Received: 29 December 2023 | Revised: 17 March 2024 | Accepted: 27 March 2024

DOI: 10.1002/btpr.3466

BIOTECHNOLOGY PROGRESS

RESEARCH ARTICLE

Cell Culture and Tissue Engineering

A systemic approach to identifying sequence frameworks that decrease mAb production in a transient Chinese hamster ovary cell expression system

Alana C. Szkodny | Kelvin H. Lee

Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, Delaware, USA

Correspondence
Kelvin H. Lee, Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE 19716, USA.
Email: khl@udel.edu

Funding information
National Science Foundation, Grant/Award Number: 1624698; National Institute of General Medical Sciences, Grant/Award Number: P20GM103446; National Institute of Standards and Technology, Grant/Award Number: 70NANB17H002; State of Delaware; Delaware Biotechnology Institute

Abstract

Monoclonal antibodies (mAbs) are often engineered at the sequence level for improved clinical performance yet are rarely evaluated prior to candidate selection for their "developability" characteristics, namely expression, which can necessitate additional resource investments to improve the manufacturing processes for problematic mAbs. A strong relationship between primary sequence and expression has emerged, with slight differences in amino acid sequence resulting in titers differing by up to an order of magnitude. Previous work on these "difficult-to-express" (DTE) mAbs has shown that these phenotypes are driven by post-translational bottlenecks in antibody folding, assembly, and secretion processes. However, it has been difficult to translate these findings across cell lines and products. This work presents a systematic approach to study the impact of sequence variation on mAb expression at a larger scale and under more industrially relevant conditions. The analysis found 91 mutations that decreased transient expression of an $IgG_1\kappa$ in Chinese hamster ovary (CHO) cells and revealed that mutations at inaccessible residues, especially those leading to decreases in residue hydrophobicity, are not favorable for high expression. This workflow can be used to better understand sequence determinants of mAb expression to improve candidate selection procedures and reduce process development timelines.

KEYWORDS
CHO cells, developability, difficult-to-express, monoclonal antibodies, transient expression

1 | INTRODUCTION

The study of sequence–structure–function relationships in monoclonal antibodies (mAbs) has led to significant improvements in clinical performance which have contributed to the therapeutic success of these products in recent decades.[1–3] Through advancements in genetic and cellular engineering, many methods now exist to aid in the search for antibodies with optimal antigen affinity and immune system activation, but improvements to these therapeutically relevant characteristics may come at the expense of other properties that can affect mAb behavior in biopharmaceutical manufacturing and formulation processes.[4] Recently, the acceleration of drug development timelines due to market competitiveness has led to increased focus on understanding the structural basis of mAb "developability" characteristics such as thermal stability, aggregation, degradation, and expression. These characteristics can define the ease with which a mAb candidate can be produced, purified, and formulated through today's biomanufacturing platform processes. Often, "poorly-behaved" mAbs are not identified until the early stages of process development, which necessitates significant time and resource investments to troubleshoot and subsequently modify manufacturing processes to overcome production challenges. Currently, no consensus

rules or thresholds have been established to define what characteristics make a mAb amenable to development and manufacture, making early identification of challenging mAbs difficult.[5] Preliminary attempts to predict mAb developability have been inspired by previous work with small molecule drugs that identified a set of general guidelines (Lipinski's "rule of five") that help predict the therapeutic success of new candidates based on molecular structure.[6] Similar efforts in the mAb space have led to various in silico tools that attempt to predict biophysical characteristics or overall developability using the primary amino acid sequence or a structural model.[7] These studies have relied on large datasets of sometimes more than 100 mAbs where each antibody is subjected to a battery of in silico measures, such as calculated hydrophobicity, viscosity, and charge metrics, as well as in vitro assays for measuring aggregation, chemical modifications, colloidal stability, and thermal stability.[5,8–11] However, no single assay, tool, or parameter has been universally accepted for developability prediction. Additionally, mAb productivity remains a key, but largely unstudied, developability metric, and for the few developability studies that have included measures of expression, they have found either no correlation or weak correlation of titer with other metrics.[5,11,12]

Cellular engineering efforts to increase mAb expression have taken two complementary approaches to address both transcriptional and post-transcriptional bottlenecks in protein production. The maximal transcriptional output of mAb genes has been achieved by improved gene cassette designs, clone selection procedures, and for targeted integration cell lines, the identification of genomic "hot spots" that confer consistently high levels of transcription.[13–17] Other studies have addressed post-transcriptional bottlenecks by targeting various components of protein processing pathways through overexpression of folding chaperones, ER stress and apoptosis regulators, transcription factors, or non-coding RNAs.[18–25] However, no single engineering approach has yielded universal improvements, with productivity gains depending on the cell line, product, and experimental approach.[26] Furthermore, even subtle differences in mAb primary sequence can result in large changes in expression in optimized systems, with single amino acid differences between mAb candidates resulting in titers differing by up to an order of magnitude in both transient and stable expression systems.[27–30] A growing body of work studying these "difficult-to-express" (DTE) mAbs has revealed that most DTE phenotypes are driven by post-translational bottlenecks in signal peptide processing, protein folding, or protein assembly, and can cause the upregulation of pathways such as the unfolded protein response (UPR) and ER-associated degradation (ERAD).[20,27,31–36] However, many of these studies focus on a small number of mAbs (<12) with varying levels of primary sequence similarity and use a variety of cell culture systems and antibody formats, making it difficult to translate their findings across cell lines and products. In general, the bottlenecks for DTE mAbs are thought to be cell line- and product-specific, requiring tailored engineering solutions for the system's unique challenges.[20,37]

Large-scale efforts to investigate sequence–structure–function relationships in mAbs have relied on high-throughput phage display systems which can achieve high levels of sequence diversity with libraries reaching >10^{10} variants, making these display systems invaluable tools for antibody discovery and the identification of key residue interactions involved in antigen binding, aggregation, and thermal stability.[38–42] However, the bacterial expression systems used for phage display are limited to the expression of individual antibody domains or single-chain variable fragments and lack the protein folding pathways and checkpoints present in mammalian cells. Therefore, the expression information arising from these studies only accounts for whether the variant is expressed at sufficient levels to be detected in the selected library. These measures of "functional" expression may not be reflective of expression trends for full-length mAbs, as mAbs derived from phage display have been found to exhibit more developability "red flags" than mAbs derived from mammalian systems.[5] Display systems using mammalian cells can account for mammalian-specific protein production pathways but these systems are constrained by low cell densities (10^6–10^7 cells) that limit the achievable sequence diversity.[43,44]

To increase the throughput of mAb mutational studies in industrially relevant mammalian cell systems, this work presents a systematic approach for selecting and evaluating the impacts of single amino acid mutations on mAb expression, with particular focus on amino acid changes that are detrimental to expression. mAb structural information and biophysical characteristics were used to select 178 single amino acid mutations in the humanized $IgG_1\kappa$ mAb trastuzumab with the intention and expectation that the selected mutations would decrease mAb expression. Each variant was then transiently expressed in a Chinese hamster ovary (CHO) cell line. Dozens of novel mutations that decrease mAb expression were identified and it was found that combining expression information with the physical and chemical changes resulting from changes in sidechain chemistry revealed distinct combinations of primary sequence mutations and molecular contexts that affect mAb expression.

2 | MATERIALS AND METHODS

2.1 | Determination of biophysical characteristics

Kabat numbering is used throughout the text, with sequence numbering used where indicated. Computational alanine scanning calculations were performed with FoldX, using the trastuzumab crystal structure (PDB: 1n8z) as an input file.[45,46] The PDB file was repaired using the "RepairPDB" function prior to use. Alanine scanning on protein complexes was performed with the "Pssm" function. Antigen-contacting residues were identified by considering the interaction between the light chain (LC)/heavy chain (HC) complex and HER2, and antibody interface residues were identified through the interaction of the LC and HC. Alanine scanning for the individual LC and HC structures were determined using the "AlaScan" function. Free energy values for mutations to amino acids other than alanine were calculated with the "PositionScan" function. For all functions, the default parameters were used. Free energy changes calculated in Rosetta used the

"ddg_monomer" function and the "protocol 13" sampling methodology presented by Kellogg and colleagues.[47] The free energy of the wild-type structure was minimized prior to use. Differences in amino acid molecular weight, hydropathy, and percent solvent-accessible surface area (SASA) upon mutation were calculated by subtracting the wild-type value from the mutated value.[48,49] Both the molecular surface area and the SASA were determined using the "get_area" function in Pymol, using mutated structures generated by Rosetta. The conservation of each mutation was extracted from the abYsis database using amino acid distributions from *Homo sapiens* (retrieved May 4, 2020).[50]

2.2 | Cell culture

A CHO-K1 host cell line (a gift from NIH NIAID) was cultured at 37°C, 5% CO_2, and 80% humidity (Infors MultiTron, orbital diameter = 25.4 mm) in ActiPro medium (Cytiva) supplemented with 6 mM L-glutamine (Fisher). Cells were maintained in 50 mL mini bioreactor tubes (NEST Chemglass) at 200 rpm with a 10 mL working volume or 125 mL shake flasks at 135 rpm with a 20–30 mL working volume.

