialized drug products exists due to lack of approval as excipients for parenteral administration or concerns on toxicity.

Highly viscous solutions can be a major roadblock in the development of an mAb DP. Disadvantages are high costs,⁴ due to high loss and low recovery in purification, difficult manufacturing or filling,¹⁵ and ultimately issues during product administration due to the need for high injection forces and slow administration with potential sensation of pain. In general, solutions with a dynamic viscosity above 15–20 mPa*s may be considered problematic.^{16,17} The desire to develop a high concentration formulation may not be apparent a priori for a new molecule; it may appear also as a consequence of the need for unexpectedly high doses or change of target route of administration. Phase 1 clinical trials are usually started with lower concentrated (<50 mg/mL protein) formulations and high protein concentrations are only explored in later phases once safe and efficacious dose levels are established. To evaluate Chemistry, manufacturing and control (CMC) issues related to DP later in development when dose ranges are established, it is essential to forecast the viscosity of a new molecule during an early development stage.

CONTACT Christoph Grapentin  christoph.grapentin@lonza.com

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/19420862.2023.2169440>

© 2023 Lonza AG/Ltd.. Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

For high concentration mAb solutions, extensive work has been done in the past two decades to understand the factors resulting in high viscosity.^{17–24} Based on this body of knowledge, some themes emerge: 1) Reversible self-association is led by Fab-Fab or Fab-Fc interactions and not by Fc-Fc interactions; 2) The differences in solution viscosity are mainly driven by changes in the complementarity-determining region (CDR) of different mAbs; 3) Both hydrophobic and electrostatic interactions contribute to self-association in the form of surface patches; and 4) The size and anisotropy of these patches influences the extent of self-association.

In early development, multiple candidates are often available only in small quantities and need to be tested in pre-formulation studies for their stability and solubility. A substantial amount of work has been done recently to use experimental data from low concentration experiments to predict the viscosity behavior of high concentration solutions (Table 1). Prior to Roberts and colleagues report, experimental data of colloidal interactions,²⁸ the diffusion interaction parameter (kD) and the second virial coefficient (A2), were found to at least qualitatively predict mAbs with potentially high viscosity (i.e., problematic mAbs), but in this report, the predictions fail in many cases. Most reports also include only a low number of samples and describe the relationship between the experimental data and the viscosity linearly. The complexity of the origin of the solution viscosity at high mAb concentrations draws into question the validity of using low concentration experimental data as a predictive tool and suggests that a non-linear modeling approach may be more appropriate.

Experimental approaches used to forecast viscosity behavior of antibodies have recently been complemented by *in silico* methods, which aim to identify decisive molecular

descriptors like solvent exposure, local charge, hydrophobic effects, and surface patches. Publications that use computational approaches are listed in Table 2. One of the most promising models to predict viscosity uses a spatial charge map (SCM) to develop a score that can predict mAbs with high viscosity.³¹ The principle of SCM was used further in an approach testing machine learning to predict mAb solution viscosity.^{38,39}

The technique of machine learning aims at identifying certain patterns or descriptors that may be connected to a specific characteristic of the protein or its behavior in solution. Usually, large data sets are separated in each one training and validation subset. The training subset is then screened and patterns identified, a process known as *feature extraction*. These learnings are afterward applied in the confirmation of the model using the validation subset. Artificial neural networks (ANNs) are a subclass of machine learning, also known as deep learning, and use unstructured data sets. Similar to normal machine learning, the data set is divided into a training and a validation subset, but ANNs do not aim at identifying patterns. Instead, they use hidden layers, similar to the human brain, to establish a model.⁴⁰ An ANN with input of the amino acid composition of antibodies as only *in silico* data was used to predict kD, apparent melting temperature and onset temperature of aggregation (T_{agg}).³⁵ The models showed strong correlations, and a similar approach may be used for viscosity.

Our work describes the use of an ANN to model the viscosity of solutions of mAbs. We measured the viscosity of 27 mAbs of IgG1 or IgG2 subtype in a concentration range from 30 to 180 mg/mL in histidine-HCl buffer at pH 6. Histidine-HCl, a common buffer for mAbs, is used in > 80% of formulations of highly concentrated approved mAb DPs.¹⁰ The isoelectric point (pI) of the mAbs used in our study is in

Table 1. Selection of studies describing factors derived from experimental data correlation with solution viscosity.

First author, year	Test material and conditions for viscosity	Experimental parameters used (analytical technique)	Correlations to viscosity (model, conclusions)
Yadav, 2011 ²⁵	4 IgG1 at 125 mg/mL in diverse buffers at pH 4–10 plus NaCl	kD (DLS), Zeta potential, structure by circular dichroism	kD and Zeta potential utile as a qualitative high throughput screening tool for high concentration viscosity behavior
Saito, 2011 ²⁶	3 IgG1 at 150 mg/mL in diverse buffers at pH 5–8 plus NaCl	kD (DLS), B2 (AUC-SE)	kD and B2 show a qualitative relationship with viscosity behavior and aggregation
Connolly, 2012 ²⁷	29 IgG1 and IgG4 at 175 mg/mL, four solvent conditions at pH 5–6	kD (DLS), Zeta potential, Effective antibody charge (zone electrophoresis)	kD and B2 show a linear correlation ($R \geq 0.8$) with viscosity. Effective charge shows no linear correlation ($R \leq 0.6$) with viscosity.
Neergaard, 2013 ⁴⁰	3 IgG1 and IgG4 at 200 mg/mL in diverse buffers at pH 4–9 plus NaCl	kD (DLS), Zeta potential for effective pI; Calculated values: Net charge and theoretical pI	kD (if negative) and net charge (if low at pI) indicate potential high viscosity. Zeta potential predicts pH of maximum viscosity
Binabaji, 2015 ¹⁹	1 IgG1 at max 269 mg/mL, pH 5–7 in diverse buffers; plus excipients sucrose, proline, arginine-HCl, NaCl	Net protein charge and zeta potential by (electrophoretic light scattering); B2, B3 (osmotic pressure data)	B3 shows a strong linear correlation ($R^2=0.97$) with the viscosity parameter from the Mooney equation
Woldeyes, 2019 ²⁸	4 mAbs at 1–150 mg/mL, pH 5–8 in diverse buffers, plus NaCl	A2 (SLS), $S_{q=0}$ (SLS), kD (DLS), $H_{q=0}$ (DLS/SLS)	kD and A2, determined at low mAb concentration, are often not qualitative predictive of viscosity at high concentration. PPI are complex especially at high concentration and influence viscosity in nonlinear fashion.
Pathak, 2021 ²⁹	3 mAbs at max 90 mg/mL, pH 5.8–6.8 in diverse buffers; plus excipients arginine-HCl, NaCl, sorbitol, sucrose, polysorbate 80	B2 (SLS), hydrated protein molecular volume (simulation)	The Huggins coefficient (kH), derived from viscosity, can serve as a measure of PPIs in relation to the increase of solution viscosity, it can be used to identify solution conditions minimizing the increment of viscosity increase with protein concentration.

