CLAIMS

What is claimed is:

1. A method comprising:

determining experimental data associated with one or more monoclonal antibodies (mAbs);

determining computationally-derived data associated with the one or more mAbs, wherein the computationally-derived data comprises one or more computational parameters weighted based on accessible surfaces (ASAs) of one or more residues of the one or more mAbs;

determining, based on the experimental data and the computationally-derived data, a plurality of candidate predictive models;

determining an optimal predictive model from the plurality of candidate predictive models; and

outputting the optimal predictive model.

2. The method of claim 1, wherein the one or more mAbs comprise one or more of an IgG1 antibody or an IgG4 antibody.

3. The method of claim 1, wherein the experimental data comprises experimental viscosity data and wherein the experimental viscosity data comprises one or more of dynamic viscosity values or kinematic viscosity values and wherein determining the experimental data associated with the one or more mAbs comprises measuring, based on a solution of each of the one or more mAbs and a viscometer, at least one of a dynamic viscosity value or a kinematic viscosity value.

4. The method of claim 1, wherein the computationally-derived data comprises charge data associated with one or more regions associated with a sequence of the one or more mAbs, modified charge data associated with the one or more regions based on a solvent accessible surface of a residue in a homology model of the one or more mAbs, a hydrophobicity index (HI), a dipole moment, or an isoelectric point (pI).

5. The method of claim 1, wherein determining the computationally-derived data associated with the one or more mAbs comprises full-antibody homology modeling of a sequence of the one or more mAbs or antigen-binding fragment (Fab) region modeling of the Fab sequence of the one or more mAbs.

6. The method of claim 1, wherein determining the computationally-derived data associated with the one or more mAbs comprises:

determining, based on a homology model of the one or more mAbs, one or more charge values associated with one or more residues in one or more regions of the one or more mAbs;

determining, based on the homology model of the one or more mAbs, a solvent accessible surface (SAS) of the one or more residues in the one or more regions;

adjusting, based on a weighting factor calculated using the SAS of the one or more residues relative to a total SAS associated with the one or more mAbs, the one or more charge values associated with the one or more residues; and

determining, based on the homology model of the one or more mAbs and the adjusted one or more charge values associated with the one or more residues, a charge value associated with each region of the one or more regions.

7. The method of claim 1, wherein determining, based on the experimental data and the computationally-derived data, the plurality of candidate predictive models comprises:

identifying one or more experimental parameters of the experimental data as dependent variables;

identifying one or more computational parameters of the computationally-derived data as independent variables; and

determining, based on a stepwise regression algorithm, based on the dependent variables, and based on the intendent variables, the plurality of candidate predictive models.

8. The method of claim 1, wherein determining the optimal predictive model from the plurality of candidate predictive models comprises:

determining, for each candidate predictive model of the plurality of candidate predictive models, an Akaike Information Criterion (AIC) score; and

determining, as the optimal predictive model, the candidate predictive model of the plurality of candidate predictive models associated with the highest AIC score.

9. The method of claim 1, wherein determining the optimal predictive model from the plurality of candidate predictive models comprises:

determining, as the optimal predictive model, the candidate predictive model of the plurality of candidate predictive models associated with a lowest error in predicting a viscosity score of a mAb excluded from the experimental data and the computationally-derived data.

10. The method of claim 1, further comprising:

receiving computationally-derived data associated with a query mAb;

providing, to the optimal predictive model, the computationally-derived data; and

determining, based on the optimal predictive model, a viscosity score associated with the query mAb.

11. The method of claim 10, further comprising:

adjusting, based on the viscosity score, an appropriate formulation composition or protein engineering strategy to mitigate specific challenges with the drug candidate in development, for example, adjusting an amount of viscosity reducer of a solution associated with the query mAb.

12. The method of claim 1, wherein the experimental data comprises experimental aggregation data and wherein the experimental aggregation data comprises high-molecular-weight (HMW) species formation data for each mAb of the one or more mAbs, and wherein determining the experimental data associated with the one or more mAbs comprises:

measuring, based on a solution of each of the one or more mAbs and size-exclusion chromatography (SEC), an amount of HMW species formation over time.

13. The method of claim 1, wherein the computationally-derived data comprises charge data associated with one or more regions associated with a sequence of the one or more mAbs, modified charge data associated with the one or more regions based on a solvent accessible surface of a residue in a homology model of the one or more mAbs,

a hydrophobicity index (HI), a dipole moment, an isoelectric point (pI), an aggregation propensity (AP), or a descriptor of conformational stability.

14. The method of claim 13, wherein the descriptor of conformational stability comprises a backbone root mean square deviation (RMSD) of a conformational structure relative to an initial structure after rigid-body alignment.

15. The method of claim 1, wherein determining the computationally-derived data associated with the one or more mAbs comprises one or more Molecular Dynamics (MD) simulations associated with the one or more mAbs.

16. The method of claim 1, wherein determining the optimal predictive model from the plurality of candidate predictive models comprises:

determining, as the optimal predictive model, the candidate predictive model of the plurality of candidate predictive models associated with a lowest error in predicting an aggregation score of a mAb excluded from the experimental data and the computationally-derived data.

17. The method of claim 1, further comprising:

receiving computationally-derived data associated with a query mAb;

providing, to the optimal predictive model, the computationally-derived data; and

determining, based on the optimal predictive model, an aggregation score.

18. The method of claim 17, further comprising:

adjusting, based on the aggregation score, an appropriate formulation composition or protein engineering strategy to mitigate specific challenges with the drug candidate in development, for example, adjusting an amount of aggregation reducer of a solution associated with the query mAb.

19. A method comprising:

receiving computationally-derived data associated with a monoclonal antibody (mAb); and

providing, to a predictive model, the computationally-derived data; and

determining, based on the predictive model, an aggregation score associated with the mAb.

20. The method of claim 19, further comprising:

adjusting, based on the aggregation score, an appropriate formulation composition or protein engineering strategy to mitigate specific challenges with a drug candidate in development.