2.3 | Molecular cloning

A plasmid expressing the wild-type trastuzumab sequence (referred to as pACS-cTRMb throughout this work) was designed and built in-house using LC and HC DNA sequences obtained by reverse transcribing the trastuzumab amino acid sequence deposited in the International Immunogenetics Information System (IMGT) database (Entry code 7637).[51] Both genes were codon optimized for expression in CHO cells using the Codon Optimization Tool from Integrated DNA Technologies (IDT). Both LC and HC genes were expressed from separate cytomegalovirus promoters, with human serum albumin signal peptides and Chinese hamster growth hormone poly-adenylation signals (Figure S1A). Variant plasmids containing the mutations of interest were generated using Gibson assembly.[52] Briefly, restriction enzyme cut sites in the LC and HC coding sequences of pACS-cTRMB were used to excise portions of the variable domain sequences. EcoRI and KpnI were used to excise the LC variable domain, and HindIII and PstI were used to excise the HC variable domain (Figure S1B,C). All restriction enzymes were purchased from New England Biolabs. Gene fragments (eBlocks™, Integrated DNA Technologies) containing the desired mutation were inserted into the digested plasmid using NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs). Assembly was performed according to the manufacturer's protocol, using a basis of 70 ng of digested backbone plasmid and a 1:2 molar ratio of backbone to insert. A common codon was used for all mutations to the same amino acid (Table S1). Plasmids were transformed in TOP10 *Escherichia coli* (Thermo Fisher Scientific) and prepared using the QIAprep 96 Plus Miniprep Kit (Qiagen) according to the manufacturers' instructions. Variant sequences were confirmed individually by Sanger sequencing at the University of Delaware DNA Sequencing and Genotyping Center. The concentration of each sequence-confirmed variant plasmid was determined with the Quant-iT™ dsDNA broad range kit (Thermo Fisher Scientific). Plasmids were diluted 10× or 50× with water and the concentrations were determined from technical triplicates using a standard curve generated with the pACS-cTRMB plasmid. The complete sequence of the pACS-cTRMb plasmid and all gene fragments used for assembly can be found in Supplemental DNA Sequences folder.

2.4 | Batch screening of variants

Variants were split into four groups, two LC groups and two HC groups, and each group was transfected and grown separately. CHO-K1 host cells were passaged to 1×10^6 cells/mL with a complete media exchange on the day prior to transfection. Transfections were performed with a SG Cell Line 96-well Nucleofector™ Kit (Lonza) in a 4D-Nucleofector® unit equipped with the 96-well Shuttle™ Add-On (Lonza). For each transfection, 2×10^6 cells were transfected with 500 fmol of plasmid DNA (250 fmol DNA/million cells). Prior to transfection, plasmid DNA was aliquoted and dried in a SpeedVac® (Savant) for 20–40 min at 45°C then resuspended in supplemented SG Solution overnight at 4°C. For each group, all variants were transfected in duplicate, and the wild-type plasmid was transfected in triplicate. Due to concerns of promoter interference and resource competition, the wild-type plasmid transfections were also used as external transfection controls to monitor the reproducibility of our transfection methods.[53,54] Transfections were performed according to the manufacturer's protocol, using transfection code FF-137. After transfection, cells recovered in fresh media for 20 min, then each transfection was split between two 96-deep well plates (96-DWP), using a working volume of 500 µL (square wells, U-bottom, Biotix), for a total of four biological replicates per variant and six biological replicates for the wild-type mAb. Well plates were covered with BreatheEasy® membranes (Sigma-Aldrich) and grown at 350 rpm for 6 days. Starting on Day 1, every 24 h, culture samples were taken and assayed for growth by measuring adenosine triphosphate (ATP) using the CellTiter-Glo® 2.0 Cell Viability Assay (Promega). Cell culture samples were diluted 100x in PBS and measured in a 384-well plate (Corning) with four technical replicates according to the manufacturer's protocol. A 100 nM ATP standard (Sigma-Aldrich) was included as a positive control. A pseudo-integral of viable cell density (IVCD, RLU-day) for each culture was calculated from the luminescence values measured every day using a trapezoidal approximation. On Day 6, all remaining culture volume was centrifuged to remove cells and supernatant samples were retained for titer analysis. Titer (mg/L) was measured with an Easy-Titer™ Human IgG kit (Thermo Fisher Scientific) according to the manufacturer's protocol. Samples were diluted 50× in kit Dilution Buffer and measured in triplicate, with concentrations determined from a standard curve generated using the 100 µg/mL IgG standard from an Octet Protein A Calibrator Set (Sartorius). Productivities (mg/RLU/L/day) for each culture were calculated by

dividing the final Day 6 titer by the pseudo-IVCD. All cultures were then normalized by setting the average of the wild-type cultures expressed on the same well plate to 1.

2.5 | Statistical Analyses

R (v 4.0.4) and the tidyverse family of packages (v 1.3.0) were used for all statistical analyses and to generate plots. Unless otherwise noted, error bars represent one standard deviation from the mean of *n* biological replicate cultures, with *n* indicated in figure captions. The statistical tests used are provided in the figure captions or main text.

3 | RESULTS

3.1 | Systematic selection and manual curation of antibody variants in a model $IgG_1\kappa$ mAb generates a panel of mutations covering a wide range of biophysical changes

The humanized $IgG_1\kappa$ antibody trastuzumab was selected as a model mAb because of its therapeutic relevance, recognition as an "easy-to-express" mAb, and the availability of a high-quality crystal structure.[31,46,55] We aimed to compile a panel of single amino acid mutations containing $\sim 10^2$ trastuzumab variants that could be evaluated for expression in an industrially relevant CHO-K1 host, with particular focus on mutations that would minimally affect antigen binding, immune system interactions, and core antibody structure, but decrease mAb expression. The final panel, consisting of 178 single amino acid mutations, was chosen systematically and manually curated using published antibody structure information and computational protein analysis tools.

Filters were applied to select a subset of residues for mutation (File S2). Only the variable domains of the LC and HC were considered because these regions are more heavily engineered for improved affinity and are outside of the regions that interact with the immune system.[56] Residues identified as part of trastuzumab's functional paratope based on alanine scanning calculations of the antibody–antigen complex and previously published paratope information were also eliminated from consideration to avoid direct disruption of the antibody–antigen interface.[57] A similar filtering method using a combination of alanine scanning and mAb structural information was applied to identify residues involved in the LC-HC interface.[58,59] Next, any residue considered "invariant" or observed to contain "closely related" residues in more than 91% of sequences across species by Chothia and colleagues was also eliminated to avoid disruption of the common structural core of the antibody.[59] Applying these two criteria eliminated 37 residues from the LC and 38 residues from the HC. The free energy contribution of each residue on the individual antibody chains was then determined through computational alanine scanning to determine the changes in the free energy of unfolding upon mutation ($\Delta\Delta G$), where large free energy changes indicate that a given residue may contribute to protein stability. Calculations were performed using the FoldX package due to its computational efficiency for large datasets, and residues with large free energy changes ($\Delta\Delta G > 1$ kcal/mol) upon mutation to alanine were considered positions of interest, as described previously.[60] Twenty residues in the LC with $\Delta\Delta G > 1$ kcal/mol were selected, but for the HC, additional filtering was needed to reduce the number of positions evaluated due to the large number (45 residues) with alanine mutations resulting in $\Delta\Delta G > 1$ kcal/mol. The free energy change for all possible mutations at the 45 remaining HC positions was determined with FoldX, and any residue with an average $\Delta\Delta G < 2$ kcal/mol across all 20 mutations (16 residues) was eliminated. Residues in CDR-H3 were subsequently eliminated (6 residues) due to the role that CDR-H3 plays in antigen binding and its highly variable nature.[56,61] These additional filtering criteria resulted in the selection of 23 positions in the HC as positions of interest. The chosen residues in both chains represent a broad cross-section of positions throughout both the complementarity determining regions (CDRs) and framework regions (FWRs), including solvent-exposed and buried positions in and around the upper and lower hydrophobic cores and within inter-strand loops, all of which may contribute to mAb stability (Figure 1).[41,42,59,62] Applying these criteria independently to both chains also resulted in the selection of four structurally equivalent residue pairs between the LC and HC, based on Kabat definitions (LC11 = HC10, LC19 = HC18, LC33 = HC34, LC62 = HC67).[59]

Mutations at each position were selected based on the wild-type amino acid, with at least four mutations manually curated at each position to assess changes in sidechain chemistry, amino acid size, and hydropathy. Amino acids that are observed in the variable domains at a frequency of less than 2.5% (LC: W, H, E, M, N; HC: H, M, F) and any unique amino acids (C, G, P) were not used for mutation. The remaining amino acids were divided into four groups based on chemistry: aromatic (F, W, Y), charged (D, E, K, R), hydrophobic (A, I, L, V), and polar (N, Q, S, T). At least one amino acid from each chemistry group was chosen, with preference given to mutations predicted to be destabilizing ($\Delta\Delta G > 1$ Rosetta energy units) from Rosetta free energy calculations and mutations not deemed as "safe."[63] The final panel of trastuzumab variants included 178 single amino acid mutations, with 86 mutations in the LC and 92 mutations in the HC (Table 1). The wild-type trastuzumab sequence was used as a benchmark for a well-expressed mAb.