The table contains a selection of studies in chronological order which describe for monoclonal antibodies the correlation of factors derived from experimental data with their solution viscosity.

Abbreviations in table: kD = Diffusion interaction parameter; DLS = Dynamic light scattering; B2/A2 = Second virial coefficient; AUC-SE = Analytical ultracentrifugation – sedimentation equilibrium; pI = isoelectric point; B3 = Third virial coefficient; $S_{q=0}$ = Zero-q limit static structure factor; $H_{q=0}$ = hydrodynamic factor; SLS = Static light scattering; PPIs = Protein-protein-interactions

Table 2. Selection of studies describing computational models for prediction of mAb solution viscosity.

First author, year	Materials	<i>In silico</i> parameters used (analytical technique)	Correlations to viscosity
Sharma, 2014 ²⁰	14 IgG1 in 200 mM arginine buffer pH 5.5	Descriptors calculated solely from sequence information. These were Fv net charge, Fv charge symmetry parameter and Fv hydrophobicity	Principal components regression was used to obtain a model predicting mAb viscosity at 180 mg/mL from the 3 sequence derived descriptors (Pearsons R = 0.8)
Li, 2014 ²¹	3 IgG1 κ , 4 IgG1 λ , 2 IgG2 κ and 2 IgG4 κ in a histidine-HCl pH 5.8 formulation including excipients	Molecular descriptors calculated from the Fv region. These are sequence-based aggregation propensity (Waltz) and Fv pI	Linear Regression model predicting relative viscosity of mAbs at 150 mg/mL using the molecular descriptors normalized by number of residues in the Fv region ($R^2 = 0.93$)
Agrawal, 2016 ³¹	3 sets of mAbs used (1) 5 IgG1 κ , 1 IgG1 λ at pH 5.8 (2) 3 IgG1 κ , 4 IgG1 λ at pH 5.8 (3) 6 IgG1 κ at pH 6.0	Homology modeling of Fv to obtain negative surface patches. Creation of a spatial charge map	SCM scores are compared to viscosity data at 150 mg/mL, a threshold score is set at which mAbs can be considered problematic
Tomar, 2017 ³²	5 IgG1 κ , 4 IgG1 λ , 4 IgG2 κ and 3 IgG4 κ in a histidine-HCl pH 5.8 formulation including excipients	Full length homology modeling of mAbs, then computational obtaining of the Vh-charge, VI-charge, hinge-charge and hydrophobic solvent accessible surface area	Equation obtained by stepwise linear regression that combines net charge on Vh domain, net charge on the hinge region, net charge on the VI domain and the solvent accessible hydrophobic surface area, to predict the slope parameter B ($R^2 = 0.885$).
Apgar, 2020 ³³	40 variants of an IgG1 mAb in histidine-HCl pH 5.8 with sucrose	Used the scoring models developed by Agrawal, ³¹ Sharma, ³⁰ Tomar ³² and a simple correlation to Fv charge	Correlation of the concentration at which the mAbs show a viscosity of 20 mPa*s with Fv-charge ($R = 0.709$), SCM-score ($R = 0.695$), Sharma-score ($R = -0.716$) and Tomar-score ($R = -0.702$)
Lai, 2021 ³⁴	27 FDA approved mAbs (21 IgG1, 4 IgG2, and 2 IgG4) in histidine-HCl buffer pH 6.0	4 previously developed models were used, from Sharma, ³⁵ Li, ²¹ SCM by Agrawal ³¹ and the Therapeutic Antibody Profiler developability tool. ³⁶ A decision tree model using machine learning was also created.	The SCM model predicted the high viscosity mAbs correct but had 8 false positives. The linear viscosity prediction models by Sharma and Li had r values of 0.64 and 0.40 respectively. The developability tool only flagged one high viscosity mAb correctly while giving false positive for 5 low viscosity mAbs. The decision tree model predicted 26/27 mAbs correctly.
Thorsteinson, 2021 ³⁷	2 published datasets were used. The first contained 38 IgG1 mAbs, ³³ the second contained 21 FDA approved IgG1 mAbs	Fv charge (ens_charge_Fv) based on the forcefield average over 100 confirmations of Fv homology model	Linear correlation of ens_charge_Fv to the mAb concentration where viscosity crosses the 20 mPa*s threshold ($R^2 = 0.52$). Then classification decision with 84% accuracy
Lai, 2022 ³⁸	18 IgG1 and 2 IgG4 in histidine-HCl buffer pH 6.0	35 different molecular properties were used to create predictive models	Multiple classification models (linear regression, decision tree, k-nearest neighbor or support vector machine models) using one-feature or two feature combinations predicting high or low viscosity (>30 mPa*s at 150 mg/mL) with the best models showing an accuracy of $R^2 = 0.86$

The table contains a selection of studies in chronological order which describe for monoclonal antibodies the correlation of factors derived from computational data or *in silico* modeling with their solution viscosity. Abbreviations in table: Fv = Fragment variable; pI = Isoelectric point; SCM = Spatial charge map; Vh = variable heavy; VI = variable light

the 6.8–9.5 range, resulting in an overall net positive charge of all mAbs at the formulation pH 6. A large set of experimental-derived and computational data is fed into the ANN with the goal to model viscosity. The data set includes input parameters from colloidal stability assessment, kD and A2, as well as apparent surface hydrophobicity, measured by hydrophobic interaction chromatography (HIC). As computational data, the mAbs' pI and Fv-charge, calculated from the primary sequence, are used. Parameters derived from *in silico* modeling, such as hydrophobic and charged patch sizes, are also included.

We show that such ANNs can be used to not only make categorical predictions of problematic or unproblematic mAbs, meaning to predict whether viscosity is above a certain threshold at a specific protein concentration, but can also serve to calculate viscosity curves.

Results

To evaluate the impact of various input parameters on the predictive power of ANNs, different models were created containing the data of mAbs 1–25 (Table S1, Figure 1), in

which either only experimental inputs (retention time in HIC; kD, A2), or only computational (pI and Fv-charge) and *in silico*-derived inputs (patch sizes and numbers), or all available inputs were fed into the modeling. Each model was created by splitting the input data into a training and a validation set, which contained the information from randomized mAbs for each new ANN model. The training sets contained data, input variables, and viscosity descriptors, of 18–20 mAbs. The validation sets contained as data only the input variables of the remaining 7–5 mAbs, with the goal to predict their viscosity descriptors. These viscosity descriptors are derived from the linearization of the viscosity-concentration-curves of each mAb, the intercept A and the slope B.

