Cell Systems

CellPress
OPEN ACCESS

Review

Engineering mammalian protein secretion: Toward the convergence of high-throughput biology and computational methods

Jacopo Gabrielli,[1,2] Nathan E. Lewis,[3,4] Cleo Kontoravdi,[1,2,5,*] and Francesca Ceroni[1,2,5,*]
[1]Department of Chemical Engineering, Imperial College London, London, UK
[2]Imperial College Centre for Engineering Biology, Imperial College London, London, UK
[3]Complex Carbohydrate Research Center, Center for Molecular Medicine, & Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, USA
[4]Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark
[5]Bezos Centre for Sustainable Protein, London, UK
*Correspondence: cleo.kontoravdi98@imperial.ac.uk (C.K.), f.ceroni@imperial.ac.uk (F.C.)
https://doi.org/10.1016/j.cels.2025.101424

SUMMARY

Protein secretion in mammalian cells is the active transport of proteins from the cytoplasm to the extracellular space. It plays a fundamental role in mammalian physiology and signaling, as well as biotherapeutics production and cell and gene therapies. The efficacy of protein secretion, however, is impacted by features of the secreted protein itself, and the host-cell machinery that supports each step of the secretion process. High-throughput techniques such as microfluidics, cell display, and cell encapsulation assays for the study and engineering of secreted proteins are transforming biomedical knowledge and our ability to modulate protein secretion. In addition, computational advances, including signal peptide modeling, whole-protein machine learning models, and genome-scale simulations, are opening new pathways for rational design of protein secretion. Here, we highlight recent developments in secretion engineering that are leading to the convergence of high-throughput experimentation and machine learning methods and can help address current challenges in bioproduction and support future efforts in cell and gene therapy while enabling new modalities.

INTRODUCTION

Protein secretion is pivotal to mammalian cell-to-cell communication, immune response, and physiology.[1–3] Consequently, secreted proteins play a fundamental role in human health and can be harnessed for therapeutic applications.[1–3] Heterologous expression of secreted proteins (e.g., antibodies, hormones, and cytokines) represents a growing proportion of the pharmaceutical market.[4–6] Similarly, protein secretion engineering is relevant for the development of novel vaccines, viral vectors, and chimeric antigen receptor T cells (CAR-Ts)[7–11] for cell therapy. Fine-tuning of *in vivo* secreted proteins can balance therapeutic benefit and toxicity in cell therapy,[7,8,12,13] for example, by modulating the immune response elicited by secreting cell therapies.[14–16]

It is important to be able to engineer protein secretion when developing and manufacturing therapeutics. To do so, knowledge of the mechanisms that regulate the synthesis and secretion of proteins is required. Protein secretion is controlled by features of the secreted protein itself, and the expression patterns of the cell machinery that support each step of the secretory pathway within the host. To address these different aspects, many computational and experimental tools have been developed to facilitate the study and engineering of the protein secretion pathway in mammalian cells, such as hamster ovary cells (CHO), human embryonic kidney (HEK), and a variety of types of primary cells.[17–25]

As both computational and experimental methods for secreted protein and host-cell engineering improve, new opportunities arise for their integration to accelerate foundational and applied mammalian cell bioengineering. In this review, we examine recent developments in machine learning and systems biology models for mammalian protein secretion and complementary high-throughput experimental methods. We aim to highlight synergies, limitations, and opportunities for future developments. We will also discuss how the synergies of these approaches will unlock new opportunities for secretion engineering and biological capabilities.

COMPUTATIONAL METHODS FOR PROTEIN AND HOST-CELL ENGINEERING

Computational approaches are integral to the field of protein engineering, facilitating the rational design and optimization of proteins for various applications.[26–29] Traditional methods, such as computational site-directed mutagenesis and molecular modeling, enabled researchers to predict and modify protein structures and functions with increasing precision.[30,31] For

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

instance, computational tools like Rosetta have been pivotal in predicting protein folding and stability and have been used to engineer function as well.[32–35] On the other hand, algorithms for directed evolution have accelerated the identification of beneficial mutations.[36] These methods enhanced our understanding of protein mechanics and streamlined the development of proteins with improved binding affinities, catalytic activities, and overall functionality.[26,27,37,38]

Deep learning has paved the way for improved protein engineering.[39] More recent applications of transformers,[40] tools like the AlphaFold models,[41] the evolutionary scale models (ESM) family,[42] and the derivative inverse folding models like ProteinMPNN,[28] have accelerated advances in protein engineering. Indeed, the field is being revolutionized by transformer-based protein language models and their ability to generate realistic protein sequences for specific functions paired with AlphaFold's accurate predictions of protein structures from amino acid sequences.[29,41,43]

The concurrent surge in computational power and algorithmic sophistication led to innovative strategies that extend beyond traditional aims of protein engineering such as stability, catalytic efficiency, or binding affinity. For instance, the application of deep learning has led to substantial improvements in signal peptide prediction, annotation, and generation[21,44,45] (Figure 1A). Furthermore, novel protein modeling methods allow the identification and removal *in silico* of cryptic transmembrane domains, which prevent correct secretion[22,46] (Figure 1B).

While the sequence of a protein is a significant contributor to protein secretion efficiency, it is not the only one. The host genetic background and the factors contributing to the regulation of secretion within the cell (e.g., metabolism, expression levels of enzymes involved in secretion control and regulation) also play an essential role. Machine learning together with the increasing availability of large omics datasets have enabled researchers to model and computationally simulate the mammalian secretion pathway. Integration of these new approaches to existing models of the mammalian cell's metabolic network enables the understanding and prediction of targeted changes on protein secretion[18,47] (Figure 1C).

Modeling of the protein sequence, its signal peptide, and the expression of the secretion cellular machinery constitute a powerful suite of tools for *in silico* secretion engineering.

Signal peptide prediction and generation with deep learning

The signal peptide is a cleavable sequence generally located at the N terminus of both transmembrane and secreted proteins, needed for the translocation of eukaryotic proteins to the endoplasmic reticulum (ER). Type I membrane proteins employ an N-terminal signal peptide that engages the canonical signal recognition particle (SRP)-Sec61 pathway and is cleaved before the first transmembrane helix acting as an anchor that impedes transfer. Therefore, the computational frameworks surveyed here are, in practice, applicable to this class of proteins for the purpose of engineering their ER translocation, however, not their membrane retention, which is determined by the transmembrane anchor features.[48,49] Signal peptides are co-translationally bound by the cytosolic SRP, then recognized by the ER-bound SRP's receptor (SR). The SRP-SR complex then coordinates translocation via the Sec61 complex into the ER lumen for further post-translational processing. A recent study has highlighted the importance of signal peptides and their biophysical properties in determining the cotransin that mediates ER translocation via Sec61 and how these differences can be exploited therapeutically.[50] Signal peptides are often characterized by a hydrophobic core, a positively charged N terminus and a cleavage site in their C terminus, generally between 10 and 50 amino acids long and conserved across organisms and kingdoms with minor differences.[51]

Due to their large sequence space, high physicochemical property conservation, low sequence similarities,[52] and similarity with small transmembrane domains, signal peptide modeling has been a long-standing bioinformatic problem.[53–56] The generation of novel sequencing datasets over the last decades has improved our knowledge of and ability to adopt computational methods for recognizing signal peptides and their cleavage site. Weight matrices were the first approach to achieve 70%+ accuracy for signal peptide recognition in very small datasets (<1,000 data points).[57] While artificial neural networks (ANNs) utilizing competitive learning failed to show substantial improvements in signal peptide detection,[58] ANNs trained with supervision and backpropagation with gradient descent resulted in the next step change in the field.[59] Further contributions were made by integrating hidden Markov models (HMMs) to the existing ANN-based methodologies[60–62] and improving weight matrices with frequency correction.[63] Other methods not relying on ANNs were developed exploiting multiple layers of classifiers with voting systems.[64,65] In 2011, SignalP 4.0, trained on protein amino acid sequences labeled by secretion system and host, established the advantage of ANNs across Gram-positive and Gram-negative bacteria, and eukaryotes by distinguishing signal peptides for the respective organisms.[66]

DeepLoc, followed by DeepSig and SignalP 5.0, implemented deep neural networks to the problem of signal peptide prediction and subcellular localization for the first time, achieving state-of-the-art (SOTA) accuracies.[21,67,68] The implementation of protein language models based on transformers has led to the next improvement in prediction accuracy. For example, SignalP 6.0 adopts the bidirectional encoder representations from transformers (BERT) model from Protein Transformer (ProtTrans), which is particularly suited to low-data settings. The improvements brought by this architecture allowed SignalP 6.0 to expand the predictable categories of signal peptides to Sec signal peptide III (Sec/SPIII) and twin-arginine translocation signal peptide II (Tat/SPII).[21] Other groups achieved similar results using approaches like TSignal, another BERT-based transformer trained on large public datasets.[69] However, many of these models suffer from large class imbalances and require sequence annotation to achieve high accuracy. To solve this, Shen et al. proposed USPNet, which uses a multi-sequence alignment (MSA) transformer to learn information, normally annotated, from the sequence alone and label distribution-aware margin loss to mitigate class imbalances.[70]

While signal peptide recognition is fundamental to studying different secretion mechanisms and their dependence on N-terminal amino acid sequences, generation of novel signal peptides