The distributions of differences in hydropathy, amino acid size, and percent SASA between the wild-type position and each mutation show that the chosen mutations cover a wide spectrum of biophysical changes, spanning much of the theoretical range of each characteristic (Figure S2). A panel of random mutations was also chosen for the same subset of positions to determine whether this rational selection method was effectively enriching for DTE variants. Random mutations were chosen to reflect the natural distribution of each amino acid (excluding unique amino acids) within the variable domains by weighing each potential amino acid change based on the amino acid's frequency of observation. Twenty-two mutations were chosen for the LC and 24 mutations were chosen for the HC (Table S2).

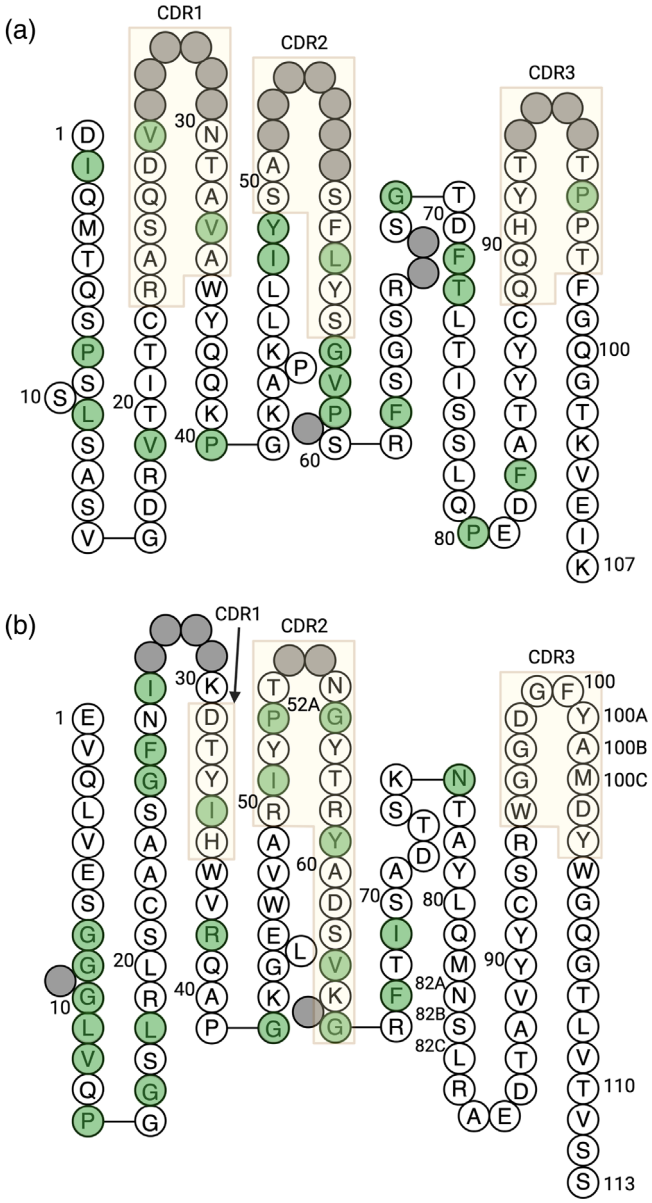

FIGURE 1 Collier-de-Perles (adapted from International Immunogenetics Information System [IMGT]) diagrams for trastuzumab indicating the chosen residues (highlighted in green) for mutation in (a) the trastuzumab light chain and (b) the trastuzumab heavy chain. Complementarity determining regions (CDRs) based on Kabat definitions are marked in gold boxes. Residues are numbered according to Kabat numbering. Gray circles indicate placeholder residues in the IMGT system that are not present in trastuzumab.

3.2 | Transient expression of trastuzumab variants identifies dozens of mutations that impact expression

All trastuzumab variants were transiently expressed in a CHO-K1 host cell line and titers were measured at the end of a 6-day batch process. Productivities for each biological replicate were determined using titer and pseudo-IVCD values, then normalized to the average expression of the wild-type samples expressed on the same 96-DWP (File S3). The average normalized productivity (NQP) of four biological replicates of each variant was compared to the overall distribution of all wild-type cultures (24 total wild-type replicates). In general, the expression distributions of all LC and HC variants were statistically different from the wild-type distribution (Figure 2a, Table S3).

Ninety-one variants (51.1%) had a lower NQP compared to wild-type trastuzumab (false discovery rate (FDR) < 0.05), with low-expressing mutations divided evenly between the two chains (48 in the HC, 43 in the LC). On average, the LC variants had lower productivities compared to the HC variants, with low-expressing LC variants ranging from 0.11× to 0.79× of wild-type and low-expressing HC variants ranging from 0.26× to 0.79× of wild-type. The highest-expressing LC variant (V58A) showed a NQP of 1.27, and the highest-expressing HC variant (G8W) showed a NQP of 2.69 but this variant had the lowest average IVCD of all variants by more than four-fold.

A total of 54.3% of variants in the randomly chosen panel were low-expressing (25 out of 46 mutations), with the distribution of LC variant expression showing a statistical difference to the wild-type distribution, as was observed with the rationally-chosen panel (Figure 2b). However, a smaller proportion of low-expressing variants was identified from the randomly chosen HC variants (9 out of 24 HC mutations, 37.5%) compared to the rationally-chosen panel (52.1%), with the random HC variants showing an overall higher distribution of expression values, demonstrating that the rational method used to select mutations in the HC was better at enriching for low-expressing constructs (Figure 2b, Table S3). This result also suggests that the filtering methods used to select positions of interest succeeded in finding positions with the potential to affect expression, regardless of the mutations studied. No significant difference in expression distribution was observed between the wild-type constructs expressed between the two panels, demonstrating a well-controlled transfection and cell culture process across assay occasions (Table S3).

Many positions showed clear preferences for amino acids with certain chemical and physical properties (Figures 3 and 4). LC-P8, an exposed residue, preferred small, polar amino acids (S and D), which aligns with the general observation that polar residues tend to reside on the surfaces of proteins. Preferences at LC-I48 and LC-V58 agreed with hydropathy conservation data with both positions showing expression similar to wild-type upon mutation to hydrophobic and aliphatic (Y and A) sidechains.[64] While conservation data shows no clear size preference at LC-F83, both mutations with small sidechains (S and A) resulted in low expression. In the HC, a short stretch of consecutive residues (HC-G8, HC-G9, HC-G10, HC-L11, HC-V12) did not tolerate mutation well (11/20 mutations were low-expressing) with HC-G9, HC-L11, and HC-V12 all showing preference for their wild-type amino acids. At HC-I34, hydrophobic residues (A and Y) were preferred over polar residues (K and S), again reflecting hydropathy conservation data. At HC-G42, larger amino acids (W and D) led to decreased expression, but mutations to threonine and valine, which are almost identical in shape and volume, showed no significant difference in expression. At HC-I69, all mutations were detrimental except for the mutation to alanine.

Comparing the results across structurally equivalent pairs showed similar sequence preferences across chains. At LC-L11/HC-G10,

TABLE 1 Rationally-chosen trastuzumab variant panel.