The quality of the created models was evaluated by plotting the predicted values of either viscosity descriptor, A or B, against the respective descriptor obtained from the experimental data and assessing the linearity of these values. The R^2 values of the created models show the interdependency of validation and training set (Table 3), as in several cases the quality of one set is excellent ($R^2 > 0.99$) while the second set is less good ($R^2 < 0.95$). This interdependency of the models may

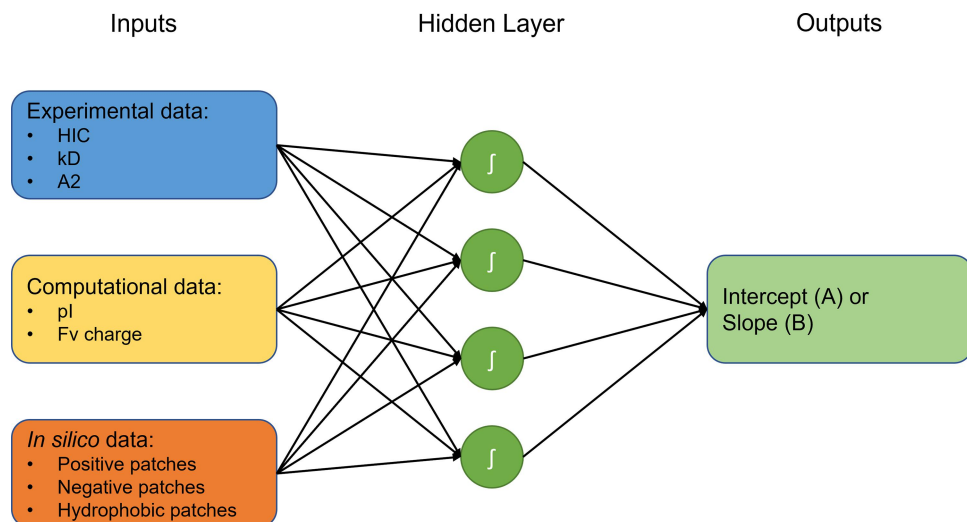


Figure 1. Artificial neural network (ANN) model scheme describing the input variables and setup of the ANN. Experimental inputs in blue, computational inputs derived from amino acid sequence in yellow, *in silico* simulation-derived input in orange. Inputs are combined in one layer with four hidden nodes with tanh activation functions and target determining viscosity descriptors, the intercept A or slope B.

originate in the amount of data, i.e., the number of mAbs, used in this study. The predictive power of the ANN depends on the amount and the range of viscosity behaviors featured by the mAbs in both the training and validation set. Ideally both sets should include well-behaved as well as problematic mAbs to cover the whole range of viscosity behaviors. This would ensure an effective training of the model with the first set and prevent the ANN from creating an overfitted model with the second set. With a total of 25 mAbs used for training and validation, the data set is substantial. Nevertheless, the number of problematic mAbs, which are likely the ones the model learns from most, is still limited. An uneven distribution between the two sets can thus influence their interdependency. This effect should decrease with an increasing size of available data, especially if more data of problematic mAbs are added.

Comparing the quality of the different models with regard to the input variables, the highest quality of model was achieved for the ANN in which all input variables (experimental and computational plus *in silico*) were used. This model has the lowest $R^2 > 0.92$ (intercept A of training set). The ANN using only computational and *in silico*-derived inputs achieves a minimally lower “worst” R^2 of > 0.90 (slope B of training set). Finally, the ANN containing only experimental data achieves

the lowest R^2 of > 0.75 , which is comparable to results published using linear correlations of kD to viscosity (see Table 1 for references) and better than a linear correlation of A2 and kD to the intercept and slope parameter done for this data set (Figure S1).

The slope parameter B describes the steepness of the viscosity curve, so the exponential increase of viscosity with the protein concentration, which is important to describe potentially problematic mAbs. Such problematic mAbs may show a moderate viscosity at low protein concentration, but experience a substantial increase of viscosity above values usually regarded as acceptable (e.g., 15–20 mPa*s) for drug products. The hyperbolic nature of the viscosity curve is common, yet for problematic mAbs the acceptable threshold value may be reached already at intermediate protein concentration, potentially disqualifying them from achieving the necessary dose in a volume of 1–2 mL. We decided to use the ANN with all input variables in the following, as it provided the best prediction for the slope B.

Of all available data, mAbs 26 and 27 were not used in ANN model creation, but were kept separately for additional verification tests.

Table 3. Quality of ANN models created with various input parameters.

Inputs			Training set			Validation set		
Experimental	<i>In silico</i>	Viscosity parameter	R^2	RMSE	SSE	R^2	RMSE	SSE
x	x	A (intercept)	0.9213	0.0267	0.0142	0.9999	0.0014	0.0000
x	x	B (slope)	0.9980	0.0003	0.0000	0.9380	0.0013	0.0000
-	x	A	0.9420	0.0265	0.0147	1.0000	0.0006	0.0000
-	x	B	0.9003	0.0017	0.0001	1.0000	0.0000	0.0000
x	-	A	0.7476	0.0574	0.0693	0.7532	0.0396	0.0078
x	-	B	0.8964	0.0016	0.0001	0.9824	0.0009	0.0000

Comparison of models created using different inputs predicting either the intercept (A) or slope (B) of the concentration-dependent viscosity curve for the mAbs. “x” indicates that a specific set of inputs was used for the model, while “-” indicates that the set of input parameters was not used. Experimental inputs refer to HIC retention time, kD and A2, while *in silico* includes the data from the surface patch analysis which is the information of size, score, and count of positive, negative, and hydrophobic surface patches. Computational data Fv-charge and pI was included regardless of the selected inputs. The quality parameters of the models shown were obtained from plotting the values of A or B that were calculated from the measured viscosity against the predicted values of the respective models.

R^2 = Coefficient of determination; RMSE = Root-mean-square error; SSE = Standard square error.

Table 4. Confusion matrix classification predictions.

Problematic at 120 mg/mL	Problematic at 150 mg/mL		Problematic at 180 mg/mL	Problematic at 150 mg/mL		Problematic at 180 mg/mL	Problematic at 150 mg/mL	
	No	Yes		No	Yes		No	Yes
No	18 (4)	0 (0)	No	16 (3)	0 (0)	No	8 (2)	0 (0)
Yes	0 (0)	2 (1)	Yes	0 (0)	4 (2)	Yes	0 (0)	12 (3)

Confusion matrices of the categorical models using experimental and *in-silico* data as inputs. Results of the validation sets are in brackets. Correct predictions are counted in the cells where the horizontal and vertical description is the same (No/No and Yes/Yes). False-positive and false-negative predictions are counted in the cells where the horizontal and vertical descriptions is different (No/Yes and Yes/No). The threshold to categorize a mAb as problematic was set to be >15 mPa*s for each selected concentration.

Categorical classification

Models for categorical classification were created by training the models on whether the mAbs show a viscosity above a threshold of 15 mPa*s at a certain concentration. This was done for concentrations of 120, 150, and 180 mg/mL, where of the 25 mAbs used in the model creation at 120 mg/mL 3 mAbs, at 150 mg/mL 6 mAbs and at 180 mg/mL 15 mAbs exhibited a viscosity of above 15 mPa*s. The confusion matrices of these models are shown in Table 4. It is apparent that both the training sets and validation sets contained problematic (viscosity ≥ 15 mPa*s) and unproblematic (viscosity < 15 mPa*s) mAbs. The models created have excellent predictive power regardless of the inputs used, with all of them exhibiting a misclassification rate of 0 (Table S3). To further evaluate the predictive power of these models, two mAbs, which were neither used in the training nor validation sets, were chosen for verification. mAb 26 shows unproblematic behavior at 120 and 150 mg/mL, but exceeds 15 mPa*s at 180 mg/mL, which was correctly predicted by our models (No/No/Yes). mAb 27 shows no problematic behavior at either of the concentrations, which was again correctly predicted by our models (No/No/No).