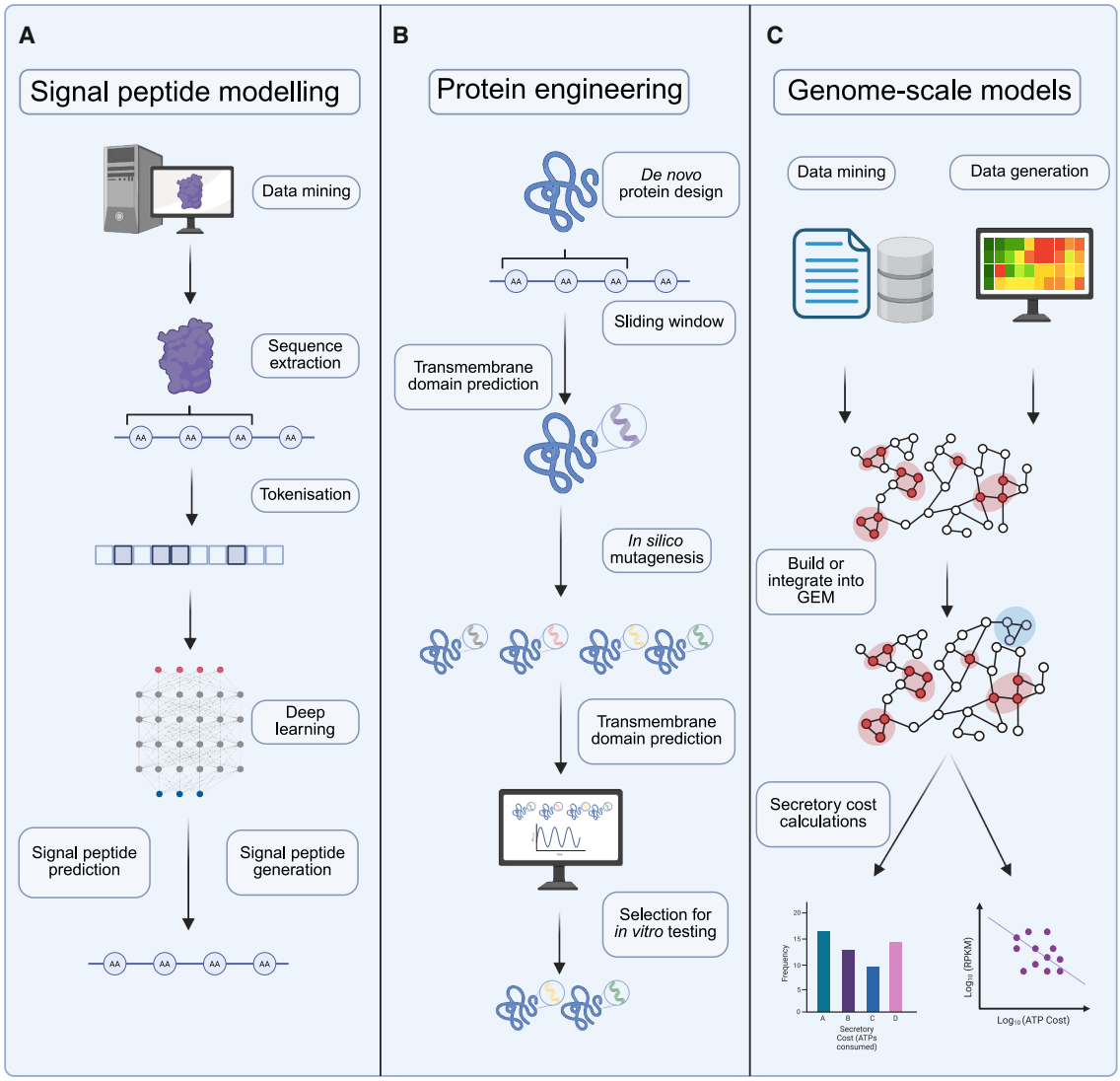

Figure 1. Computational tools facilitate the study and engineering of protein secretion

(A) Machine learning modeling can aid in the prediction and generation of signal peptides. Signal peptide sequences are extracted from secreted protein sequences annotated on public databases like UniProt. The sequences are then tokenized for the training of either predictive or generative deep-learning models. Adapted from Wu et al.[45] and Teufel et al.[21]

(B) Protein engineering methods can guide efforts to improve secretability in mammalian cells by identifying cryptic transmembrane domains. The likelihood of the presence of transmembrane domains is calculated on sliding windows. *In silico* mutagenesis is then applied to mutate and remove the identified domains. Adapted from Wang et al.[22]

(C) Systems biology approaches can be used to identify and model the host-cell machinery in the secretion pathway. This can support identification and mitigation of interactions between the secreted protein and the cellular microenvironment that can impact secretion. Data from the literature are integrated with new information derived from omics approaches to update the parameters of a genome-scale model enabling secretion cost calculations. Adapted from Gutierrez et al.[18] and Masson et al.[47]

can further advance protein engineering efforts and their applications in biotherapeutics and biomanufacturing. The architecture of transformer-based models is particularly suited to generative tasks. As of today, there is only one study, restricted to Gram-positive bacteria, where the application of these architectures to the *in silico* generation of new signal peptides starting from a large public dataset of native signal peptides is reported. Such peptides were tailored to the protein of interest or the expression host. The designs were then tested in the lab and were shown to achieve secretion levels similar or higher to natural secretion peptides even though the model was not trained specifically to optimize secretion levels.[45] Early work harnessing experimental secretion datasets and the perplexity metric characteristic of language models has started implementing a quantitative aspect to the signal peptide generation task. However, the early results were still suboptimal, something that can be ascribed to the quality and heterogeneity of the experimental data the model was evaluated on.[44] Even though substantial improvement has been made on the task of signal peptide prediction, a model for *de novo* generation of signal peptides for

therapeutic proteins expressed in mammalian cells either for bioproduction, cell-, or gene-therapy applications is still missing.

In silico engineering of the protein sequence for improved secretion

While signal peptides are well-defined and localized domains, protein secretion is impacted by variations in other neighboring elements within the protein sequence as well. Some of these elements are also difficult-to-engineer targets due to their variable location and sequence conservation. Examples include tRNA codon choice,[71] aggregation prone regions (APRs),[72] glycosylation sites,[73] and transmembrane domains. Each of these warrant modeling and engineering of the whole protein to understand how changes will affect secretion while maintaining other properties of interest.

Synonymous codon choice adds a dynamic layer of control to secretion. Introducing strings of non-optimal codons ~35 amino acids downstream of the signal sequence slows elongation just long enough for SRP engagement, boosting Sec61 translocation and increasing immunoglobulin G (IgG) secretion up to 4-fold in CHO cells.[71] Which codons are "slow" is host dependent, because the anticodon pool differs between tissues and remodels under stress. Yeast, for example, rewires its tRNA repertoires within minutes to favor stress-response transcripts, demonstrating that tRNA supply is a dynamic regulator of translation and secretion.[74] In higher eukaryotes, proliferating versus differentiated cells deploy mutually exclusive tRNA programs that mirror their codon biases, meaning that a coding sequence tuned for HEK293 might translate sub optimally in other cell types or under different conditions.[75] Current codon optimization algorithms are incorporating progressively more context as they evolved from simple maximizers of the codon adaption index to taking into account organism-specific information, mRNA abundance, and more.[76–78] In the future, codon optimization algorithms for *in silico* design of secreted proteins' coding sequence should incorporate local sequence dependencies and phenotypic ones.

APRs are short stretches of mostly hydrophobic amino acids with low net charge, which form highly stable intermolecular β sheet structures. These are usually solvent exposed as, when they are buried within globular proteins, they do not exhibit aggregation tendencies. Therefore, they are a tough hurdle when present in the Fab or even in the complementarity-determining region (CDR) of an antibody as they can form inclusion bodies within the ER and prevent or diminish the efficiency of protein secretion.[72,79–81] Recently, a study used the Solubis method,[82] which merges the TANGO algorithm to predict APRs[83] and the FoldX force field to conserve stability of the native structure,[84] to engineer a monoclonal antibody (mAb) *in silico*. A single amino acid mutation in the APR located in the light chain led to a reduction in inclusion bodies within the CHO host cells' ER lumen. This reduction correlated with an improvement in production and complexity of the glycosylation profile with unchanged cell viability suggesting improved secretion.[85] Finally, the SOTA model for *in silico* solubility engineering is Aggrescan4D, which added pH-dependent analysis to the parent model Aggrescan3D.[86] Aggrescan4D beats Solubis, the aforementioned model, on computational benchmarks; however, this still needs extensive experimental validation.[87]

Not only the protein context but also a wide range of post-translational modifications can affect protein secretion efficiency. These include glycosylation,[88] phosphorylation,[89] and gamma carboxylation.[90] Glycosylation is of particular interest due to its key role in secretion regulation and its impact on the functionality and activity of several proteins of therapeutic interest. Protein glycosylation is the process of addition of glycans to proteins often during processing in the ER.[73] Protein glycosylation and secretion are dynamically interlinked. For example, protein secretion of some therapeutically relevant proteins is abrogated by tunicamycin,[91] which is an inhibitor of key glycosylation enzymes in mammalian cells. While inhibition of glycosylation enzymes prevents secretion of therapeutic proteins, the addition of glycosylation sites can be used to upregulate it.[92] Glycosylation engineering influences not only secretion levels but also protein function and protein serum half-life. For example, hyper-glycosylation of human erythropoietin, up to five N-linked glycosylation sites versus the natural three, leads to 3-fold increase of erythropoietin serum half-life.[93] Thus, glycosylation not only plays a crucial role in promoting secretion but is also influenced by secretion dynamics itself. In fact, models of secretion in CHO showed that high levels of secretion cause limited ER residence time and affect the complexity of glycosylation due to availability constraints of glycosylation enzymes.[94] Controlling glycosylation levels and complexity is thus of interest for the modulation of secretion.

Recent advances in computational glycoengineering that could aid secretion engineering include CHOGlycoNET,[95] a comprehensive glycosylation reaction network in CHO to accelerate model development, and an ANN model able to predict site-specific glycoform distributions with an average absolute error of 1.1%.[96] Finally, while glycoengineering is often carried out at the cellular level, a recent extensive study established the relationship between primary protein structure and glycosylation beyond sequons.[97] Their "glycoimpact" metric could be used in the future to merge glycoengineering and protein engineering for improved secretion.