Wild-type residue [Kabat Positions]	Mutations				
	Aromatic (F, W, Y)	Charged (D, E, H, K, R)	Hydrophobic (A, I, L, M, V)	Polar (N, Q, S, T)	Additional mutations
Light chain					
F [62, 71, 83]	Y	K	A	S	F62T, F71T, F83T
G [57, 68]	F	K	V	T	G68D
I [2, 48]	Y	D	A	S	
L [11, 54]	F	D	A	Q	L11T
P [8, 40, 59, 80, 95]	F	D	I	S	
T [72]	Y	D	A	Q	
V [19, 29, 33, 58]	Y	K	A	Q	V19D, V19S, V33S
Y [49]	F	D	A	S	
Heavy chain					
F [27, 67]	W	K	A	S	
G [8, 9, 10, 16, 26, 42, 55, 65]	W	D	V	T	
I [29, 34, 51, 69]	Y	K	A	S	
L [11, 18]	Y	D	A	S	
N [76]	Y	K	I	T	
P [14, 52A]	Y	D	I	S	
R [38]	W	D	L	S	
V [12, 63]	W	D	I	N	
Y [59]	W	D	A	S	

Note: Positions of interest were grouped by wild-type amino acid and mutated to the same four or five amino acids, covering a range of sidechain amino acid chemistries, sizes, and hydropathies. Additional mutations were added to the light chain panel to match mutations chosen for the heavy chain at structurally equivalent locations. Positions are given based on Kabat numbering.

mutations to aromatic and charged residues had no impact on expression, and mutations to threonine resulted in decreased expression. At LC-V19/HC-L18, aromatic mutations were tolerated, and mutations to aspartic acid decreased expression at both locations. Hydrophobic mutations were tolerated at LC-V33/HC-I34, but charged and polar mutations were not. Almost all mutations made at positions LC-F62/HC-F67 resulted in low expression. Three of the mutations chosen (LC-I2S, LC-V19A, and LC-F71A) are observed frequently in antibodies, with observation rates of the mutated amino acid greater than 30%.[50] As expected, all three of these mutations did not show a significant difference in expression compared to wild-type.

3.3 | Twelve positions are less tolerant to mutations

Twelve of the positions studied (LC-V33, LC-F62, LC-G68, LC-F71, LC-P80, HC-G9, HC-L11, HC-V12, HC-R38, HC-V63, HC-F67, HC-I69) had three or more mutations identified as low-expressing, suggesting that these positions are less tolerant to mutations. Included in this subset is the LC-F62/HC-F67 structural pair, in which 8 of the 9 studied mutations decreased productivity. LC-F62 is a conserved residue, with phenylalanine observed at this location in >80% of human antibody sequences, and hydrophobic or neutral residues observed in >99% of sequences across species.[59,64] A similar intolerance for mutations other than alanine at the equivalent HC position suggests a common structural role of this position in antibody variable domains although HC-F67 shows less stringent conservation preferences, with valine commonly observed.[50,65] Position LC-V33 was also identified as less tolerant to mutation, possibly due to its importance in determining the conformation of LC-CDR1.[66] The residue in position HC-9 is a key residue for defining the backbone conformation of the N-terminal segment of VH domains, with the preferred residues at this position defined by the amino acid identities at positions HC-6 and HC-7.[67] The other identified positions have not been previously implicated as important for expression.

3.4 | Enrichment of low-expressing variants was observed in certain molecular contexts

Despite the large number of low-expressing variants identified in this study, attempts to predict expression based on biophysical characteristics did not yield practically useful models. Pearson correlation

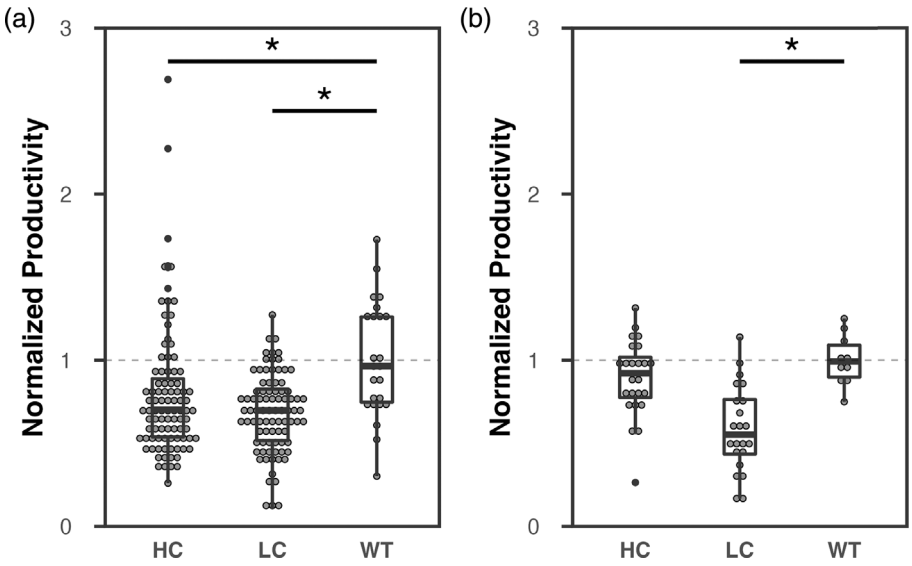

FIGURE 2 Normalized productivity distributions for the (a) rationally-selected panel and the (b) randomly-selected panel for heavy chain (HC) variants, light chain variants (LC) and wild-type trastuzumab (WT). Asterisks indicate $p < 0.05$, two-sample, two-tailed Mann–Whitney U test.

coefficients showed weak associations between the measured parameters, and principal component analysis could only account for approximately 56% of the variation observed in the data (data not shown). Previous studies on mutational landscapes have shown that sensitivity to amino acid change is dependent on the extent of side-chain burial so variants were grouped based on the surface accessibility of the wild-type residue to investigate position-dependent mutation preferences.[41,68] Positions were divided into three categories, exposed (>50 Å^2), partial (20–50 Å^2), and inaccessible (<20 Å^2), based on areas calculated in Pymol, using the area ranges defined by Chothia.[59] This grouping revealed certain molecular contexts that appear to influence a position's tolerance to mutation (Table 2). Low-expressing variants were enriched at inaccessible residues (hypergeometric, $p < 0.1$), and further grouping of the data based on the chemistry of the mutated sidechain showed that mutating inaccessible residues to polar or charged sidechains often led to low expression. This trend was observed in the LC, with both polar and charged mutations showing enrichment, but only mutations to polar sidechains were identified as enriched for the HC (hypergeometric, $p < 0.1$). These findings align with previous observations that protein cores strongly prefer hydrophobic residues but can tolerate alternative hydrophobic amino acids at those positions.[69] In general, large decreases in hydrophobicity upon mutation resulted in reduced expression, regardless of solvent accessibility.

4 | DISCUSSION

The rational selection and expression of 178 single amino acid mutations in trastuzumab led to the identification of dozens of novel mutations that can impact mAb expression, with approximately 50% of all studied mutations resulting in lower expression compared to the wild-type mAb. By considering mutations that encompass a wide range of different biophysical changes at each selected location, a comprehensive picture of mutation tolerance was constructed for many amino acid positions, including some that have not been previously linked to expression. As a result of this systematic selection process, chemical and physical amino acid preferences could be identified at each location studied which would not have been possible with random mutagenesis, with the strongest preferences observed for hydropathy and size. As hydrophobic effects are believed to be the primary driving forces of protein folding, mutations that significantly disrupt key interactions may affect the folding process.[70] Indeed, previous work with a simple α/β protein showed that aggregation rates of protein variants correlated with hydrophobicity changes in the regions containing mutations, driving continued work to understand and predict the sequence-level factors influencing protein aggregation.[71,72] The observation of size preferences in the absence of chemical class conservation, such as in the case of HC-G42, suggests that protein folding accommodations made to account for steric hindrance can also affect expression.

The selection criteria used to build the variant panel consciously prioritized potentially detrimental mutations at residues hypothesized to contribute to mAb stability, all with the intent to identify mutations that decrease mAb expression. As expected, our work did not identify any mutations that meaningfully improve the expression of trastuzumab, which is already considered an "easy-to-express" mAb and is likely already well-engineering for high expression in CHO cells.[31] The variants that did show nominal increases in productivity (such as HC-G8W and HC-G16W) also had dramatically reduced IVCD values, suggesting that these mutations are significantly detrimental to cell growth and therefore do not represent a practical mechanism for improving expression.