Viscosity curve prediction

To obtain more extensive information of mAbs' viscosity behavior, the full concentration-dependent viscosity curves were predicted, or more specifically, the intercept (A) and slope (B) of the linearized exponential function. Using the predicted values for A and B obtained from the predictive models with both experimental and *in silico* inputs, theoretical viscosity curves can be constructed and compared to the actual measured values, examples of such comparison are in Figure 2 A-D, where four mAbs with different viscosity behavior were chosen as a representative sample. The full predicted concentration-dependent viscosity curves for all mAbs used in this study are shown in Figure S2. Despite not exactly matching the actual values, the predicted values are very close and the predicted curve reflects the actual concentration-dependent viscosity behavior in a similar fashion. The percentage and absolute differences of predicted compared to measured viscosities of all mAbs are presented in Figure 3. Calculated average differences across all 27 mAbs between predicted and measured viscosity values are in Table 5. The average absolute difference, calculated to measured, is between 0.1 and 4.1

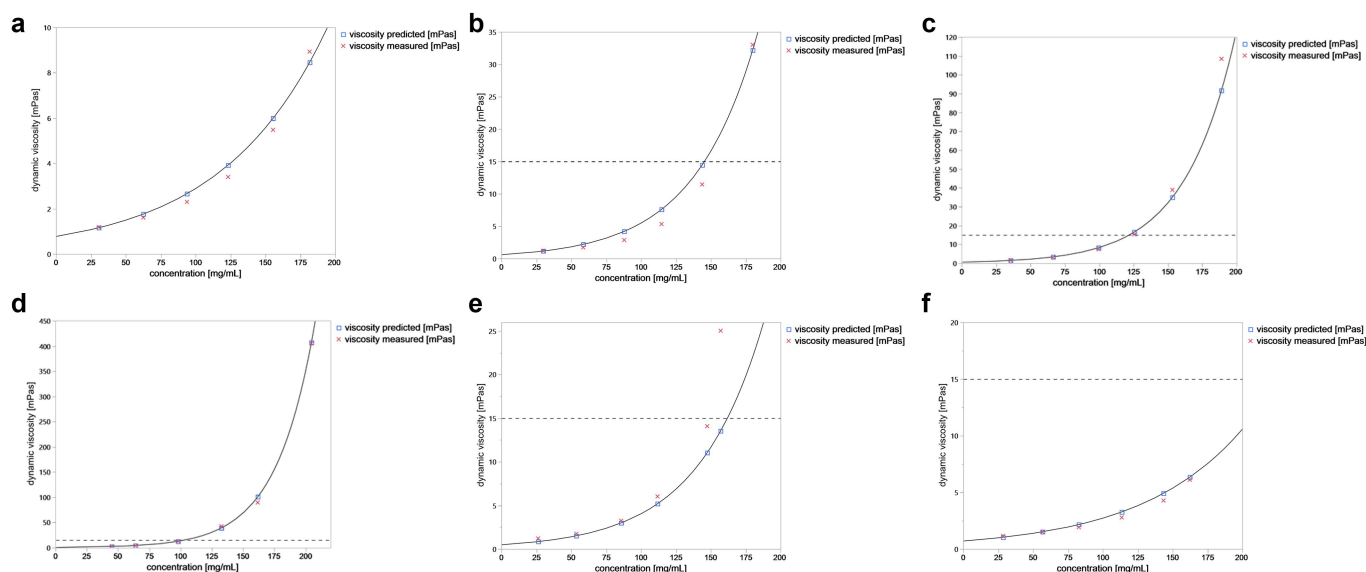


Figure 2. Examples for predicted viscosity curves. The predicted viscosity obtained from models using both experimental and *in silico* inputs versus the measured viscosity of selected mAbs; red crosses indicate the measured values, the black line was drawn using the predicted values for the slope (b) and intercept (a) inserted into the formula $y = A * e^{(B * x)}$, blue squares indicate the predicted viscosities at the same concentrations as the measured viscosities, the dashed line indicates the viscosity threshold for problematic mAbs at 15 mPa*s A) is mAb 13 (used in validation set) B) is mAb 2 (used in training set) C) is mAb 17 (used validation set) D) is mAb 24 (used in training set) E) is mAb 26 (used as verification mAb) F) is mAb 27 (used as verification mAb). Viscosity curves for all mAbs are in Figure S2.