Transmembrane domains, stretches of alpha-helical hydrophobic residues, in the signal peptide are required for conventional protein secretion in eukaryotes to trigger ER translocation. These are later cleaved, and the protein is released for further processing in the ER, followed by secretion.[66] Consequently, transmembrane domains that do not get cleaved can impair secretion of the target protein by preventing its release from the membrane. While earlier work focused on computationally identifying transmembrane domains, a recent study used deep-learning models to model proteins *in silico* and remove cryptic transmembrane domains to improve protein secretion in mammalian cells.[22,98] Their degreaser protocol takes a sliding window of 19 amino acids and calculates $\Delta G_{ins;pred}$ with the von Heijne model for transmembrane-helix recognition.[99] Based on a threshold they empirically determined by applying the model to transmembrane and non-transmembrane proteins, they find the cryptic transmembrane domains and perturb them using Rosetta until a stable structure is obtained without them. The model was also experimentally validated by significantly improving the secretion of stable nanoparticles in mammalian cells.[22,98]

Protein secretion is shaped by multiple sequence elements beyond signal peptides, including APRs, glycosylation sites,

Table 1. Overview of the experimental methods examined in this review and adopted for secretion engineering

	Methodology		
Objective	Well-based assays	Cell-display	Encapsulation-based methods
Improving secretion for productivity	Yu et al.,[111] Pinto et al.,[112] and Diep et al.[113]	Dyson et al.[25] and Aebischer-Gumy et al.[114]	Rutkauskaite et al.[115] and Gérard et al.[116]
Screening secreted protein sequence libraries	Gaa et al.,[24] Winters et al.,[117] Jorgolli et al.,[118] and Zhang et al.[119]	Dyson et al.,[25]Javanmardi et al.,[120] Crook et al.,[121] Gaa et al.,[122] and Waldmeier et al.[123]	Fischer et al.,[108] Akbari and Pirbodaghi,[124] and Shembekar et al.[125]
Identifying host factors contributing to secretion	George and Wang,[126] Chen et al.,[127] Ma et al.,[128] and Norman et al.[129]	N/A	Cheng et al.,[20] Yu et al.,[111] Udani et al.,[130] and Wu et al.[131]

For each of the selected methods, three principal use cases are reported accompanied by relevant publications. The selected methods include optofluidic-based ones, cell-surface display, and nanovials and other microgels, which allow greater multiplexing capabilities.

and transmembrane domains. Advances in computational tools now enable targeted engineering to mitigate secretion bottlenecks, from reducing aggregation to optimizing glycosylation and removing cryptic transmembrane domains. In the future, integration of these tools into protein engineering workflows might improve secretion engineering through synergistic modification of multiple protein domains.

Modeling context effects in the secreting host

Protein secretion efficiency is not only determined by the protein sequence or its post-translational modifications but is extensively impacted by the expression host, its secretome, and the growing conditions.[17,80,100] Considering the host context is therefore vital for a holistic approach in secretion engineering in mammalian cells. Substantial computational work has gone into media optimization and host optimization, either assisted by genome-scale models or machine learning.[101–103] However, most of the work in this area has focused on generally increasing titers rather than specifically boosting protein secretion. For instance, the first genome-scale models did not incorporate the secretion pathway.[104] Recent progress in omics techniques has provided the data to model the secretion pathway of several mammalian cell lines, including CHO cell, human, and mouse systems. This allowed model-guided secretion engineering of bioproduction hosts, though quantitatively accurate modeling still represents a hurdle.[18]

The reconstruction of the mammalian secretory pathway and integration to genome-scale models started from a secretory pathway model from yeast.[105] The model was then supplemented with information of the mammalian regulatory pathway lacking in yeasts (e.g., calnexin/calreticulin cycle) and information from public databases and literature. For the mouse and human versions of the model, orthologs were found via homology-based search. A protein-specific information matrix (PSIM) was then created using UniProt information to estimate its secretion cost in ATP equivalents. The ATP equivalents calculation is performed using protein length and ATP costs of each reaction in the twelve-component system modeling the pathway from literature.[18] The reconstruction of the mammalian secretory pathway was also included in the user-friendly computational toolbox CellFie in its SecCellFie expansion.[106]

Following on from this, Kol et al. used their genome-scale models with the protein secretion pathway in their effort to engineer the cells to improve secretion of therapeutic proteins. Specifically, the authors identified proteins whose secretion constituted a burden on the cell as a function of secretion cost and expression levels. They filtered our essential genes and then proceeded to knock out 14 "expensive" and/or problematic host-cell proteins. This resulted in a significant improvement in secretion of their protein of interest in a relevant host. Furthermore, the knockouts increased ease of purification as host-cell proteins constitute important contaminants during downstream processing. Nonetheless, boosting the secretory pathway drains resources, they in fact predicted a Pareto frontier for specific productivity and growth.[17] To maintain production levels at the Pareto frontier, a recent study employed real-time feedback on unfolded protein response (UPR) reporters increasing secreted yield by 70% without further genome edits.[107]

Recent work on genome-scale secretion modeling has been successfully translated into knockout experiments improving yields and simplifying downstream processing of antibody-producing CHO cells. However, while genome-scale metabolic models (GEMs) serve as a useful guide to secretion optimization via genome engineering, they still struggle to accurately predict the quantitative outcome of perturbations. Future improvements in genome annotation and omics datasets from perturbation studies could bridge this gap between useful guide and "digital twin."

INCREASING EXPERIMENTAL THROUGHPUT IN SECRETION ENGINEERING

Advancing experimental throughput is essential for the systematic engineering of protein secretion in mammalian cells (Table 1). High-throughput experimental techniques enable the efficient screening of vast libraries of protein variants, facilitating the identification of those with desired attributes.[19,108–110] Consequently, they can accelerate the discovery of protein variants with optimal secretion parameters for a variety of use-cases (Figure 2).

Miniaturization of well-based assays via microfluidics has allowed the scaling of throughput of traditional methods while maintaining control over the cellular environment. These platforms allow researchers to manipulate variables such as nutrient supply, temperature, or chemical inducers with high accuracy, ensuring consistent and reproducible conditions. These

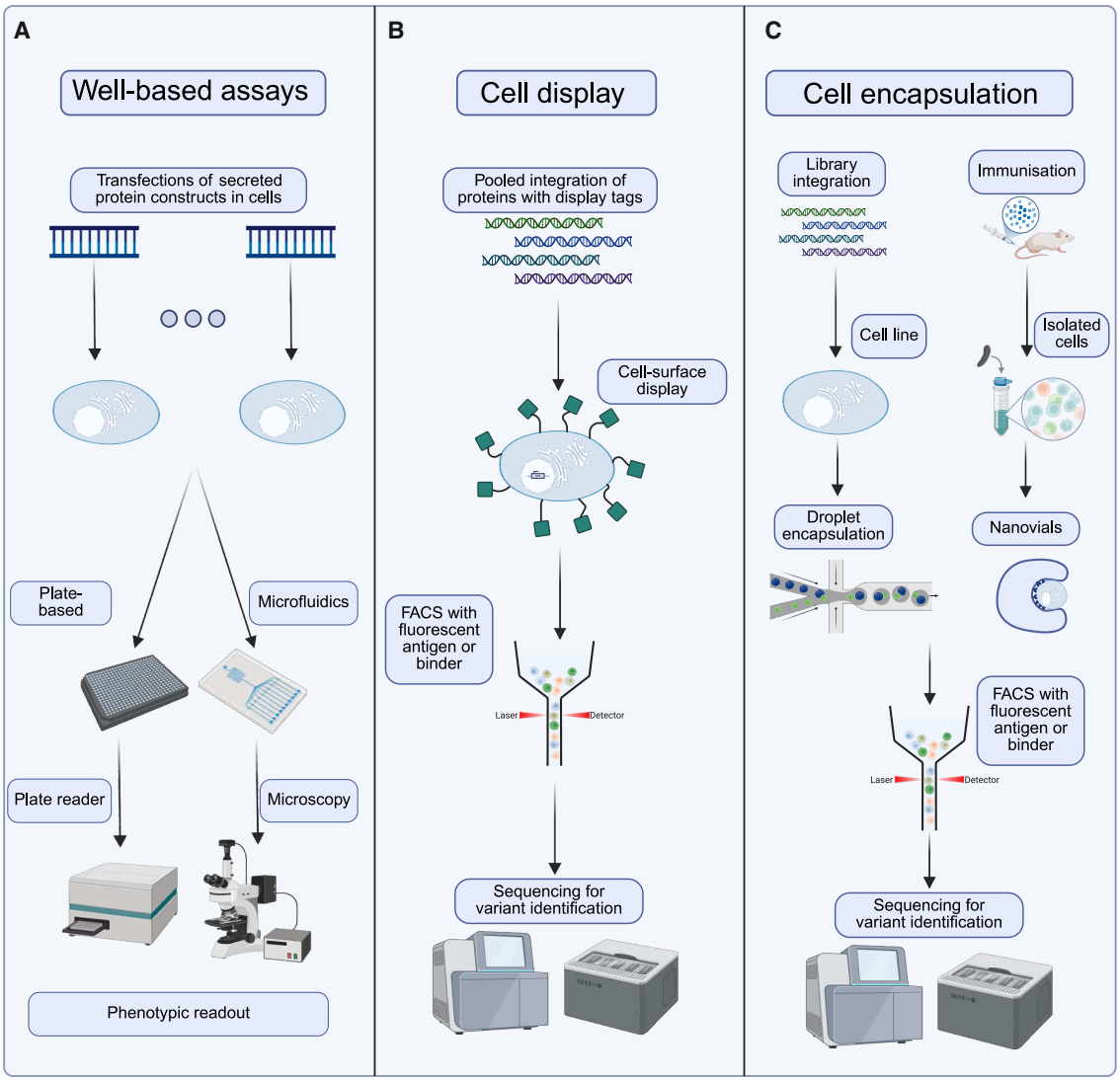

Figure 2. Experimental methods to scale out secretion engineering

(A) Examples of well-based assays for secreted proteins. DNA constructs are transfected into cells in either multi-well plates or microwells via microfluidic devices. To quantify secreted protein abundance or function, plate reader or high-throughput imaging setups can be used. Adapted from O'Neill et al.[132] and Yu et al.[23]

(B) Library variants for a protein of interest can be screened using cell display followed by FACS and sequencing in high throughput. Adapted from Gaa et al.[24] and Dyson et al.[25]

(C) Libraries of protein variants are first designed and integrated or retried from mouse immunization. Isolated cells are then encapsulated in droplets with custom microfluidics or pre-prepared microgels. The encapsulated cells are screened via FACS, and the sequences of secreted protein variants are retrieved by sequencing. Adapted from de Rutte et al.[19] and Cheng et al.[20]

experiments can either be carried out on a microwell plate reader or via microscopy and automated image analysis.[23,111,132] Additionally, miniaturization of experiments leads to a reduced experimental cost per condition tested due to lower reagent costs. This level of control is advantageous to dissect the complex interactions that govern protein secretion where the cellular and environmental context can considerably influence experiments (Figure 2A).