Some of the positions and mutations studied here have been previously identified as detrimental to expression in other mAb formats and cell lines. Mutations at LC-F62 in a lambda LC decreased production of IgM in COS-1 cells, where disruption of the LC structure prevented the LC from being transported from the ER to the Golgi.[32] Different combinations of amino acids at HC-6, HC-7, and HC-9 were found to impact the expression of single-chain variable fragment

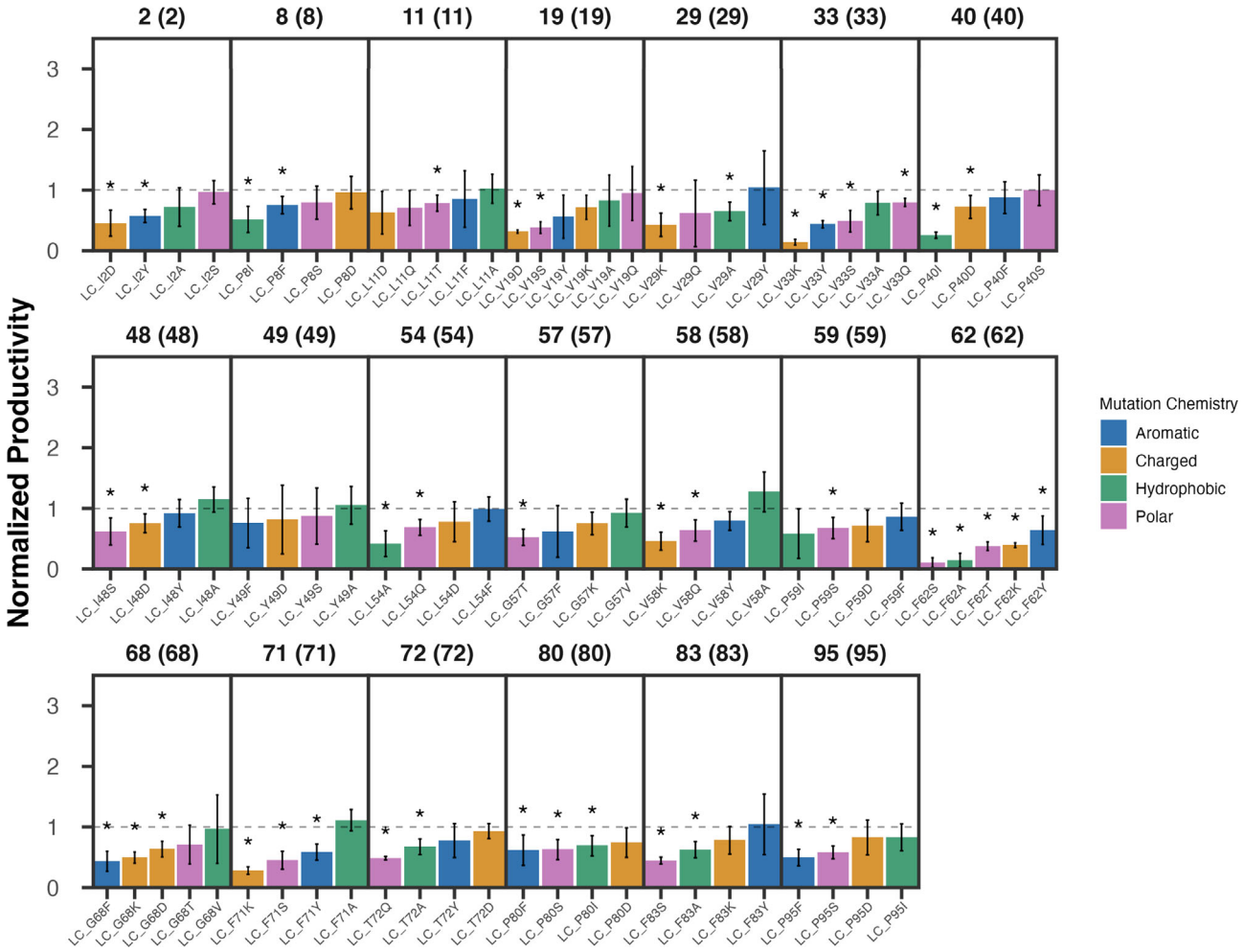

FIGURE 3 Average normalized productivities for all mutations in the rationally-chosen panel in the light chain (LC) colored by the sidechain chemistry of the mutated amino acid. Error bars represent one standard deviation from four biological replicates. Asterisks indicate mutations identified as lower-expressing than wild-type trastuzumab (FDR < 0.05). Position numbers above the subplots are given based on sequence order with the corresponding Kabat number in parentheses. Bar labels use the format [chain]_[wild type amino acid][sequence number][mutated amino acid].

constructs in *E. coli*, with yields differing by more than ten-fold.[67] Multiple mutations made at HC-I51 in an IgG_2 expressed in hybridoma cells decreased secretion by impairing LC-HC pairing, but also caused persistent association of the HC with ER folding chaperones.[34] While our data showed that the HC-G42V mutation had no detrimental impact on expression, mutating valine back to the more frequently-observed glycine in combination with other mutations improved titer and drug product stability, suggesting that HC-42 may work in concert with other residues to influence developability properties.[73] Other studies identified detrimental mutations in positions not studied here but in positions known to interact with residues in our panel. Changing an alanine to a glycine at HC-49 in an IgG_4 resulted in a non-secreted HC variant, with the equivalent position in trastuzumab (HC-A49) interacting with HC-34, HC-51, HC-59, and HC-69, all studied here.[28,51] The expression profiles for HC-34 and HC-69 showed preferences for aromatic and hydrophobic amino acids, and the clear preference for aliphatic residues at HC-69 suggests that this chemistry class may be necessary at this location for high expression. The conservation of many amino acid networks in antibodies has been shown, and in some cases, the disruption of either residue involved in a key interaction can lead to destabilizing changes to protein structure.[74,75]

Early work to determine amino acid preferences relied on sequence and structural alignments from large datasets reflecting the diversity of naturally occurring antibodies.[59,66,76] However, the sequences contained in these datasets are inherently limited by evolution, where certain mutations may be less common due to the number of base pair changes required or limited functional benefits preventing them from becoming prevalent.[77] This work moves beyond studies focused on natural diversity and better aligns with the diversity that would be observed in synthetic antibody discovery platforms, which are not limited by evolution and can evaluate many amino acids at each position of interest.[38] These platforms use antigen affinity as a primary screening technique, which can result in the

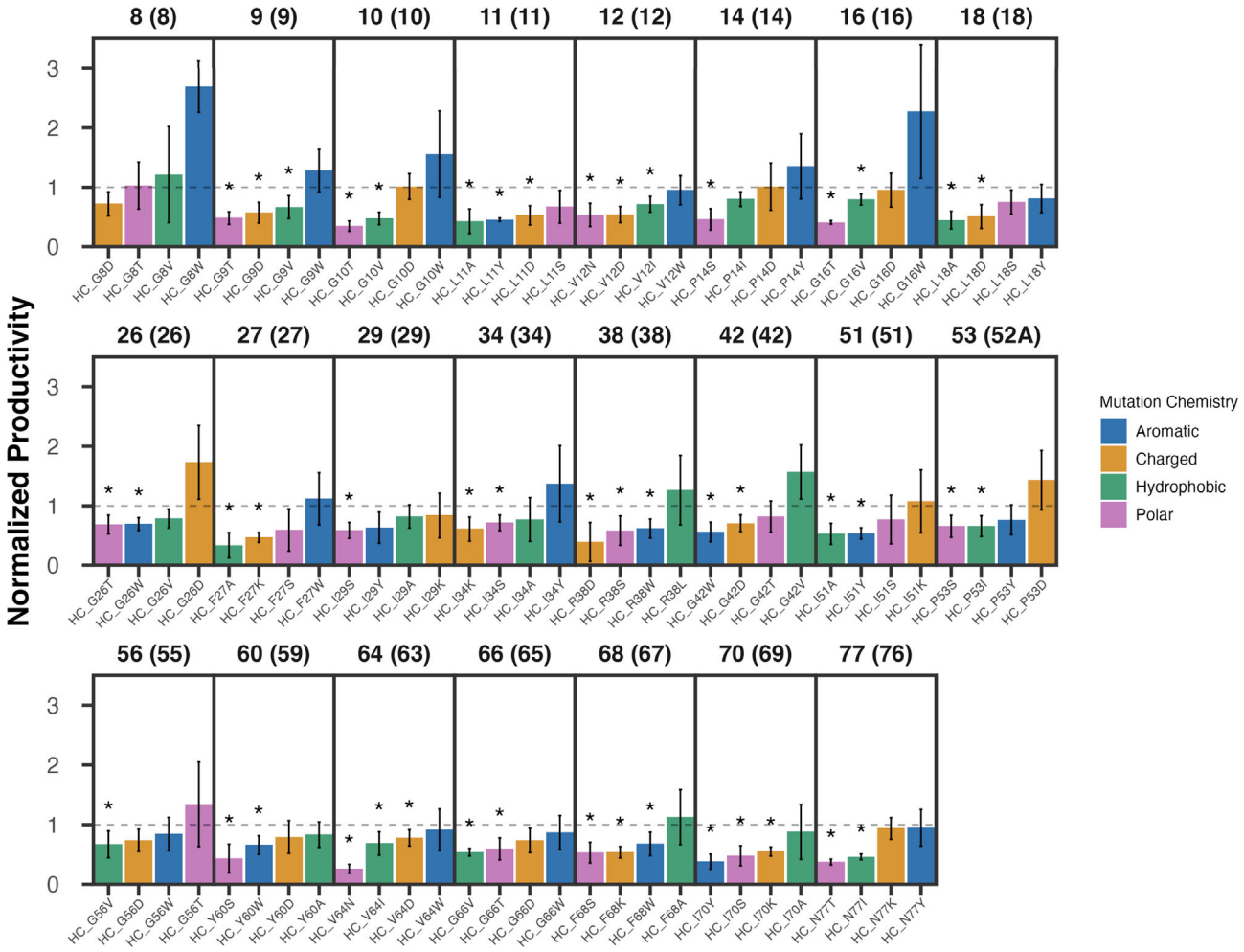

FIGURE 4 Average normalized productivities for all mutations in the rationally-chosen panel in the heavy chain (HC) colored by the sidechain chemistry of the mutated amino acid. Error bars represent one standard deviation from four biological replicates. Asterisks indicate mutations identified as lower-expressing than wild-type trastuzumab (FDR < 0.05). Position numbers above the subplots are given based on sequence order with the corresponding Kabat number in parentheses. Bar labels use the format [chain]_[wild type amino acid][sequence number][mutated amino acid].