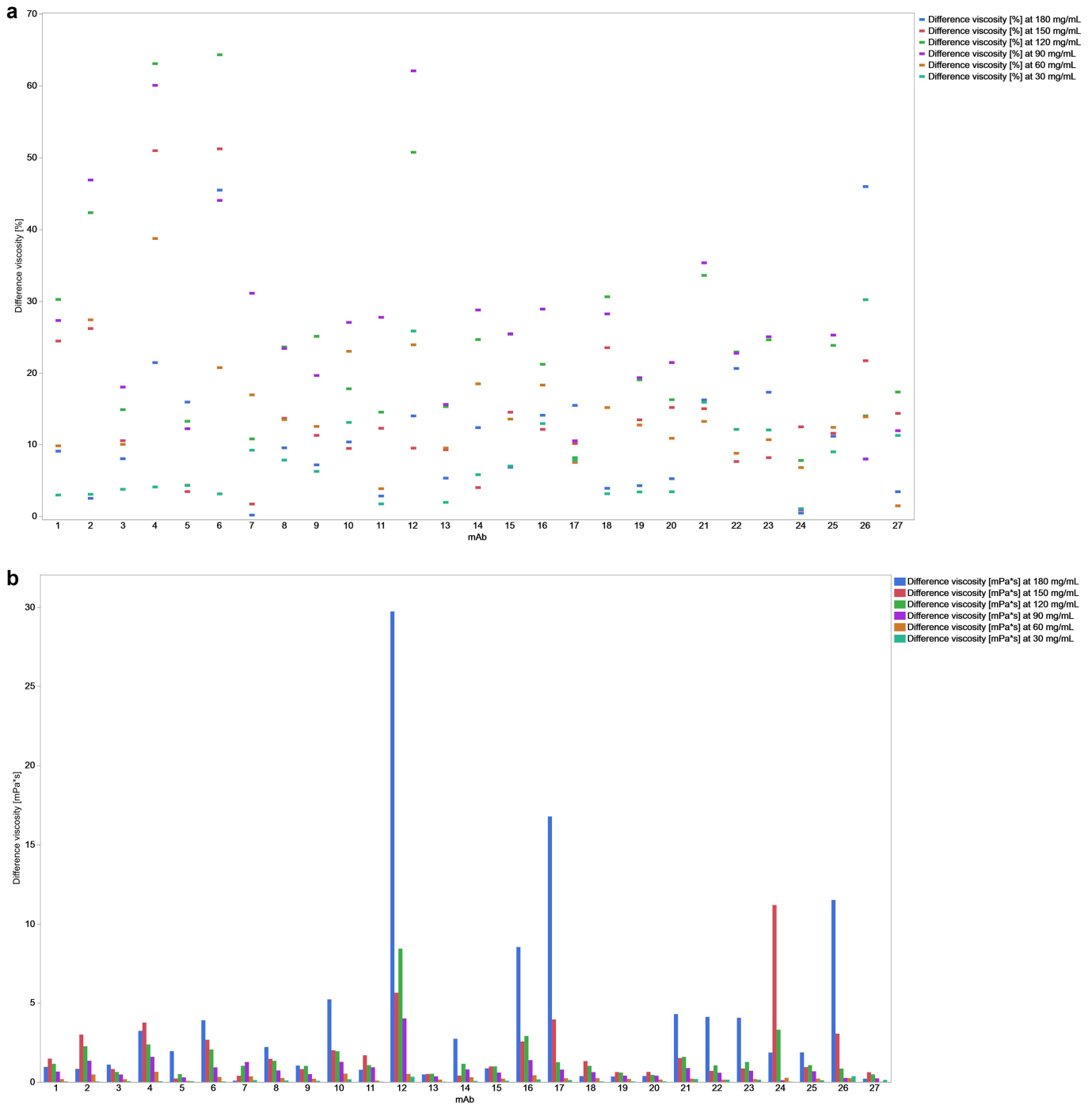


Figure 3. Difference of predicted viscosity values compared to measured viscosity values at the same protein concentration. The percentage (%) of difference is presented in A with squared points and the absolute difference (mPa*s) is presented in B with bars.

Table 5. Comparison of predicted and measured viscosity.

	Average difference between predicted and measured viscosity values at different protein concentration					
Concentration [mg/mL]	180	150	120	90	60	30
Absolute difference [mPa*s]	± 4.1	± 2.0	± 1.6	± 0.8	± 0.3	± 0.1
Relative difference [%]	12.2	15.5	25.0	26.2	14.0	8.2

Includes data from all mAbs in this study. Predicted values for viscosities were calculated using equation 2 and using the predicted values for intercept A and slope B from the model using all input variables. To get an accurate comparison the values of the concentrations for each mAb were taken from the actual measurements as shown in Table S2.

mPa*s, with a gradual increase of difference with increasing protein concentration. The relative difference % are between 8.2–26.2%, but follow a curved function, with the highest differences at concentrations of 90–120 mg/mL and decreasing difference with decreasing and increasing protein concentration.

For verification of the model, mAb 26 and 27 were again used; **Figure 2 E and F** show the predicted viscosity curves for both mAbs. While the course of the predicted curve for mAb 27 (**Figure 2 F**) seems very similar to the measured values, the predicted curve for mAb 26 (**Figure 2 E**) does not correctly reflect the steepness of the measured curve above 150 mg/mL.

Discussion

Prediction or forecasting of the viscosity behavior of mAbs at higher concentration is highly relevant for the pharmaceutical industry due to a growing interest in subcutaneously injected, and thus often highly concentrated, DPs and the potential implications of a late-phase discovery of a candidate's unsuitability for this route of administration. Because only limited amounts of protein are usually available for testing during early development, typically data from colloidal stability measurements at low concentration, usually < 10 mg/mL, are used, indicating attractive or repulsive PPI often in the form of A2 or kD, to correlate with viscosity at high concentration. Due to the complexity of molecular interactions, such linear, direct correlations of colloidal stability descriptors often have a low predictive power, and may thus eventually be limited to solution conditions of strong attractive PPI,²⁴ in line with historic publications (**Table 1**) and our own data (**Figure S1**).

Further descriptors of PPI can be added based on sequence analyses (e.g., pI, net charge) as well as *in silico* modeling (e.g., size, charge, location of charged or hydrophobic patches). Both types of data have the advantage of being relatively easily accessible and do not require material and laboratory work. With increasing scope of input data also more sophisticated data analysis tools may be required. A common approach described in literature for the prediction of viscosity³⁴ and aggregation,³⁸ solubility,⁴¹ or oxidation propensity of certain residues⁴² is machine learning, which aims at identifying distinctive features determining the degree of a trait of interest. A recent publication applying machine learning uses data from molecular modeling of 27 mAbs, specifically the Fv-charge and an SCM, as well as viscosity data obtained for protein concentrations up to 200 mg/mL in histidine-HCl pH 6.0.³⁴ The set includes mostly mAbs of isotype IgG 1 (21 molecules), but also IgG 2 (4 molecules) and IgG 4 (2 molecules), and uses a threshold of 30 mPa*s for definition of high or low viscosity. The machine learning approach using a decision tree model identifies certain features determining high viscosity, namely the net charge in the mAbs and the amino acid composition in the Fv. The viscosity categorization, below or above threshold, of most mAbs can be correctly predicted, but viscosity curves are not predicted.

In our approach, we use ANNs, which represent a subclass in machine learning, to predict the viscosity behavior of mAbs. The models can be used to predict a viscosity categorization, in our case above or below a threshold of 15 mPa*s, and to

predict viscosity curves. Whilst the categorical classification is flawless, the models for viscosity curve prediction show a high power and good viscosity curve forecast. The use of a multitude of input variables, derived from experimental data as well as from calculations or *in silico* modeling, appears advantageous. However, the data set can be expanded and the model can be applied for predictions also if only selected input data are available. The experimental data added in our model (kD, A2, HIC) can be easily obtained with low effort and protein consumption, the instruments required for their generation are widely distributed across academia and industry, and often these data are generated by default in early development phases of novel mAb candidates. Nevertheless, if only computational or *in silico*-derived inputs are available, the model can still be used for predictions. Potential applications are for example in candidate selection or early formulation development studies. In summary, ANNs appear to be a powerful tool to precipitate the complexity of molecular interactions that determine mAb viscosity at high concentration. In the future, the predictions may be further improved by the expansion of the experimental data set as well as further fine-grained definition, such as by including the location of charged patches or hydrophobic patches or their proximity. The work presented here offers an example of how ANNs can be used to predict complex protein parameters or behavior, but it does not explain the cause of it. A general limit of ANN is their lack of understanding or quantification of exact relationships. Nevertheless, as this brief report shows, their predictive power can be impressive and can be potentially improved.

Materials and methods

Monoclonal antibodies

All mAbs were obtained in-house from Lonza AG/Ltd. Double gene vectors containing the heavy and light chains were transfected into CHOK1SV GS-KO cells and cultured under selection conditions as stable pooled cultures. Clarified supernatant was obtained by centrifugation followed by filter sterilization using 0.22 μ m filters. Protein A chromatography was used for mAb purification. All proteins were concentrated to final concentration of 10 mg/mL and buffer exchanged into the formulation buffer (20 mM histidine-HCl, pH 6.0) by tangential flow filtration. mAbs are of different subtypes IgG1 or IgG2 (**Table S1**).

Formulation buffer

All lab experiments described (HIC, dynamic light scattering (DLS), static light scattering (SLS), viscosity) were performed in the formulation buffer, 20 mM histidine-HCl, pH 6.0.