Single-cell assays further enhance experimental throughput by providing detailed insights into the variability of secretion levels among individual cells, unlike bulk assays, which average the responses of entire cell populations. This is crucial for the identification and optimization of protein variants that may be overlooked in population-averaged measurements. In cell-display systems, proteins of interest are genetically fused to a cell-surface anchor, enabling their detection and analysis directly on individual cells via techniques like fluorescence-activated cell sorting (FACS)[24,25,120] (Figure 2B). However, modification of the protein of interest affects cell secretion and biases the single-cell measurements with variants better displayed on the cell surface.[25]

On the contrary, cell encapsulation enables single-cell analysis of secretion with native or unmodified proteins. Techniques like droplet-based microfluidics and microgel encapsulation

enables high-throughput screening of 10^3–10^6 protein variants in parallel. Similar to cell display, cell encapsulation supports the integration of high-throughput analytical processes, such as FACS followed by next-generation sequencing. Therefore, cell encapsulation enables improved capabilities to identify and isolate protein variants from a library or specific cells within a population with optimal secretion profiles[19,20,124] (Figure 2C).

Miniaturizing and scaling up well-based assays

The amount of protein secreted by a given secreting host can be quantified by measuring the quantity of the protein of interest that is present in the growth medium. Depending on the protein produced, methods to detect protein abundance in the medium can vary. Fluorescence detection is adopted for fluorescent reporters, chromogenic assays for specific enzymes, and antibody-based detection when the previous approaches are not applicable but an antibody is available. Customarily, these methods are carried out in well plates usually in 96, 384, or in some cases 1,536 wells.[45,132] To increase throughput beyond 1,536 wells in a standard plate format, microfluidics using custom chips are required.

Microfluidic technologies have considerably advanced secretion analysis by enabling the confinement of individual cells within uniform compartments, thereby preventing crosstalk and allowing precise measurement of secreted proteins. Microengraving, which used lithography to sediment and assay thousands of cells, has been proposed as one of the first approaches to quantify secretion from individual cells. Immunospot array assays scaled the traditional ELISPOT to chips and made the cells recoverable.[133] Single-cell barcode chips (SCBCs) offer the ability to analyze thousands of cells multiplexing measurements and recovering the secreting cells.[128] These methods have reached exquisite levels of sensitivity, throughput, and multiplexing. In a recent study, a chip with >200,000 wells was used to profile seroconversion in blood of COVID-19 patients as early as day-1 after a positive real-time PCR result.[129]

High-throughput microfluidic methods often come with the drawback of laborious cell recovery. This is usually performed either by hand, well-by-well processing, adopting liquid handlers, or transferring samples to a separate instrument for analysis (e.g., a fluorescence-activated cell sorter). As of today, optoelectronic tweezers (OETs) represent the most efficient class of methods used to sort cells on a chip. These use the photoelectric effect to sort individual cells on a chip via high-intensity light beams creating patterned virtual electrodes on a photoconductive material.[134] OETs allow clonal selection of producer CHO cells for bioproduction or secreting B cells for antibody discovery.[117,118] Moreover, they are starting to be employed in cell therapy where they allow short term cell culturing, detection of antigen binding, as well as cytokine secretion and cell isolation on the same device.[135] Commercial implementations of these technologies exist where 10,000 chambers per chip are used. However, due to their associated costs or requirements of microfluidic expertise, these methods are still out of reach for most labs.[118]

Cell display for single-cell secretion analysis

Cell display has emerged as a valuable tool for single-cell secretion analysis, offering an alternative to microfluidic techniques with distinct advantages and limitations. Unlike microfluidic methods that often require complex fabrication and specialized equipment to confine cells and prevent crosstalk, cell-display approaches leverage genetic engineering to anchor the protein of interest on the cell surface. This approach facilitates direct detection and quantification of secreted proteins on individual cells using techniques such as FACS. Compared with other display methods such as yeast and phage display, they allow us to interrogate "developability" (i.e., the likelihood for a drug to become a biomanufacturing product) as part of the screen. The fusion of proteins to membrane anchors may however alter their transport across the native secretion pathways or their functionality, potentially introducing biases in the analysis. In contrast, microfluidic encapsulation methods allow for the assessment of unmodified, native proteins but may be less accessible due to the need for specialized expertise.[136,137]

The first important consideration in preparing mammalian cell-display libraries is to pick the membrane tether. Previously tested ones can be split into N-terminal, from smallest to largest, glycosylphosphatidylinositol (GPI) anchor,[138] platelet-derived growth factor receptor transmembrane domain (PDGFR-TM),[138] C-C chemokine receptor type 5 (CCR5),[139] and C-terminal FasL.[121] The carrier selected for introducing the DNA encoding the library in the host cell is also important. Early versions of the method used chemical transformation with poor efficiencies.[140] This can be improved by adopting lentiviruses but with a risk of multiplicity of integration and are limited to a ~10 kb cargo.[141] While transposases are easier to use and allow integration library fragments beyond 10 kb, they also lead to random integration.[123] On the contrary, CRISPR-Cas9 and recombinase-mediated integration allow for the site-specific integration of long cargoes and constitute the current SOTA in the preparation of display libraries.[142,143]

Binding, folding, and stability are the main properties of secreted proteins that have been investigated with mammalian cell display. Binding is assessed by displaying the binder on the surface and detecting binding via a sandwich assay on the antigen.[126,142] The disadvantage to this approach is that some binders need to be secreted to tightly bind their target. An alternative is to display the antigen on the surface and secrete the candidate binders whose binding will still be detected via sandwich assays.[128] One of the caveats to this method is its low signal-to-noise ratio arising from crosstalk between cells in the pooled population. Moreover, the actual affinity is intrinsically confounded by the level of surface expression: variants that fold or traffic poorly are displayed at lower density and therefore appear to bind weakly even when their intrinsic affinity is high.

Multiplexed variant-effect assays (MAVEs) are a family of experimental techniques where many mutations are performed on a protein or genetic sequence, and their effect on a specific phenotype is measured in a multiplexed setting. These have recently been scaled to analyze up to five phenotypes of secreted protein variants in a multiplexed manner exploiting cell-surface display. Popp et al. combined oligo pool-based cloning, saturation mutagenesis, cell-surface display, and next-generation sequencing to track the extracellular abundance and post-translational modifications of thousands of variants of coagulation factor IX. The study yielded dense quantitative maps that link local sequence changes to

secretion efficiency of the target protein.[144] A similar method combining cell-surface display and multiplexed next-generation sequencing is multiple antigens and multiple antibodies (MAGMA-seq). MAGMA-seq is used to display multiple variants of multiple antibodies on the surface of yeast cells. Then, fluorescent antigens are added at different concentrations to infer quantitative biophysical parameters. These two methods exemplify the combination of cell-surface display with next-generation sequencing and ingenuous cell sorting to characterize secreted proteins in high throughput.[109]

Finally, Dyson et al. found a correlation between the developability of known antibodies and the display abundance on the cell's surface assessed by flow cytometry.[25] This suggested that the approach could be used to simultaneously screen for developability and binding. They demonstrated that in a subsequent study in CHO cells expanding the use-cases of cell display.[143] Finally, other recent innovations in mammalian cell display include an alternative splicing system to screen both displayed and secreted product called SPLICELECT,[114] secretion-switchable libraries,[122] and integrations with microfluidics technologies.[24]

Mammalian cell-display technologies offer a simple experimental method for single-cell measurements of secreted proteins. Nonetheless, the modification of the target protein for membrane tethering or the potential crosstalk in post-secretion capture offers avenues for future improvements via cell encapsulation or alternative membrane tethering methods.

Tag-free secretome assays via cell encapsulation

Encapsulation methods provide an alternative approach to single-cell secretion analysis by isolating individual cells within microdroplets or in gel microparticles, creating discrete microenvironments that preserve cellular integrity and prevent crosstalk.[145,146] Unlike cell-display systems, which require genetic modifications to tether secreted proteins to the cell surface, encapsulation does not disrupt the native state of secreted proteins, allowing for unbiased and accurate measurement of secretion profiles. When compared with microfluidic compartmentalization techniques that confine cells to single wells, encapsulation offers enhanced scalability and flexibility. Additionally, encapsulation methods facilitate the capture and retention of secreted proteins within confined spaces, improving the sensitivity and reliability of detection assays.[136,147] However, these methods also present challenges, such as the need for precise control over encapsulation conditions to ensure consistent cell viability and secretion measurements or the integration with downstream analytical processes. The latter can be technically demanding, requiring specialized microfluidic device designs and expertise.

Droplet-based microfluidics for secretion analysis generally entails the encapsulation of single cells in droplets via single-step emulsification. This happens at the interface of a cell sample, capture and detection reagent, aqueous solution, and an oil or surfactant one.[145] The secreted protein is then either captured on the surface of the cells for assaying via immunofluorescence and flow cytometry,[148] detected via a Förster resonance electron transfer (FRET)-based assay through microscopy,[115] or an enzyme-linked immunosorbent assay (ELISA) followed by FACS.[124] After sorting, the genotype of each cell can be recovered either via next-generation DNA sequencing, RNA sequencing, or both.[116]

Recent advances in droplet microfluidics have motivated its adoption in the analysis of mammalian secretion systems. Examples include the characterization of membrane retention of proteins of interest on single cells via droplet encapsulated FRET assays,[115] and co-encapsulation of secreting and target cells to perform functional screening of immunotherapies. The latter led to the discovery of CD40 agonists present at low frequencies in the test library (<0.02%) and CD3-HER2 bispecifics.[149] Moreover, a recent study presented the use of high-throughput microfluidics to accumulate interleukin-2 (IL-2) onto single T cells, while avoiding diffusion, using on-cell cytokine capture before staining and sorting with FACS. Their technique involves capturing the protein of interest on the cell membrane followed by encapsulation in a droplet to avoid spurious capture from neighboring cells. This prevents the experimental noise from crosstalk observed when the same cell-capture technique is used in mammalian cell-display studies, discussed above.[148]