identification of functionally favorable mAb candidates that may possess additional amino acid motifs that negatively impact developability characteristics.[2,4]

Algorithms and workflows for calculating and predicting various developability factors have become increasingly popular within the biopharmaceutical industry, as early identification of unfavorable mAb candidates can better guide candidate selection procedures to address process development challenges such as low titers, precipitation during purification steps, and drug product aggregation.[11,30,73] In silico and in vitro methods are of particular interest, as these methods are amenable to high throughput analysis of mAb candidates.[9] In silico methods to predict developability characteristics, such as aggregation propensity and chemical modifications, have set predictive thresholds for these quantities that may indicate poor developability.[5,8,10,72,78,79] In addition, the emergence of comprehensive mAb databases such as abYsis and IMGT have enabled extensive structural and statistical analyses of clinically-relevant mAbs.[50,51] While biopharmaceutical companies may have proprietary tools for screening mAb candidate expression in the context of their platforms, no algorithm to predict mAb expression in CHO as a function of sequence has currently been published for public use. Incorporating early expression predictions into candidate selection would help prevent resource-intensive process development activities in the future. Additional work to assess the findings of this study in a representative stable expression system would further strengthen these conclusions by confirming that the low-expressing phenotypes observed here are preserved with stable systems, which are the manufacturing platform of choice for the biopharmaceutical industry.

5 | CONCLUSION

This work presents the largest dataset of expression-impacting mAb mutations as measured by secreted IgG_1 titers in CHO published to

TABLE 2 Hypergeometric tests to identify molecular contexts detrimental to mAb expression.

Category	Number of Variants	*p*-Value	Significant	Not significant
All	178	n/a	91	87
LC	86	0.556	43	43
HC	92	0.330	48	44
Inaccessible*	80	**0.011**	48	32
LC*	36	**0.063**	22	14
HC*	44	**0.082**	26	18
Aromatic	19	0.722	8	11
LC	8	0.385	4	4
HC	11	0.757	4	7
Charged*	20	**0.059**	13	7
LC*	9	**0.020**	7	2
HC	11	0.294	6	5
Hydrophobic	19	0.542	9	10
LC	8	0.154	5	3
HC	11	0.294	6	5
Polar*	22	**0.0003**	18	4
LC*	11	**0.034**	8	3
HC*	11	**0.0005**	10	1
Partial	54	0.641	26	28
LC	22	0.455	11	11
HC	32	0.632	15	17
Aromatic	13	0.892	4	9
LC	5	0.521	2	3
HC	8	0.875	2	6
Charged	14	0.425	7	7
LC	6	0.363	3	3
HC	8	0.385	4	4
Hydrophobic	13	0.143	8	5
LC	5	0.521	2	3
HC*	8	**0.037**	6	2
Polar	14	0.425	7	7
LC	6	0.116	4	2
HC	8	0.665	3	4
Exposed	44	0.959	17	27
LC	28	0.942	10	18
HC	16	0.639	7	9
Aromatic	11	0.757	4	7
LC	7	0.796	2	5
HC	4	0.328	2	2
Charged	11	0.976	2	9
LC	7	0.948	1	6
HC	4	0.706	1	3
Hydrophobic	11	0.757	4	7
LC	7	0.525	3	4
HC	4	0.706	1	3

TABLE 2 (Continued)

Category	Number of Variants	*p*-Value	Significant	Not significant
Polar	11	0.121	7	4
LC	7	0.241	4	3
HC*	4	**0.066**	3	1

Note: Variants were divided by accessibility (inaccessible, partial, and exposed), mutation chemistry (aromatic, charged, hydrophobic, polar), and chain (light chain, LC; heavy chain, HC). Total number of variants, the number of variants in each grouping identified as significantly lower expressing than wild-type (Significant) and variants not significantly lower than wild-type (Not significant) are shown. *P*-values represent the probability of observing more than the identified number of significant variants according to the hypergeometric distribution. Groups with *p*-values <0.1 (*p*-values shown in bold and category marked with an asterisk) were considered enriched for low-expressing variants.

date. Previous studies of DTE mAbs in mammalian hosts have included less than a dozen mAb variants, whereas this study increases throughput by an order of magnitude while assessing natural IgG expression through mammalian protein processing pathways in CHO cells, the major cell type used for mAb production in the biopharmaceutical industry.[3] Using a small-scale transient expression system allowed for rapid evaluation of 178 mAb variants in parallel and careful curation of the variant panel ensured that mutations were evaluated at positions throughout the antibody variable domains with measurable contributions to protein structure but minimal predicted impact on affinity and immune system activation. Mutating each position to a variety of amino acids with different sidechain chemistries and biophysical characteristics provided a more complete mutational landscape at many residues than had been previously documented and identified 12 positions with clear amino acid requirements for maintaining high mAb expression. Although combining the expression data with calculated biophysical characteristics failed to generate a useful predictive model, the data showed that inaccessible residues in the protein do not tolerate mutation well, and mutations at inaccessible residues to amino acids with charged or polar sidechains generally reduced mAb expression. Exploring this observation further with other clinically relevant mAbs will be an important next step towards sequence-level rational design of well-expressed mAbs.

The use of a transient expression system in this work presents an experimental framework that allows for flexibility in host cell choice and creates a workflow amenable to further throughput improvements with automation. Application across many clinically-relevant antibodies from different industry platforms and the incorporation of combinatorial studies to identify crucial dependencies in amino acid networks could eventually generate enough data to create an in silico model capable of predicting expression from primary sequence or identifying sequence motifs that are undesirable for high productivity. Application of predictive tools can improve candidate selection procedures, provide early warnings of molecules with low developability, or construct mAb sequence frameworks optimized for high expression, all of which could contribute to accelerated process development timelines and reduced resource requirements for bringing novel antibody therapies into the clinic.

AUTHOR CONTRIBUTIONS

Kelvin H. Lee: Funding acquisition; writing – review and editing; project administration; supervision; resources. **Alana C. Szkodny:** Writing – original draft; methodology; writing – review and editing; formal analysis; conceptualization; investigation.

ACKNOWLEDGMENTS

Computational work was performed on the BIOMIX computing cluster with support from the University of Delaware Center for Bioinformatics and Computational Biology Core Facility and funding from Delaware INBRE (NIGMS P20GM103446), the State of Delaware, and the Delaware Biotechnology Institute. The CHO-K1 host cell line used for this work was obtained from the NIH (NIAID). This work was supported by funding from the National Science Foundation under grant 1624698, and in part by the financial assistance award 70NANB17H002 from the U.S. Department of Commerce and National Institutes of Standards and Technology.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Alana C. Szkodny https://orcid.org/0000-0002-9071-9363
Kelvin H. Lee https://orcid.org/0000-0003-0908-2981

REFERENCES

1. Winter G, Milstein C. Man-made antibodies. *Nature*. 1991;349(6307):293-299. doi:10.1038/349293a0
2. Tiller KE, Tessier PM. Advances in antibody design. *Annu Rev Biomed Eng*. 2015;17(1):191-216. doi:10.1146/annurev-bioeng-071114-040733
3. Walsh G, Walsh E. Biopharmaceutical benchmarks 2022. *Nat Biotechnol*. 2022;40(12):1722-1760. doi:10.1038/s41587-022-01582-x
4. Gilliland GL, Luo J, Vafa O, Almagro JC. Leveraging SBDD in protein therapeutic development: antibody engineering. In: Tari LW, ed. *Structure-Based Drug Discovery*. Vol 841. Methods in Molecular Biology. Humana Press; 2012:321-349. doi:10.1007/978-1-61779-520-6_14