Protein concentration

For concentration determination, an Agilent Cary 60 UV-Spectrophotometer with a variable path length extension SoloVPE was used. For each measurement 30 μ L of sample were loaded into a cuvette and measured at 280 nm using the appropriate specific extinction coefficient.

Hydrophobic interaction chromatography

The hydrophobic surface properties of all mAbs were determined by HIC. Proteins were analyzed at 10 mg/mL in formulation buffer, 5 μ L were injected on a ProPac Hic-10 column (ThermoScientific) and separated using a Waters HPLC system. The start condition of 95% mobile phase A (1 M ammonium sulfate in 20 mM sodium phosphate pH 7.0) was linearly reduced over 39 min to 95% mobile phase B (20 mM sodium phosphate pH 7.0). Flow rate was set to 1 mL/min at a column temperature of 24°C.

Dynamic and static light scattering

DLS and SLS measurements were performed on a DynaPro PlateReader III (with software Dynamics; Wyatt Technologies). Stock solutions of the antibodies were filtered through 0.22 μ m PVDF filters (Millex GV) and serial dilutions with seven concentrations from 10 to 2 mg/mL protein with the formulation buffer were prepared. Samples were transferred to 384-well plates (Aurora) in triplicate and the plates were centrifuged at 750*g for 2 min to remove air bubbles. The temperature for the measurement was set at 25°C. Laser power was set to 20% and attenuation level to 0%, 20 acquisitions of 5 s length were made for each well. Assessment of the diffusion interaction parameter kD (mL/g) was performed via dynamic light scattering (DLS). The mutual diffusion coefficient Dm (m^2/s) was plotted against the protein concentration (g/mL) and kD was obtained from the slope of a linear fit. The second virial coefficient $A2$ ($mol*mL/g$) was obtained from SLS measurements. Calibration of the plates was performed using Dextran (Sigma) with a predetermined molecular weight of 36.9 ± 0.1 kDa. Solvent offsets were measured in triplicates for the formulation buffer. The reciprocal molecular weight (mol/g) was plotted against the protein concentration (g/mL) and $A2$ was obtained from the slope of a linear fit.

Calculation of pI and Fv charge

The pI of the full mAbs was calculated from the amino acid sequence. To calculate the Fv charge, the variable heavy- and variable light chains of each antibody were analyzed with the prot pi protein tool (<https://www.protpi.ch/Calculator/ProteinTool>). The two chains were each defined as a subunit of the entire protein. The set modifier for post-translational modifications was global disulfide bridges for the cysteine residues in the Fv. Charge was calculated at pH 6.0.

In silico modeling of Fv

Modeling of mAbs was performed using the software BioLuminate (version 3.80; Schrödinger, LLC, New York, NY). Homology modeling of the Fv region was done by use of the antibody prediction tool. Framework templates for isotypes IgG 1 or IgG 2 were selected based on the highest composite score from the PDB database. The best CDR loop cluster was selected automatically. For modeling the standard presets of the software were kept, except the pH was set to 6.0 to represent the experimental settings. The surface of the

modeled mAb Fv-regions were analyzed with the protein surface analyzer tool in the BioLuminate software, to obtain size (in Å^2), count, score (sum of all contributing patch scores associated with this patch), and type (pos = positively charged; neg = negatively charged; hyd = hydrophobic) of the patches. The descriptors derived from modeling are thus “size pos/neg/hyd Fv total”, “count pos/neg/hyd Fv total”, and “score pos/neg/hyd Fv total”.

Rheometry and viscosity descriptors

mAbs were concentrated to approximately 180 mg/mL using spin filters with a 30 kDa molecular weight cut off and then diluted to six concentrations from 180 to 30 mg/mL. Concentration-dependent viscosity data were generated using a VROC viscometer (Rheosense) equipped with the B05 chip. The measurement protocol used in this study measures each sample 10 times at a temperature of 25°C whilst applying an automatic shear rate. This shear rate is determined by the instrument's software to produce a pressure inside the chip that results in the most precise measurement values for each sample. The shear rates applied were all significantly below the values where shear thinning in the samples would be observable. To obtain descriptive information of the concentration-dependent viscosity behavior of the mAb samples, the experimentally measured viscosity data were processed according to previous work by Li²¹ and Tomar.³² Equation 1 is used to calculate the relative viscosity of the mAb samples. The viscosity of the formulation buffer was measured to be 0.92 mPa*s at 25°C.

$$\eta_{rel} = \frac{\eta}{\eta_0}$$

Equation 1: Relative viscosity; η_{rel} =relative viscosity; η_0 =buffer viscosity; η =sample viscosity

Equation 2 was used to describe the exponential viscosity behavior of mAb solutions. This equation can be linearized to obtain the intercept A and the slope B using the natural logarithm as in equation 3.

$$\eta_{rel} = A * e^{Bc}$$

Equation 2: Exponential viscosity behavior; η_{rel} =relative viscosity; A =intercept; B =slope; c =mAb concentration

$$\ln \eta_{rel} = \ln A + Bc$$

Equation 3: linearized viscosity behavior; η_{rel} =relative viscosity; A =intercept; B =slope; c =mAb concentration

Data analyses and artificial neural network modeling

The ANN model creation follows the approach illustrated in Figure 1. All input parameters (experimental data, calculated values from sequence, *in silico*-derived data, and viscosity descriptors) are listed in Table S1. Each model is trained on a categorical response aiming to identify mAbs that show a viscosity value above the threshold value of 15 mPa*s at a distinct protein concentration. The ANNs are generated using software JMP v.16.0.0 (SAS Institute Inc.). The activation

function used for all nodes is the *tan h*-function, which transforms values to be between -1 and 1 . For all models, one hidden layer is sufficient with the number of nodes being four. To prevent the network from overfitting the model and in turn losing predictive power, the data is split into a training and a validation set. The method used is K-fold where the original data is split into K subsets. Each of the K sets is used to validate the model fit on the rest of the data, fitting a total of K models. The value of K was set to 5 for each model. Each model was created with a random seed (0). The model reported by the JMP software is based on the best log likelihood. The quality of the ANNs was determined using the coefficient of determination (R^2), the standard square and the root-mean-square error for training and validation datasets.

Abbreviations

Abbreviation	Meaning
A2, B2, B22	Second virial coefficient
ANN	Artificial neural network
AUC-SE	Analytical ultracentrifugation – sedimentation equilibrium
CDR	Complementarity-determining region
CMC	Chemistry, manufacturing and control
DLS	Dynamic light scattering
DP	Drug product
Fab	Fragment antigen binding
Fc	Fragment crystallizable
FDA	U.S. Food and Drug Administration
Fv	Fragment variable
HIC	Hydrophobic interaction chromatography
KD	Diffusion interaction parameter
mAb	Monoclonal antibody
mPa*s	Millipascal second
pI	Isoelectric point
PPI	Protein-protein-interaction
s.c.	Subcutaneous
SLS	Static light scattering

Acknowledgments

The authors thank Lonza Basel employees Marigone Lenjani, Jill Werner, Sonja Rutz, Thomas Zech, Emie-kim Ngotan, and Katja In-Albon for support in experiments and data generation. The authors thank Lonza Slough employee Jean Aucamp for the purification and generous provision of the monoclonal antibodies. The support of Schrödinger for technical support with the software BioLuminate and its provision free of charge for an extended period is greatly acknowledged.