Cell encapsulation in gel microparticles is an emerging technology in the analysis of mammalian cell secretion. Gel microparticles are often made of hydrogels, three-dimensional networks of crosslinked hydrophilic polymer chains, using various biopolymers. They are fabricated via droplet-based microfluidics either with cells encapsulated during fabrication or as hollow structures where cells can be inserted in.[146,150,151] If the cell is inside the gel microparticle, the advantage of microparticles over droplets in emulsion is that they can be exposed to flow shear stress; therefore, cell-surface capture is not required to assay single cells. On the other hand, the advantage of hollow microparticles is that cells can be encapsulated by mixing rather than microfluidics. Therefore, they can be produced, stored empty, and filled at the point and time of use with the downside of potential crosstalk due to the opening in the microparticles.[152]

Gel microparticles were recently adopted to advance secretion screening of mammalian proteins. In a first study by de Rutte et al., "nanovials," crescent-shaped hydrogels, are developed by crosslinking hydrogel with multi-arm polyethylene glycol (PEG) + UV-crosslinked acrylate and gelatin and are functionalized with biotin. Nanovials can house single secreting cells within a pocket where the secreted protein of interest is captured by an antibody pre-coated on the surface of the pocket by strep-biotin interaction.[19] The protein is then detected by fluorescent antibody and FACS or oligo-barcoded antibody and single-cell RNA sequencing (scRNA-seq) and oligo sequencing. In a follow up, the authors proposed secretion-coupled sequencing (SEC-seq), a method that enables high-throughput association of secretion levels and the transcriptome of single mammalian cells by employing oligo-barcoded antibodies followed by scRNA-seq.[9] Because SEC-seq indexes each nanovial with a unique oligonucleotide barcode, a single run can profile secretion and transcripts of specific proteins for $>10^4$ individual cells, providing great multiplexing capacity.[12,20]

A different approach has instead recently been proposed, which uses droplet microfluidics to encapsulate secreting cells in (bioglass) BG-agarose, which can then be functionalized with a SNAP-tagged capture to retain the protein of interest within the hydrogel. Fluorescently labeled antigens can then be

applied to establish binding specificity followed by fluorescent Fc-capture antibodies to establish secretion levels via FACS.[108]

Encapsulation in BG-agarose is compatible with flow-based analysis, high-throughput encapsulation (10^7 cells/h), and multiplexed measurements of phenotypic parameters with no cross-talk, as opposed to open crescent-shaped microgels.[108] However, it requires access to microfluidics devices, contrary to nanovials, which can be purchased as a reagent.

Outlook—Enabling secretion engineering through the convergence of AI and systems biology with high-throughput experimentation

The convergence of recent efforts in computational approaches and high-throughput experimentation is going to accelerate our ability to dissect and engineer mammalian protein secretion. Although the field has seen considerable progress in recent years, challenges remain in optimizing the interplay between protein features and the cellular context for efficient and predictable secretion.

The proliferation of deep-learning methods and protein language models for protein engineering,[29,43] combined with *in silico* tools for rationally modulating protein features such as cryptic transmembrane domains,[22] glycosylation patterns,[96,97] and APRs,[82,87] has already resulted in secretion improvements of several challenging proteins like hard-to-secrete mAbs[85] and *de novo* complex protein nanoparticles.[98] Integration of the secretory pathway in genome-scale models with detailed, mechanistic descriptions of host metabolism is giving rise to frameworks that capture metabolic burdens, regulatory networks, and the holistic effects of secretory demands on the host cell.[18,104,106] These will be fundamental in complementing computational protein engineering to predict the effects of modifications in the right cellular or environmental context.

Novel methods for *in silico* secretion engineering have been scarcely applied to highly specialized secretory cell types that present unique challenges and opportunities, such as the investigation of granule targeting or regulated release. While major focus has been given to bioproduction chassis, we believe that future work should incorporate considerations on a wider range of cell types and specialized functions.

On the experimental front, miniaturization of conventional well-based assays into microfluidic-, well-, or droplet-based workflows offers increasingly precise and scalable platforms for secretion screening.[116,124] Single-cell assays such as droplet- or gel-encapsulation preserve protein secretion in its native form, while more traditional display methods and microfluidics-based compartmentalization provide complementary high-throughput solutions.[108,115,130,149] However, expanding the use of these methods will depend on the further maturation of microfluidics manufacturing and integration with accessible instrumentation for both cell sorting and subsequent molecular analyses like scRNA-seq and proteomics.[145,151]

A promising avenue for accelerating progress lies in the lab-in-the-loop paradigm, in which experimental data continually refine *in silico* models, closing the cycle with directed experimental campaigns. Some early examples of high-throughput lab data feeding back into sophisticated computational models already exist in the recent literature discussed in this review. Notably, the ~160-variant secretion dataset of Wu et al.[45] now serves as an independent dataset for the first predictive model of signal peptide efficiency, SecretoGen.[44] In Kol et al., the multiplexed genome edits were suggested by an updated GEM of CHO cells including the secretion pathway based on the cellular cost of commonly secreted proteins.[17]

In the future, we expect further integration of high-throughput experimentation with advanced computational models. For example, CRISPR-based functional screening with microfluidics[111] could synergize with genome-scale models[18] to dissect bottlenecks in secretion (e.g., identification of host factors by multiplex perturbation and then iterative investigation of their interactions) (Figure 3A). Machine learning-based signal peptide generation[44,45] with droplet encapsulation systems[108,130] could also be adopted for *de novo* peptide validation (Figure 3B). Simultaneous functional and secretory output screening via microfluidics could synergize with the "degreaser" pipeline[22] and protein deep-learning models[43,98] "in the loop" to optimize protein nanoparticle vaccines (Figure 3C). While these examples illustrate the substantial potential of synergistic, data-driven pipelines, challenges persist around standardization of quantitative readouts, scalability to a range of protein classes, and reliable interpretation of single-cell measurements of secretory output.

Furthermore, recent synthetic-biology approaches now enable post-translational, protease-gated export of pre-synthesized cargo from the ER. In these systems, canonical ER-retention motifs (e.g., KDEL or KKXX) are appended to the C terminus of the target protein and flanked by an orthogonal protease-cleavage site; activity of a genetically engineered cognate protease, delivered to the cells and chemically induced, mediates motif excision and triggers passage of the folded cargo through the Golgi to the plasma membrane. Three complementary implementations illustrate the breadth of the approach: a tobacco etch virus (TEV) protease-based design that couples cargo release to programmable protease logic, thereby coordinating intercellular signaling networks[154]; a chemically inducible variant in which a split protease reconstitutes upon small-molecule addition, enabling rapid, dose-dependent secretion pulses[155]; and a modular architecture that layers orthogonal proteases to multiplex secretion of distinct cargoes in mammalian cells.[12] These tools could synergize with high-throughput experiments and modeling frameworks surveyed in this review to illuminate rate-limiting steps within the ER-Golgi-vesicle continuum and to inform model-guided secretion engineering.

CHALLENGES AND OPPORTUNITIES

Four obstacles still limit progress in the field:

(1) data standardization: secretion datasets vary in units and metadata; common ontologies and a central repository would make models portable.
(2) Quantitative single-cell readouts: droplet and gel assays[106,128] need calibration standards to translate fluorescence into absolute flux.
(3) Context dependence: rules learned in CHO often fail in HEK or primary cells; host-swap benchmark panels are required.

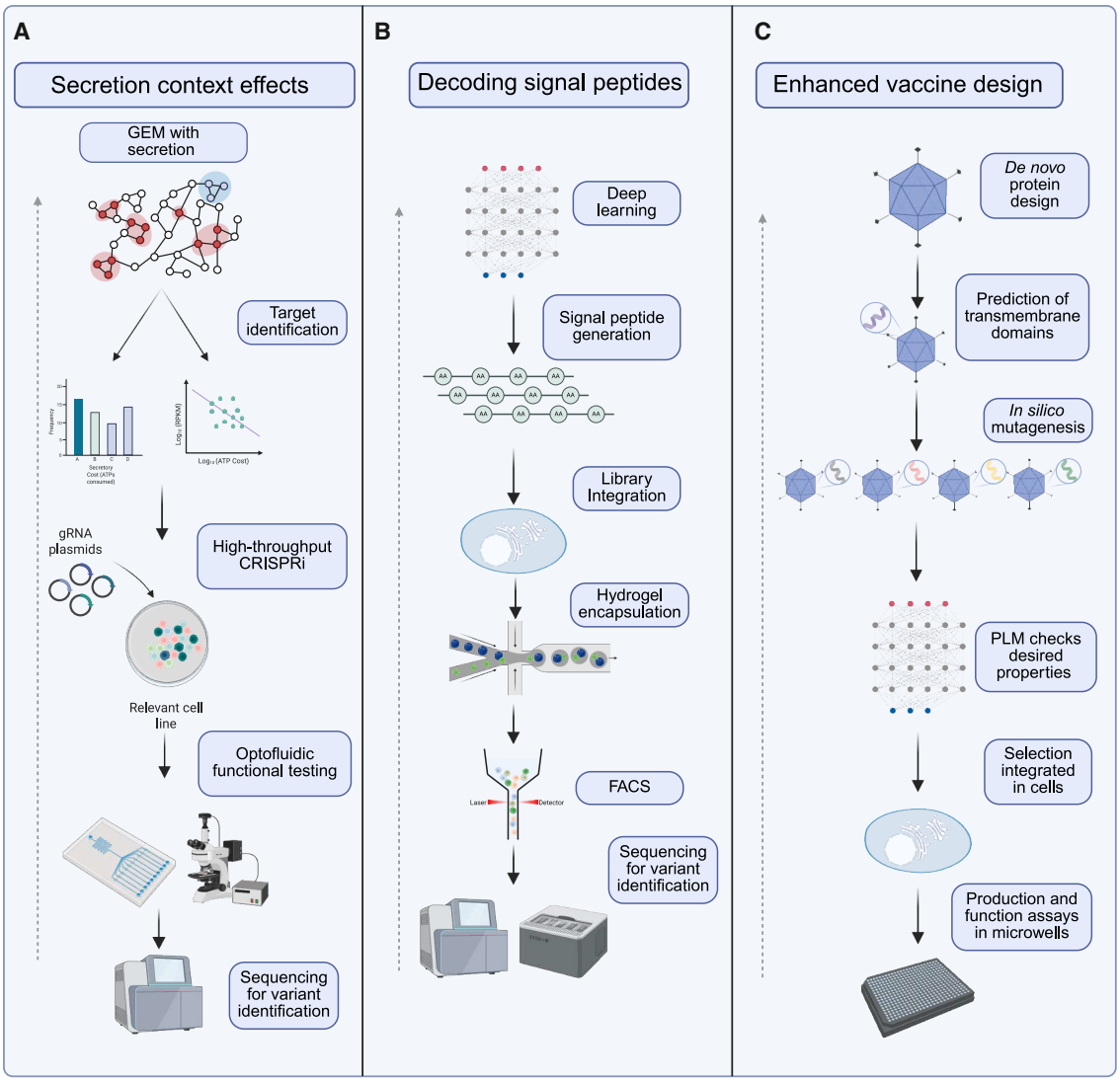

Figure 3. Computational and high-throughput experimental methods can be coupled to accelerate the study and engineering of mammalian protein secretion