5. Jain T, Sun T, Durand S, et al. Biophysical properties of the clinical-stage antibody landscape. *Proc Natl Acad Sci*. 2017;114(5):944-949. doi:10.1073/pnas.1616408114
6. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 1997; 23(1–3):3-25. doi:10.1016/S0169-409X(96)00423-1
7. Khetan R, Curtis R, Deane CM, et al. Current advances in biopharmaceutical informatics: guidelines, impact and challenges in the computational developability assessment of antibody therapeutics. *MAbs*. 2022;14(1):2020082. doi:10.1080/19420862.2021.2020082
8. Sharma VK, Patapoff TW, Kabakoff B, et al. In silico selection of therapeutic antibodies for development: viscosity, clearance, and chemical stability. *Proc Natl Acad Sci*. 2014;111(52):18601-18606. doi:10.1073/pnas.1421779112
9. Jarasch A, Koll H, Regula JT, Bader M, Papadimitriou A, Kettenberger H. Developability assessment during the selection of novel therapeutic antibodies. *J Pharm Sci*. 2015;104(6):1885-1898. doi:10.1002/jps.24430
10. Raybould MIJ, Marks C, Krawczyk K, et al. Five computational developability guidelines for therapeutic antibody profiling. *Proc Natl Acad Sci*. 2019;116(10):4025-4030. doi:10.1073/pnas.1810576116
11. Bailly M, Mieczkowski C, Juan V, et al. Predicting antibody developability profiles through early stage discovery screening. *MAbs*. 2020; 12(1):1743053. doi:10.1080/19420862.2020.1743053
12. Kelley B. Industrialization of mAb production technology: the bioprocessing industry at a crossroads. *MAbs*. 2009;1(5):443-452. doi:10.4161/mabs.1.5.9448
13. Ho SCL, Koh EYC, van Beers M, et al. Control of IgG LC:HC ratio in stably transfected CHO cells and study of the impact on expression, aggregation, glycosylation and conformational stability. *J Biotechnol*. 2013;165(3):157-166. doi:10.1016/j.jbiotec.2013.03.019
14. Carver J, Ng D, Zhou M, et al. Maximizing antibody production in a targeted integration host by optimization of subunit gene dosage and position. *Biotechnol Prog*. 2020;36(4):e2967. doi:10.1002/btpr.2967
15. Lee CJ, Seth G, Tsukuda J, Hamilton RW. A clone screening method using mRNA levels to determine specific productivity and product quality for monoclonal antibodies. *Biotechnol Bioeng*. 2009;102(4): 1107-1118. doi:10.1002/bit.22126
16. Zhang L, Inniss MC, Han S, et al. Recombinase-mediated cassette exchange (RMCE) for monoclonal antibody expression in the commercially relevant CHOK1SV cell line. *Biotechnol Prog*. 2015;31(6): 1645-1656. doi:10.1002/btpr.2175
17. Hilliard W, Lee KH. A compendium of stable hotspots in the CHO genome. *Biotechnol Bioeng*. 2023;120:2133-2143. doi:10.1002/bit.28390
18. Borth N, Mattanovich D, Kunert R, Katinger H. Effect of increased expression of protein disulfide isomerase and heavy chain binding protein on antibody secretion in a recombinant CHO cell line. *Biotechnol Prog*. 2005;21(1):106-111. doi:10.1021/bp0498241
19. Tigges M, Fussenegger M. Xbp1-based engineering of secretory capacity enhances the productivity of Chinese hamster ovary cells. *Metab Eng*. 2006;8(3):264-272. doi:10.1016/j.ymben.2006.01.006
20. Pybus LP, Dean G, West NR, et al. Model-directed engineering of "difficult-to-express" monoclonal antibody production by Chinese hamster ovary cells: improving difficult-to-express mAb expression. *Biotechnol Bioeng*. 2014;111(2):372-385. doi:10.1002/bit.25116
21. Johari YB, Estes SD, Alves CS, Sinacore MS, James DC. Integrated cell and process engineering for improved transient production of a "difficult-to-express" fusion protein by CHO cells. *Biotechnol Bioeng*. 2015; 112(12):2527-2542. doi:10.1002/bit.25687
22. Tang D, Lam C, Bauer N, et al. Bax and Bak knockout apoptosis-resistant Chinese hamster ovary cell lines significantly improve culture viability and titer in intensified fed-batch culture process. *Biotechnol Prog*. 2022;38(2):e3228. doi:10.1002/btpr.3228
23. Kim SH, Baek M, Park S, Shin S, Lee JS, Lee GM. Improving the secretory capacity of CHO producer cells: the effect of controlled Blimp1 expression, a master transcription factor for plasma cells. *Metab Eng*. 2022;69:73-86. doi:10.1016/j.ymben.2021.11.001
24. Jadhav V, Hackl M, Klanert G, et al. Stable overexpression of miR-17 enhances recombinant protein production of CHO cells. *J Biotechnol*. 2014;175:38-44. doi:10.1016/j.jbiotec.2014.01.032
25. Fischer S, Marquart KF, Pieper LA, et al. miRNA engineering of CHO cells facilitates production of difficult-to-express proteins and increases success in cell line development. *Biotechnol Bioeng*. 2017; 114(7):1495-1510. doi:10.1002/bit.26280
26. Hansen HG, Pristovšek N, Kildegaard HF, Lee GM. Improving the secretory capacity of Chinese hamster ovary cells by ectopic expression of effector genes: lessons learned and future directions. *Biotechnol Adv*. 2017;35(1):64-76. doi:10.1016/j.biotechadv.2016.11.008
27. Hasegawa H, Hsu A, Tinberg CE, Siegler KE, Nazarian AA, Tsai MM. Single amino acid substitution in LC-CDR1 induces Russell body phenotype that attenuates cellular protein synthesis through eIF2α phosphorylation and thereby downregulates IgG secretion despite operational secretory pathway traffic. *MAbs*. 2017;9(5):854-873. doi:10.1080/19420862.2017.1314875
28. Mason M, Sweeney B, Cain K, Stephens P, Sharfstein ST. Identifying bottlenecks in transient and stable production of recombinant monoclonal-antibody sequence variants in Chinese hamster ovary cells. *Biotechnol Prog*. 2012;28(3):846-855. doi:10.1002/btpr.1542
29. Tadauchi T, Lam C, Liu L, et al. Utilizing a regulated targeted integration cell line development approach to systematically investigate what makes an antibody difficult to express. *Biotechnol Prog*. 2019; 35(2):e2772. doi:10.1002/btpr.2772
30. Bauer J, Mathias S, Kube S, et al. Rational optimization of a monoclonal antibody improves the aggregation propensity and enhances the CMC properties along the entire pharmaceutical process chain. *MAbs*. 2020;12(1):1787121. doi:10.1080/19420862.2020.1787121
31. Le Fourn V, Girod PA, Buceta M, Regamey A, Mermod N. CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion. *Metab Eng*. 2014;21:91-102. doi:10.1016/j.ymben.2012.12.003
32. Dul JL, Argon Y. A single amino acid substitution in the variable region of the light chain specifically blocks immunoglobulin secretion. *Proc Natl Acad Sci*. 1990;87(20):8135-8139. doi:10.1073/pnas.87.20.8135
33. Chen C, Martin TM, Stevens S, Rittenberg MB. Defective secretion of an immunoglobulin caused by mutations in the heavy chain complementarity determining region 2. *J Exp Med*. 1994;180(2):577-586. doi:10.1084/jem.180.2.577
34. Wiens GD, Lekkerkerker A, Veltman I, Rittenberg MB. Mutation of a single conserved residue in VH complementarity-determining region 2 results in a severe Ig secretion defect. *J Immunol*. 2001;167(4): 2179-2186. doi:10.4049/jimmunol.167.4.2179
35. Reinhart D, Sommeregger W, Debreczeny M, Gludovacz E, Kunert R. In search of expression bottlenecks in recombinant CHO cell lines—a case study. *Appl Microbiol Biotechnol*. 2014;98(13):5959-5965. doi:10.1007/s00253-014-5584-z
36. Mathias S, Wippermann A, Raab N, et al. Unraveling what makes a monoclonal antibody difficult-to-express: from intracellular accumulation to incomplete folding and degradation via ERAD. *Biotechnol Bioeng*. 2020;117(1):5-16. doi:10.1002/bit.27196
37. O'Callaghan PM, McLeod J, Pybus LP, et al. Cell line-specific control of recombinant monoclonal antibody production by CHO cells. *Biotechnol Bioeng*. 2010;106(6):938-951. doi:10.1002/bit.22769