Disclosure statement

The authors report no conflict of interest.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

References

- Kaplon H, Chenoweth A, Crescioli S, Reichert JM. Antibodies to watch in 2022. *MAbs*. 2021;14 PMID: 35030985. doi:10.1080/19420862.2021.2014296.
- Jiskoot W, Hawe A, Menzen T, Volkin DB, Crommelin DJA. Ongoing Challenges to Develop High Concentration Monoclonal Antibody-based Formulations for Subcutaneous Administration: Quo Vadis? *J Pharm Sci*. 2022;111:861–67. PMID: 34813800. doi:10.1016/j.xphs.2021.11.008.
- Mathaes R, Koulov A, Joerg S, Mahler HC. Subcutaneous Injection Volume of Biopharmaceuticals—Pushing the Boundaries. *J Pharm Sci*. 2016;105:2255–59. PMID: 27378678. doi:10.1016/j.xphs.2016.05.029.
- Shire SJ, Shahrokh Z, Liu J. Challenges in the development of high protein concentration formulations. *J Pharm Sci*. 2004;93:1390–402. PMID: 15124199. doi:10.1002/jps.20079.
- Yadav S, Liu J, Shire SJ, Kalonia DS. Specific interactions in high concentration antibody solutions resulting in high viscosity. *J Pharm Sci*. 2010;99:1152–68. PMID: 19705420. doi:10.1002/jps.21898.
- Kanai S, Liu J, Patapoff TW, Shire SJ. Reversible self-association of a concentrated monoclonal antibody solution mediated by Fab–Fab Interaction that impacts solution viscosity. *J Pharm Sci*. 2008;97:4219–27. PMID: 18240303. doi:10.1002/jps.21322.
- Wang S, Zhang N, Hu T, Dai W, Feng X, Zhang X, Qian F. Viscosity-Lowering Effect of Amino Acids and Salts on Highly Concentrated Solutions of Two IgG1 Monoclonal Antibodies. *Mol Pharmaceut*. 2015;12:4478–87. PMID: 26528726. doi:10.1021/acs.molpharmaceut.5b00643.
- Scherer TM, Liu J, Shire SJ, Minton AP. Intermolecular Interactions of IgG1 Monoclonal Antibodies at High Concentrations Characterized by Light Scattering. *J Phys Chem B*. 2010;114:12948–57. PMID: 20849134. doi:10.1021/jp1028646.
- Tilegenova C, Izadi S, Yin J, Huang CS, Wu J, Ellerman D, Hymowitz SG, Walters B, Salisbury C, Carter PJ. Dissecting the molecular basis of high viscosity of monospecific and bispecific IgG antibodies. *MAbs*. 2020;12:1692764. PMID: 31779513. doi:10.1080/19420862.2019.1692764.
- Wang SS, Yan Y, Ho K. US FDA-approved therapeutic antibodies with high-concentration formulation: summaries and perspectives. *Antib Ther*. 2021;4:262–73. PMID: 34909579. doi:10.1093/abt/tbab027.
- Tsumura K, Hsu W, Mimura M, Horiuchi A, Shiraki K. Lowering the viscosity of a high-concentration antibody solution by protein–polyelectrolyte complex. *J Biosci Bioeng*. 2022;133:17–24. PMID: 34629298. doi:10.1016/j.jbiosc.2021.09.011.
- Zeng Y, Tran T, Wuthrich P, Naik S, Davagnino J, Greene DG, Mahoney RP, Soane DS. Caffeine as a Viscosity Reducer for Highly Concentrated Monoclonal Antibody Solutions. *J Pharm Sci*. 2021;110:3594–604. PMID: 34181992. doi:10.1016/j.xphs.2021.06.030.
- Guo Z, Chen A, Nassar RA, Helk B, Mueller C, Tang Y, Gupta K, Klibanov AM. Structure-Activity Relationship for Hydrophobic Salts as Viscosity-Lowering Excipients for Concentrated Solutions of Monoclonal Antibodies. *Pharm Res*. 2012;29:3102–09. PMID: 22692671. doi:10.1007/s11095-012-0802-9.
- Srivastava A, O'Dell C, Bolessa E, McLinden S, Fortin L, Deorkar N. Viscosity Reduction and Stability Enhancement of Monoclonal Antibody Formulations Using Derivatives of Amino Acids. *J Pharm Sci*. 2022;111:2848–56. PMID: 35605688. doi:10.1016/j.xphs.2022.05.011.
- Yang Y, Velayudhan A, Thornhill NF, Farid SS. Multi-criteria manufacturability indices for ranking high-concentration monoclonal antibody formulations. *Biotechnol Bioeng*. 2017;114:2043–56. PMID: 28464235. doi:10.1002/bit.26329.
- Heise T, Nosek L, Dellweg S, Zijlstra E, Praestmark KA, Kildegaard J, Nielsen G, Sparre T. Impact of injection speed and volume on perceived pain during subcutaneous injections into the abdomen and thigh: a single-centre, randomized controlled trial. *Diabetes Obes Metab*. 2014;16:971–76. PMID: 24720741. doi:10.1111/dom.12304.
- Tomar DS, Kumar S, Singh SK, Goswami S, Li L. Molecular basis of high viscosity in concentrated antibody solutions: strategies for