(A) Genome-scale models that incorporate the secretion pathway can be used to identify genetic targets for cell engineering experiments. These can be performed in high throughput on cells within optofluidic devices to decode context effects dependent on the secretion host. Adapted from Gutierrez et al.[18] and Yu et al.[111]

(B) Generative deep-learning models, such as transformer-based protein language models, can be used to generate signal peptides to enhance secretion. The *de novo* peptides can be tested in high-throughput single-cell assays performed on hydrogel-encapsulated cells. Adapted from Teufel et al.[44] and Fischer et al.[108]

(C) *De novo* protein design can be used to perform *in silico* mutagenesis guided by the degreaser pipeline to remove transmembrane domains. The *in silico* mutagenesis is followed by heterologous production and functional assays in microwells to simultaneously optimize secretion output and function of *de novo*-designed protein nanoparticle vaccines. Adapted from Wang et al.[153] and Hendricks et al.[98] Dashed arrows indicate the potential for iterative improvement of these methods by feeding data back into the GEMs, deep-learning signal peptide generation, or the *de novo* protein design models.

(4) Multi-objective optimization: secretion, glycosylation, and growth can be antagonistic; Pareto-front design algorithms integrated with GEMs could offer a solution.

Addressing these will aid the generalization of learnings from secretion engineering studies and to scale them in biomanufacturing-relevant settings.

Looking ahead, bridging computational and experimental realms more tightly will demand the generation of larger and more standardized datasets from diverse cell lines and libraries of different classes of secreted protein (Table 2). To maximize the value of forthcoming high-throughput secretion screening approaches, we also suggest the creation of a dedicated, FAIR-compliant public repository that enforces a common schema for sequence, quantitative secretion readouts, and assay metadata. It will be essential to expand beyond the current focus on mAbs, cytokines, and hormones to obtain generalizable insights. Even partial progress in this space would help mitigate current issues of data heterogeneity and context dependences. In parallel, a stronger mechanistic understanding of the

Table 2. Roadmap for secretion engineering enabled by the convergence of computational and high throughput methods

Horizon	Modeling target	Matching experimental technique	Why is it tractable	Key references
Short-term (0–5 years)	context-aware signal-peptide generators	nanovial/FACS screens of 10^5 signal peptides	SecretoGen can propose sequences; nanovials verify secretion in <48 h	Teufel et al.[44] and Ma et al.[128]
	GEM-guided multiplexed knockout (KO) for host cell protein (HCP) secretion burden	CRISPR knockouts plus pooled secretion assay	reaction-cost updates already fed back into CHO GEMs	Kol et al.,[17] Gutierrez et al.,[18] and Praznik et al.[155]
Mid-term (2–5 years)	SRP and codon-optimized signal peptide generation	high-throughput screens with parallel SRP and secretion reporter	SRP reporter recently established as well as parallel secretion and display techniques	Mansouri et al.,[12] Wenzell et al.,[50] and Zhang et al.[119]
	quantitative ER-stress cost models	reporter-guided multiplexed screens, GEM extension	UPR reporters combined with pooled secretion screening plus GEM extension	Gutierrez et al.,[18] Parthiban et al.,[142] and Ranganathan et al.[156]
	multi-objective Pareto design (titer plus glyco plus growth)	continuous processing, reinforcement learning, or Bayesian optimization-powered feedback	live controllers already piloted for secretion stress and other bioprocess variables	Elliott et al.,[93] Feizi et al.,[105] and Ranganathan et al.[156]
Long-term (5–10 years)	closed-loop biofoundry for secreted proteins	multiplexed secretion assays in bioreactors with multiple readouts	integration of bioreactors with multiplexed secretion assays still needs to be performed; reinforcement learning for bioreactors exists but not applied to secreted protein or secreting host design	Cheng et al.[20] and Feizi et al.[105]

ER-Golgi-vesicle continuum, including unconventional secretion routes and quality control checkpoints, could significantly expand the scope of machine learning models. As these complementary insights accumulate, truly comprehensive secretome engineering strategies will emerge, enabling improved therapeutics with finely tuned pharmacokinetics, on-demand secretion of immunomodulatory agents in cell and gene therapies, and more predictable biomanufacturing processes.

ACKNOWLEDGMENTS

The authors acknowledge the support of the EPSRC Centre for Doctoral Training in BioDesign Engineering (EP/S022856/1) (to J.G., C.K., and F.C.). F.C. and C.K. were also partly funded by the Bezos Earth Fund through the Bezos Centre for Sustainable Protein (BCSP/IC/001); the UK National Alternative Proteins Innovation Centre (NAPIC), which is an Innovation and Knowledge Centre funded by the Biotechnology and Biological Sciences Research Council (BBSRC); and the Innovate UK (BB/Z516119/1). F.C. is also supported in part by the Engineering and Physical Sciences Research Council under the EE-Bio Programme Grant (EP/Y014073/1). N.E.L. was supported in part by the National Institutes of Health (R35 GM119850), the Novo Nordisk Foundation (NNF20SA0066621), and the Georgia Research Alliance.

The graphical abstract was created in BioRender. Gabrielli, J. (2025) https://BioRender.com/3bk15ti. Figure 1 was created in BioRender. Gabrielli, J. (2025) https://BioRender.com/pqp4u3u. Figure 2 was created in BioRender. Gabrielli, J. (2025) https://BioRender.com/sqoyilf. Figure 3 was created in BioRender. Gabrielli, J. (2025) https://BioRender.com/wflyi4g.

AUTHOR CONTRIBUTIONS

All authors conceived and wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT to refine readability and check the grammar and consistency of tone and vocabulary. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

1. Ferro-Novick, S., and Brose, N. (2013). Nobel 2013 Physiology or medicine: Traffic control system within cells. Nature *504*, 98. https://doi.org/10.1038/504098a.

2. Benham, A.M. (2012). Protein Secretion and the Endoplasmic Reticulum. Cold Spring Harb. Perspect. Biol. *4*, a012872. https://doi.org/10.1101/cshperspect.a012872.

3. Rabouille, C. (2017). Pathways of Unconventional Protein Secretion. Trends Cell Biol. *27*, 230–240. https://doi.org/10.1016/j.tcb.2016.11.007.

4. Kinch, M.S., Kraft, Z., and Schwartz, T. (2023). Monoclonal antibodies: Trends in therapeutic success and commercial focus. Drug Discov. Today *28*, 103415. https://doi.org/10.1016/j.drudis.2022.103415.

5. Baldo, B.A., and Pham, N.H. (2021). Biologics: Monoclonal Antibodies for Non-cancer Therapy, Cytokines, Fusion Proteins, Enzymes, and Hormones. In Drug Allergy: Clinical Aspects, Diagnosis, Mechanisms, Structure-Activity Relationships, B.A. Baldo and N.H. Pham, eds. (Springer International Publishing), pp. 533–593. https://doi.org/10.1007/978-3-030-51740-3_13.

6. McGonigle, P. (2025). How Biologics Have Changed the Drug Discovery Landscape. Annu. Rev. Pharmacol. Toxicol. *65*, 29–46. https://doi.org/10.1146/annurev-pharmtox-061724-080811.

7. Mo, C., Li, X., Wu, Q., Fan, Y., Liu, D., Zhu, Y., Yang, Y., Liao, X., Zhou, Z., Zhou, L., et al. (2023). SARS-CoV-2 mRNA vaccine requires signal peptide to induce antibody responses. Vaccine *41*, 6863–6869. https://doi.org/10.1016/j.vaccine.2023.09.059.

8. Hou, F., Zhang, Y., Liu, X., Murad, Y.M., Xu, J., Yu, Z., Hua, X., Song, Y., Ding, J., Huang, H., et al. (2023). mRNA vaccines encoding fusion proteins of monkeypox virus antigens protect mice from vaccinia virus challenge. Nat. Commun. *14*, 5925. https://doi.org/10.1038/s41467-023-41628-5.

9. Jaspers, J.E., Khan, J.F., Godfrey, W.D., Lopez, A.V., Ciampricotti, M., Rudin, C.M., and Brentjens, R.J. (2023). IL-18–secreting CAR T cells targeting DLL3 are highly effective in small cell lung cancer models. J. Clin. Invest. *133*, e166028. https://doi.org/10.1172/JCI166028.

10. Díez-Alonso, L., Falgas, A., Arroyo-Ródenas, J., Romencín, P.A., Martínez, A., Gómez-Rosel, M., Blanco, B., Jiménez-Reinoso, A., Mayado, A., Pérez-Pons, A., et al. (2024). Engineered T cells secreting anti-BCMA T cell engagers control multiple myeloma and promote immune memory in vivo. Sci. Transl. Med. *16*, eadg7962. https://doi.org/10.1126/scitranslmed.adg7962.

11. Philippidis, A. (2023). BioMarin's ROCTAVIAN Wins Food and Drug Administration Approval As First Gene Therapy for Severe Hemophilia A. Hum. Gene Ther. *34*, 665–668. https://doi.org/10.1089/hum.2023.29251.bfs.

12. Mansouri, M., Ray, P.G., Franko, N., Xue, S., and Fussenegger, M. (2023). Design of programmable post-translational switch control platform for on-demand protein secretion in mammalian cells. Nucleic Acids Res. *51*, e1. https://doi.org/10.1093/nar/gkac916.