38. Hoogenboom HR. Selecting and screening recombinant antibody libraries. *Nat Biotechnol*. 2005;23(9):1105-1116. doi:10.1038/nbt1126
39. Koenig P, Lee CV, Walters BT, et al. Mutational landscape of antibody variable domains reveals a switch modulating the interdomain conformational dynamics and antigen binding. *Proc Natl Acad Sci*. 2017;114(4):E486-E495. doi:10.1073/pnas.1613231114
40. Dudgeon K, Rouet R, Kokmeijer I, et al. General strategy for the generation of human antibody variable domains with increased aggregation resistance. *Proc Natl Acad Sci*. 2012;109(27):10879-10884. doi:10.1073/pnas.1202866109
41. Hsu HJ, Lee KH, Jian JW, et al. Antibody variable domain Interface and framework sequence requirements for stability and function by high-throughput experiments. *Structure*. 2014;22(1):22-34. doi:10.1016/j.str.2013.10.006
42. Chang HJ, Jian JW, Hsu HJ, et al. Loop-sequence features and stability determinants in antibody variable domains by high-throughput experiments. *Structure*. 2014;22(1):9-21. doi:10.1016/j.str.2013.10.005
43. Parthiban K, Perera RL, Sattar M, et al. A comprehensive search of functional sequence space using large mammalian display libraries created by gene editing. *MAbs*. 2019;11(5):884-898. doi:10.1080/19420862.2019.1618673
44. Parola C, Neumeier D, Friedensohn S, et al. Antibody discovery and engineering by enhanced CRISPR-Cas9 integration of variable gene cassette libraries in mammalian cells. *MAbs*. 2019;11(8):1367-1380. doi:10.1080/19420862.2019.1662691
45. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids Res*. 2005;33(Web Server):W382-W388. doi:10.1093/nar/gki387
46. Cho HS, Mason K, Ramyar KX, et al. Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature*. 2003;421(6924):756-760. doi:10.1038/nature01392
47. Kellogg EH, Leaver-Fay A, Baker D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability: conformational sampling in computing mutation-induced changes. *Proteins Struct Funct Bioinforma*. 2011;79(3):830-838. doi:10.1002/prot.22921
48. Lehninger AL, Nelson DL, Cox MM. *Lehninger Principles of Biochemistry*. 5th ed. W.H. Freeman; 2008.
49. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*. 1982;157(1):105-132. doi:10.1016/0022-2836(82)90515-0
50. Swindells MB, Porter CT, Couch M, et al. abYsis: integrated antibody sequence and structure—management, analysis, and prediction. *J Mol Biol*. 2017;429(3):356-364. doi:10.1016/j.jmb.2016.08.019
51. Kaas Q. IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. *Nucleic Acids Res*. 2004;32(90001):208D-210D. doi:10.1093/nar/gkh042
52. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*. 2009;6(5):343-345. doi:10.1038/nmeth.1318
53. Di Blasi R, Marbiah MM, Siciliano V, Polizzi K, Ceroni F. A call for caution in analysing mammalian co-transfection experiments and implications of resource competition in data misinterpretation. *Nat Commun*. 2021;12(1):2545. doi:10.1038/s41467-021-22795-9
54. Brown AJ, Sweeney B, Mainwaring DO, James DC. Synthetic promoters for CHO cell engineering. *Biotechnol Bioeng*. 2014;111(8):1638-1647. doi:10.1002/bit.25227
55. Mullard A. FDA approves 100th monoclonal antibody product. *Nat Rev Drug Discov*. 2021;20(7):491-495. doi:10.1038/d41573-021-00079-7
56. Chiu ML, Goulet DR, Teplyakov A, Gilliland GL. Antibody structure and function: the basis for engineering therapeutics. *Antibodies*. 2019;8(4):55. doi:10.3390/antib8040055
57. Bostrom J, Yu SF, Kan D, et al. Variants of the antibody herceptin that interact with HER2 and VEGF at the antigen binding site. *Science*. 2009;323(5921):1610-1614. doi:10.1126/science.1165480
58. Chothia C, Novotný J, Bruccoleri R, Karplus M. Domain association in immunoglobulin molecules. *J Mol Biol*. 1985;186(3):651-663. doi:10.1111/j.1742-4658.2011.08207.x
59. Chothia C, Gelfand I, Kister A. Structural determinants in the sequences of immunoglobulin variable domain. *J Mol Biol*. 1998;278:457-479.
60. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein–protein complexes. *Proc Natl Acad Sci*. 2002;99(22):14116-14121. doi:10.1073/pnas.202485799
61. Xu JL, Davis MM. Diversity in the CDR3 region of VH is sufficient for most antibody specificities. *Immunity*. 2000;13(1):37-45. doi:10.1016/S1074-7613(00)00006-6
62. Kuramochi T, Igawa T, Tsunoda H, Hattori K. Humanization and simultaneous optimization of monoclonal antibody. In: Steinitz M, ed. *Human Monoclonal Antibodies*. Vol 1060. Methods in Molecular Biology. Humana Press; 2014:123-137. doi:10.1007/978-1-62703-586-6_7
63. Bordo D, Argos P. Suggestions for "safe" residue substitutions in site-directed mutagenesis. *J Mol Biol*. 1991;217(4):721-729. doi:10.1016/0022-2836(91)90528-E
64. Pommié C, Levadoux S, Sabatier R, Lefranc G, Lefranc MP. IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. *J Mol Recognit*. 2004;17(1):17-32. doi:10.1002/jmr.647
65. Knappik A, Ge L, Honegger A, et al. Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J Mol Biol*. 2000;296(1):57-86. doi:10.1006/jmbi.1999.3444
66. Chothia C, Lesk AM, Tramontano A, et al. Conformations of immunoglobulin hypervariable regions. *Nature*. 1989;342(6252):877-883. doi:10.1038/342877a0
67. Jung S, Spinelli S, Schimmele B, et al. The importance of framework residues H6, H7 and H10 in antibody heavy chains: experimental evidence for a new structural subclassification of antibody VH domains. *J Mol Biol*. 2001;309(3):701-716. doi:10.1006/jmbi.2001.4665
68. Nisthal A, Wang CY, Ary ML, Mayo SL. Protein stability engineering insights revealed by domain-wide comprehensive mutagenesis. *Proc Natl Acad Sci*. 2019;116(33):16367-16377. doi:10.1073/pnas.1903888116
69. Bowie JU, Reidhaar-Olson JF, Lim WA, Sauer RT. Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science*. 1990;247(4948):1306-1310. doi:10.1126/science.2315699
70. Kessel A, Ben-Tal N. *Introduction to Proteins: Structure, Function, and Motion*. CRC Press; 2011.
71. Chiti F, Taddei N, Baroni F, et al. Kinetic partitioning of protein folding and aggregation. *Nat Struct Biol*. 2002;9(2):137-143. doi:10.1038/nsb752
72. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol*. 2004;22(10):1302-1306. doi:10.1038/nbt1012
73. Seeliger D, Schulz P, Litzenburger T, et al. Boosting antibody developability through rational sequence optimization. *MAbs*. 2015;7(3):505-515. doi:10.1080/19420862.2015.1017695
74. Wang N, Smith WF, Miller BR, et al. Conserved amino acid networks involved in antibody variable domain interactions. *Proteins Struct Funct Bioinforma*. 2009;76(1):99-114. doi:10.1002/prot.22319

75. Helms LR, Wetzel R. Specificity of abnormal assembly in immunoglobulin light chain deposition disease and amyloidosis. *J Mol Biol*. 1996;257(1):77-86. doi:10.1006/jmbi.1996.0148
76. Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol*. 1987;196(4):901-917. doi:10.1016/0022-2836(87)90412-8
77. Hietpas RT, Jensen JD, Bolon DNA. Experimental illumination of a fitness landscape. *Proc Natl Acad Sci*. 2011;108(19):7896-7901. doi:10.1073/pnas.1016024108
78. Chennamsetty N, Voynov V, Kayser V, Helk B, Trout BL. Design of therapeutic proteins with enhanced stability. *Proc Natl Acad Sci*. 2009;106(29):11937-11942. doi:10.1073/pnas.0904191106
79. Lauer TM, Agrawal NJ, Chennamsetty N, Egodage K, Helk B, Trout BL. Developability index: a rapid in Silico tool for the screening of antibody aggregation propensity. *J Pharm Sci*. 2012;101(1):102-115. doi:10.1002/jps.22758

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Szkodny AC, Lee KH. A systemic approach to identifying sequence frameworks that decrease mAb production in a transient Chinese hamster ovary cell expression system. *Biotechnol. Prog*. 2024;40(5):e3466. doi:10.1002/btpr.3466