- high concentration drug product development. *MAbs*. 2016;8:216–28. PMID: 26736022. doi:10.1080/19420862.2015.1128606;
18. Yadav S, Shire SJ, Kalonia DS. Factors affecting the viscosity in high concentration solutions of different monoclonal antibodies. *J Pharm Sci*. 2010;99:4812–29. PMID: 20821382. doi:10.1002/jps.22190;
 19. Binabaji E, Ma J, Zydney AL. Intermolecular interactions and the viscosity of highly concentrated monoclonal antibody solutions. *Pharm Res*. 2015;32:3102–09. PMID: 25832501. doi:10.1007/s11095-015-1690-6;
 20. Neergaard MS, Kalonia DS, Parshad H, Nielsen AD, Møller EH, van de Weert M. Viscosity of high concentration protein formulations of monoclonal antibodies of the IgG1 and IgG4 subclass – prediction of viscosity through protein–protein interaction measurements. *Eur J Pharma Sci*. 2013;49:400–10. PMID: 23624326. doi:10.1016/j.ejps.2013.04.019;
 21. Li L, Kumar S, Buck PM, Burns C, Lavoie J, Singh SK, Warne NW, Nichols P, Luksha N, Boardman D. Concentration Dependent Viscosity of Monoclonal Antibody Solutions: Explaining Experimental Behavior in Terms of Molecular Properties. *Pharm Res*. 2014;31:3161–78. PMID: 24906598. doi:10.1007/s11095-014-1409-0;
 22. Nichols P, Li L, Kumar S, Buck PM, Singh SK, Goswami S, Balthazor B, Conley TR, Sek D, Allen MJ. Rational design of viscosity reducing mutants of a monoclonal antibody: hydrophobic *versus* electrostatic inter-molecular interactions. *MAbs*. 2015;7:212–30. PMID: 25559441. doi:10.4161/19420862.2014.985504;
 23. Skar-Gislinge N, Camerin F, Stradner A, Zaccarelli E, Schurtenberger P. Cluster formation and the link to viscosity in antibody solutions. *ArXiv*. 2022. doi:10.48550/arXiv.2209.05182;
 24. Shimomura T, Sekiguchi M, Honda R, Yamazaki M, Yokoyama M, Uchiyama S. Estimation of the Viscosity of an Antibody Solution from the Diffusion Interaction Parameter. *Biol Pharm Bull*. 2022;45:b22–00263. PMID: 36047198. doi:10.1248/bpb.b22-00263;
 25. Yadav S, Shire SJ, Kalonia DS. Viscosity behavior of high-concentration monoclonal antibody solutions: correlation with interaction parameter and electroviscous effects. *J Pharm Sci*. 2012;101:998–1011. PMID: 22113861. doi:10.1002/jps.22831;
 26. Saito S, Hasegawa J, Kobayashi N, Kishi N, Uchiyama S, Fukui K. Behavior of Monoclonal Antibodies: Relation Between the Second Virial Coefficient (B_2) at Low Concentrations and Aggregation Propensity and Viscosity at High Concentrations. *Pharm Res*. 2012;29:397–410. PMID: 21853361. doi:10.1007/s11095-011-0563-x;
 27. Connolly BD, Petry C, Yadav S, Demeule B, Ciaccio N, Moore JMR, Shire SJ, Gokarn YR. Weak interactions govern the viscosity of concentrated antibody solutions: high-throughput analysis using the diffusion interaction parameter. *Biophys J*. 2012;103:69–78. PMID: 22828333. doi:10.1016/j.bpj.2012.04.047;
 28. Woldeyes MA, Qi W, Razinkov VI, Furst EM, Roberts CJ. How Well Do Low- and High-Concentration Protein Interactions Predict Solution Viscosities of Monoclonal Antibodies? *J Pharm Sci*. 2019;108:142–54. PMID: 30017887. doi:10.1016/j.xphs.2018.07.007;
 29. Pathak J, Nugent S, Bender M, Roberts C, Curtis R, Douglas J. Comparison of Huggins Coefficients and Osmotic Second Virial Coefficients of Buffered Solutions of Monoclonal aAntibodies. *Polymers (Basel)*. 2021;13:601. PMID: 33671342. doi:10.3390/polym13040601;
 30. Sharma VK, Patapoff TW, Kabakoff B, Pai S, Hilario E, Zhang B, Li C, Borisov O, Kelley RF, Chorny I, et al. In silico selection of therapeutic antibodies for development: viscosity, clearance, and chemical stability. *Proc Nat Acad Sci*. 2014;111:18601–06. PMID: 25512516. doi:10.1073/pnas.1421779112;
 31. Agrawal NJ, Helk B, Kumar S, Mody N, Sathish HA, Samra HS, Buck PM, Li L, Trout BL. Computational tool for the early screening of monoclonal antibodies for their viscosities. *MAbs*. 2016;8:43–48. PMID: 26399600. doi:10.1080/19420862.2015.1099773;
 32. Tomar DS, Li L, Broulidakis MP, Luksha NG, Burns CT, Singh SK, Kumar S. In-silico prediction of concentration-dependent viscosity curves for monoclonal antibody solutions. *MAbs*. 2017;9:476–89. PMID: 28125318. doi:10.1080/19420862.2017.1285479;
 33. Apgar JR, Tam ASP, Sorm R, Moesta S, King AC, Yang H, Kelleher K, Murphy D, D'Antona AM, Yan G, et al. Modeling and mitigation of high-concentration antibody viscosity through structure-based computer-aided protein design. *PLoS One*. 2020;15:e0232713. PMID: 32379792. doi:10.1371/journal.pone.0232713;
 34. Lai P-K, Fernando A, Cloutier TK, Gokarn Y, Zhang J, Schwenger W, Chari R, Calero-Rubio C, Trout BL. Machine Learning Applied to Determine the Molecular Descriptors Responsible for the Viscosity Behavior of Concentrated Therapeutic Antibodies. *Mol Pharm*. 2021;18:1167–75. PMID: 33450157. doi:10.1021/acs.molpharmaceut.0c01073;
 35. Gentiluomo L, Roessner D, Augustijn D, Svilenov H, Kulakova A, Mahapatra S, Winter G, Streicher W, Åsmund R, Peters GHJ, et al. Application of interpretable artificial neural networks to early monoclonal antibodies development. *Eur J Pharma Biopharm*. 2019;141:81–89. PMID: 31112768. doi:10.1016/j.ejpb.2019.05.017;
 36. Raybould MIJ, Marks C, Krawczyk K, Taddese B, Nowak J, Lewis AP, Bujotzek A, Shi J, Deane CM. Five computational developability guidelines for therapeutic antibody profiling. *Proc Nat Acad Sci*. 2019;116:4025–30. PMID: 30765520. doi:10.1073/pnas.1810576116;
 37. Thorsteinson N, Gunn JR, Kelly K, Long W, Labute P. Structure-based charge calculations for predicting isoelectric point, viscosity, clearance, and profiling antibody therapeutics. *MAbs*. 2021;13 PMID: 34632944. doi:10.1080/19420862.2021.1981805;
 38. Lai P-K, Gallegos A, Mody N, Sathish HA, Trout BL. Machine learning prediction of antibody aggregation and viscosity for high concentration formulation development of protein therapeutics. *MAbs*. 2022;14 PMID: 35075980. doi:10.1080/19420862.2022.2026208;
 39. Lai P-K. DeepSCM: An efficient convolutional neural network surrogate model for the screening of therapeutic antibody viscosity. *Comput Struct Biotechnol J*. 2022;20:2143–52. PMID: 35832619. doi:10.1016/j.csbj.2022.04.035;
 40. Helm JM, Swiergosz AM, Haerberle HS, Karnuta JM, Schaffer JL, Krebs VE, Spitzer AI, Ramkumar PN. Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. *Curr Rev Musculoskelet Med*. 2020;13:69–76. PMID: 31983042. doi:10.1007/s12178-020-09600-8;
 41. Han X, Shih J, Lin Y, Chai Q, Cramer SM. Development of QSAR models for *in silico* screening of antibody solubility. *MAbs*. 2022;14 PMID: 35442164. doi:10.1080/19420862.2022.2062807;
 42. Delmar JA, Buehler E, Chetty AK, Das A, Quesada GM, Wang J, Chen X. Machine learning prediction of methionine and tryptophan photooxidation susceptibility. *Mol Ther Methods Clin Dev*. 2021;21:466–77. PMID: 33898635. doi:10.1016/j.omtm.2021.03.023;