13. Park, H.-J., Bang, Y.-J., Kwon, S.P., Kwak, W., Park, S.-I., Roh, G., Bae, S.-H., Kim, J.-Y., Kwak, H.W., Kim, Y., et al. (2023). Analyzing immune responses to varied mRNA and protein vaccine sequences. npj Vaccines *8*, 84. https://doi.org/10.1038/s41541-023-00684-0.

14. Christodoulou, I., Ho, W.J., Marple, A., Ravich, J.W., Tam, A., Rahnama, R., Fearnow, A., Rietberg, C., Yanik, S., Solomou, E.E., et al. (2021). Engineering CAR-NK cells to secrete IL-15 sustains their anti-AML functionality but is associated with systemic toxicities. J. Immunother. Cancer *9*, e003894. https://doi.org/10.1136/jitc-2021-003894.

15. Amatya, P.N., Xiang, J., O'Neal, J., Ritchey, J.K., Carter, A.J., Cooper, M. L., and DiPersio, J. (2023). Mechanistic Studies of Cytokine Release Syndrome (CRS) with Roles of Interferon-Gamma (IFN-g) and Tumor Necrosis Factor Alpha (TNF-α) While Maintaining CAR-T Function *in Vitro*. Blood *142*, 2086. https://doi.org/10.1182/blood-2023-187449.

16. Vlahos, A.E., Call, C.C., Kadaba, S.E., Guo, S., and Gao, X.J. (2023). Compact Programmable Control of Protein Secretion in Mammalian Cells. Preprint at bioRxiv. https://doi.org/10.1101/2023.10.04.560774.

17. Kol, S., Ley, D., Wulff, T., Decker, M., Arnsdorf, J., Schoffelen, S., Hansen, A.H., Jensen, T.L., Gutierrez, J.M., Chiang, A.W.T., et al. (2020). Multiplex secretome engineering enhances recombinant protein production and purity. Nat. Commun. *11*, 1908. https://doi.org/10.1038/s41467-020-15866-w.

18. Gutierrez, J.M., Feizi, A., Li, S., Kallehauge, T.B., Hefzi, H., Grav, L.M., Ley, D., Baycin Hizal, D., Betenbaugh, M.J., Voldborg, B., et al. (2020). Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. Nat. Commun. *11*, 68. https://doi.org/10.1038/s41467-019-13867-y.

19. de Rutte, J., Dimatteo, R., Archang, M.M., van Zee, M., Koo, D., Lee, S., Sharrow, A.C., Krohl, P.J., Mellody, M., Zhu, S., et al. (2022). Suspendable Hydrogel Nanovials for Massively Parallel Single-Cell Functional Analysis and Sorting. ACS Nano *16*, 7242–7257. https://doi.org/10.1021/acsnano.1c11420.

20. Cheng, R.Y.-H., de Rutte, J., Ito, C.E.K., Ott, A.R., Bosler, L., Kuo, W.-Y., Liang, J., Hall, B.E., Rawlings, D.J., Di Carlo, D., et al. (2023). SEC-seq: association of molecular signatures with antibody secretion in thousands of single human plasma cells. Nat. Commun. *14*, 3567. https://doi.org/10.1038/s41467-023-39367-8.

21. Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gíslason, M.H., Pihl, S.I., Tsirigos, K.D., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat. Biotechnol. *40*, 1023–1025. https://doi.org/10.1038/s41587-021-01156-3.

22. Wang, J.Y.J., Khmelinskaia, A., Sheffler, W., Miranda, M.C., Antanasijevic, A., Borst, A.J., Torres, S.V., Shu, C., Hsia, Y., Nattermann, U., et al. (2023). Improving the secretion of designed protein assemblies through negative design of cryptic transmembrane domains. Proc. Natl. Acad. Sci. USA *120*, e2214556120. https://doi.org/10.1073/pnas.2214556120.

23. Yu, P., Ran, J., Yang, R., Zhu, H., Lu, S., Wu, Y., Zhao, T., and Xiong, T. (2024). Rapid isolation of pan-neutralizing antibodies against Omicron variants from convalescent individuals infected with SARS-CoV-2. Front. Immunol. *15*, 1374913. https://doi.org/10.3389/fimmu.2024.1374913.

24. Gaa, R., Kumari, K., Mayer, H.M., Yanakieva, D., Tsai, S.P., Joshi, S., Guenther, R., and Doerner, A. (2023). An integrated mammalian library approach for optimization and enhanced microfluidics-assisted antibody hit discovery. Artif. Cells Nanomed. Biotechnol. *51*, 74–82. https://doi.org/10.1080/21691401.2023.2173219.

25. Dyson, M.R., Masters, E., Pazeraitis, D., Perera, R.L., Syrjanen, J.L., Surade, S., Thorsteinson, N., Parthiban, K., Jones, P.C., Sattar, M., et al. (2020). Beyond affinity: selection of antibody variants with optimal biophysical properties and reduced immunogenicity from mammalian display libraries. mAbs *12*, 1829335. https://doi.org/10.1080/19420862.2020.1829335.

26. Jiang, L., Althoff, E.A., Clemente, F.R., Doyle, L., Röthlisberger, D., Zanghellini, A., Gallaher, J.L., Betker, J.L., Tanaka, F., Barbas, C.F., et al. (2008). De Novo Computational Design of Retro-Aldol Enzymes. Science *319*, 1387–1391. https://doi.org/10.1126/science.1152692.

27. Tinberg, C.E., Khare, S.D., Dou, J., Doyle, L., Nelson, J.W., Schena, A., Jankowski, W., Kalodimos, C.G., Johnsson, K., Stoddard, B.L., et al. (2013). Computational design of ligand-binding proteins with high affinity and selectivity. Nature *501*, 212–216. https://doi.org/10.1038/nature12443.

28. Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R.J., Milles, L.F., Wicky, B.I.M., Courbet, A., de Haas, R.J., Bethel, N., et al. (2022). Robust deep learning–based protein sequence design using ProteinMPNN. Science *378*, 49–56. https://doi.org/10.1126/science.add2187.

29. Ferruz, N., Schmidt, S., and Höcker, B. (2022). ProtGPT2 is a deep unsupervised language model for protein design. Nat. Commun. *13*, 4348. https://doi.org/10.1038/s41467-022-32007-7.

30. Bloom, J.D., Meyer, M.M., Meinhold, P., Otey, C.R., MacMillan, D., and Arnold, F.H. (2005). Evolving strategies for enzyme engineering. Curr. Opin. Struct. Biol. *15*, 447–452. https://doi.org/10.1016/j.sbi.2005.06.004.

31. Lutz, S. (2010). Beyond directed evolution—semi-rational protein engineering and design. Curr. Opin. Biotechnol. *21*, 734–743. https://doi.org/10.1016/j.copbio.2010.08.011.

32. Rohl, C.A., Strauss, C.E.M., Misura, K.M.S., and Baker, D. (2004). Protein Structure Prediction Using Rosetta. Methods Enzymol. *383*, 66–93. https://doi.org/10.1016/S0076-6879(04)83004-0.

33. Plaxco, K.W., Simons, K.T., and Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins1. J. Mol. Biol. *277*, 985–994. https://doi.org/10.1006/jmbi.1998.1645.

34. Guerois, R., Nielsen, J.E., and Serrano, L. (2002). Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations. J. Mol. Biol. *320*, 369–387. https://doi.org/10.1016/S0022-2836(02)00442-4.

35. Capriotti, E., Fariselli, P., and Casadio, R. (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. *33*, W306–W310. https://doi.org/10.1093/nar/gki375.

36. Romero, P.A., Krause, A., and Arnold, F.H. (2013). Navigating the protein fitness landscape with Gaussian processes. Proc. Natl. Acad. Sci. USA *110*, E193–E201. https://doi.org/10.1073/pnas.1215251110.

37. Lippow, S.M., Wittrup, K.D., and Tidor, B. (2007). Computational design of antibody-affinity improvement beyond in vivo maturation. Nat. Biotechnol. *25*, 1171–1176. https://doi.org/10.1038/nbt1336.

38. Röthlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J.L., Althoff, E.A., Zanghellini, A., Dym, O., et al. (2008). Kemp elimination catalysts by computational enzyme design. Nature *453*, 190–195. https://doi.org/10.1038/nature06879.

39. Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems 25 (NIPS 2012).

40. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2023). Attention is all you need. Preprint at arXiv. https://doi.org/10.48550/arXiv.1706.03762.

41. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

42. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. Science *379*, 1123–1130. https://doi.org/10.1126/science.ade2574.

43. Madani, A., Krause, B., Greene, E.R., Subramanian, S., Mohr, B.P., Holton, J.M., Olmos, J.L., Xiong, C., Sun, Z.Z., Socher, R., et al. (2023). Large language models generate functional protein sequences across diverse families. Nat. Biotechnol. *41*, 1099–1106. https://doi.org/10.1038/s41587-022-01618-2.

44. Teufel, F., Stahlhut, C., Refsgaard, J.C., Nielsen, H., Winther, O., and Madsen, D. (2023). SecretoGen: Towards Prediction of Signal Peptides for Efficient Protein Secretion. In 37th Annual Conference on Neural Information Processing Systems.

45. Wu, Z., Yang, K.K., Liszka, M.J., Lee, A., Batzilla, A., Wernick, D., Weiner, D.P., and Arnold, F.H. (2020). Signal Peptides Generated by Attention-Based Neural Networks. ACS Synth. Biol. *9*, 2154–2161. https://doi.org/10.1021/acssynbio.0c00219.

46. Vázquez Torres, S., Leung, P.J.Y., Venkatesh, P., Lutz, I.D., Hink, F., Huynh, H.H., Becker, J., Yeh, A.H.W., Juergens, D., Bennett, N.R., et al. (2024). De novo design of high-affinity binders of bioactive helical peptides. Nature *626*, 435–442. https://doi.org/10.1038/s41586-023-06953-1.

47. Masson, H., Tat, J., Di Giusto, P., Antonakoudis, A., Shamie, I., Baghdassarian, H., Samoudi, M., Robinson, C.M., Kuo, C.-C., Koga, N., et al. (2024). A Reconstruction of the Mammalian Secretory Pathway Identifies Mechanisms Regulating Antibody Production. Preprint at bioRxiv. https://doi.org/10.1101/2024.11.14.623668.

48. Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S.H., and von Heijne, G. (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature *433*, 377–381. https://doi.org/10.1038/nature03216.

49. von Heijne, G. (1992). Membrane protein structure prediction: Hydrophobicity analysis and the positive-inside rule. J. Mol. Biol. *225*, 487–494. https://doi.org/10.1016/0022-2836(92)90934-c.

50. Wenzell, N.A., Tuch, B.B., McMinn, D.L., Lyons, M.J., Kirk, C.J., and Taunton, J. (2024). Global signal peptide profiling reveals principles of selective Sec61 inhibition. Nat. Chem. Biol. *20*, 1154–1163. https://doi.org/10.1038/s41589-024-01592-7.

51. Owji, H., Nezafat, N., Negahdaripour, M., Hajiebrahimi, A., and Ghasemi, Y. (2018). A comprehensive review of signal peptides: Structure, roles, and applications. Eur. J. Cell Biol. *97*, 422–441. https://doi.org/10.1016/j.ejcb.2018.06.003.

52. Fukasawa, Y., Leung, R.K.K., Tsui, S.K.W., and Horton, P. (2014). Plus ça change – evolutionary sequence divergence predicts protein subcellular localization signals. BMC Genomics *15*, 46. https://doi.org/10.1186/1471-2164-15-46.

53. Käll, L., Krogh, A., and Sonnhammer, E.L.L. (2004). A Combined Transmembrane Topology and Signal Peptide Prediction Method. J. Mol. Biol. *338*, 1027–1036. https://doi.org/10.1016/j.jmb.2004.03.016.

54. Nielsen, H., Tsirigos, K.D., Brunak, S., and von Heijne, G. (2019). A Brief History of Protein Sorting Prediction. Protein J. *38*, 200–216. https://doi.org/10.1007/s10930-019-09838-3.

55. Hallgren, J., Tsirigos, K.D., Pedersen, M.D., Armenteros, J.J.A., Marcatili, P., Nielsen, H., Krogh, A., Winther, O., et al. (2022). DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. Preprint at bioRxiv. https://doi.org/10.1101/2022.04.08.487609.

56. Bernhofer, M., and Rost, B. (2022). TMbed: transmembrane proteins predicted through language model embeddings. BMC Bioinformatics *23*, 326. https://doi.org/10.1186/s12859-022-04873-x.

57. von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. *14*, 4683–4690. https://doi.org/10.1093/nar/14.11.4683.

58. Arrigo, P., Giuliano, F., Scalia, F., Rapallo, A., and Damiani, G. (1991). Identification of a new motif on nucleic acid sequence data using Kohonen's self-organizing map. Comput. Appl. Biosci. *7*, 353–357. https://doi.org/10.1093/bioinformatics/7.3.353.

59. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. *10*, 1–6. https://doi.org/10.1093/protein/10.1.1.

60. Nielsen, H., and Krogh, A. (1998). Prediction of signal peptides and signal anchors by a hidden Markov model. Proc. Int. Conf. Intell. Syst. Mol. Biol. *6*, 122–130.

61. Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved Prediction of Signal Peptides: SignalP 3.0. J. Mol. Biol. *340*, 783–795. https://doi.org/10.1016/j.jmb.2004.05.028.

62. Bagos, P.G., Nikolaou, E.P., Liakopoulos, T.D., and Tsirigos, K.D. (2010). Combined prediction of Tat and Sec signal peptides with hidden Markov models. Bioinformatics *26*, 2811–2817. https://doi.org/10.1093/bioinformatics/btq530.

63. Hiller, K., Grote, A., Scheer, M., Münch, R., and Jahn, D. (2004). PrediSi: prediction of signal peptides and their cleavage positions. Nucleic Acids Res. *32*, W375–W379. https://doi.org/10.1093/nar/gkh378.

64. Chou, K.-C., Shen, H.-B., and Signal, C.F. (2007). Signal-CF: A subsite-coupled and window-fusing approach for predicting signal peptides. Biochem. Biophys. Res. Commun. *357*, 633–640. https://doi.org/10.1016/j.bbrc.2007.03.162.

65. Shen, H.-B., and Chou, K.-C. (2007). Signal-3L: A 3-layer approach for predicting signal peptides. Biochem. Biophys. Res. Commun. *363*, 297–303. https://doi.org/10.1016/j.bbrc.2007.08.140.

66. Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., and Signal, P. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods *8*, 785–786. https://doi.org/10.1038/nmeth.1701.

67. Almagro Armenteros, J.J., Sønderby, C.K., Sønderby, S.K., Nielsen, H., and Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics *33*, 3387–3395. https://doi.org/10.1093/bioinformatics/btx431.

68. Savojardo, C., Martelli, P.L., Fariselli, P., and Casadio, R. (2018). DeepSig: deep learning improves signal peptide detection in proteins. Bioinformatics *34*, 1690–1696. https://doi.org/10.1093/bioinformatics/btx818.

69. Dumitrescu, A., Jokinen, E., Paatero, A., Kellosalo, J., Paavilainen, V.O., and Lähdesmäki, H. (2023). TSignal: a transformer model for signal peptide prediction. Bioinform. Oxf. Engl. *39*, i347–i356. https://doi.org/10.1093/bioinformatics/btad228.

70. Shen, J., Yu, Q., Chen, S., Tan, Q., Li, J., and Li, Y. (2024). Unbiased organism-agnostic and highly sensitive signal peptide predictor with deep protein language model. Nat. Comput. Sci. *4*, 29–42. https://doi.org/10.1038/s43588-023-00576-2.

71. Pechmann, S., Chartron, J.W., and Frydman, J. (2014). Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. Nat. Struct. Mol. Biol. *21*, 1100–1105. https://doi.org/10.1038/nsmb.2919.

72. Wang, X., Das, T.K., Singh, S.K., and Kumar, S. (2009). Potential aggregation prone regions in biotherapeutics: A survey of commercial monoclonal antibodies. mAbs *1*, 254–267. https://doi.org/10.4161/mabs.1.3.8035.

73. Reily, C., Stewart, T.J., Renfrow, M.B., and Novak, J. (2019). Glycosylation in health and disease. Nat. Rev. Nephrol. *15*, 346–366. https://doi.org/10.1038/s41581-019-0129-4.

74. Torrent, M., Chalancon, G., de Groot, N.S., Wuster, A., and Madan Babu, M. (2018). M. Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. Sci. Signal. *11*, eaat6409. https://doi.org/10.1126/scisignal.aat6409.

75. Gingold, H., Tehler, D., Christoffersen, N.R., Nielsen, M.M., Asmar, F., Kooistra, S.M., Christophersen, N.S., Christensen, L.L., Borre, M., Sørensen, K.D., et al. (2014). A dual program for translation regulation in cellular proliferation and differentiation. Cell *158*, 1281–1292. https://doi.org/10.1016/j.cell.2014.08.011.

76. Gong, Z., Jiang, Z., Gao, W., Zhuo, D., and Ma, L. (2025). A New Deep-Learning-Based Approach for mRNA Optimization: High Fidelity, Computation Efficiency, and Multiple Optimization Factors. Preprint at arXiv. https://doi.org/10.48550/arXiv.2505.23862.

77. Fallahpour, A., Gureghian, V., Filion, G.J., Lindner, A.B., and Pandi, A. (2025). CodonTransformer: a multispecies codon optimizer using context-aware neural networks. Nat. Commun. *16*, 3205. https://doi.org/10.1038/s41467-025-58588-7.

78. Ren, Z., Jiang, L., Di, Y., Zhang, D., Gong, J., Gong, J., Jiang, Q., Fu, Z., Sun, P., Zhou, B., et al. (2024). CodonBERT: a BERT-based architecture tailored for codon optimization using the cross-attention mechanism. Bioinform. Oxf. Engl. *40*, btae330. https://doi.org/10.1093/bioinformatics/btae330.

79. van der Kant, R., Karow-Zwick, A.R., Van Durme, J., Blech, M., Gallardo, R., Seeliger, D., Aßfalg, K., Baatsen, P., Compernolle, G., Gils, A., et al. (2017). Prediction and Reduction of the Aggregation of Monoclonal Antibodies. J. Mol. Biol. *429*, 1244–1261. https://doi.org/10.1016/j.jmb.2017.03.014.

80. Zhou, Y., Raju, R., Alves, C., and Gilbert, A. (2018). Debottlenecking protein secretion and reducing protein aggregation in the cellular host. Curr. Opin. Biotechnol. *53*, 151–157. https://doi.org/10.1016/j.copbio.2018.01.007.

81. Le Fourn, V., Girod, P.-A., Buceta, M., Regamey, A., and Mermod, N. (2014). CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion. Metab. Eng. *21*, 91–102. https://doi.org/10.1016/j.ymben.2012.12.003.

82. Van Durme, J., De Baets, G., Van Der Kant, R., Ramakers, M., Ganesan, A., Wilkinson, H., Gallardo, R., Rousseau, F., and Schymkowitz, J. (2016). Solubis: a webserver to reduce protein aggregation through mutation. Protein Eng. Des. Sel. *29*, 285–289. https://doi.org/10.1093/protein/gzw019.

83. Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat. Biotechnol. *22*, 1302–1306. https://doi.org/10.1038/nbt1012.

84. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005). The FoldX web server: an online force field. Nucleic Acids Res. *33*, W382–W388. https://doi.org/10.1093/nar/gki387.

85. Bauer, J., Mathias, S., Kube, S., Otte, K., Garidel, P., Gamer, M., Blech, M., Fischer, S., and Karow-Zwick, A.R. (2020). Rational optimization of a monoclonal antibody improves the aggregation propensity and enhances the CMC properties along the entire pharmaceutical process chain. mAbs *12*, 1787121. https://doi.org/10.1080/19420862.2020.1787121.

86. Zambrano, R., Jamroz, M., Szczasiuk, A., Pujols, J., Kmiecik, S., and Ventura, S. (2015). AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures. Nucleic Acids Res. *43*, W306–W313. https://doi.org/10.1093/nar/gkv359.

87. Zalewski, M., Iglesias, V., Bárcenas, O., Ventura, S., and Kmiecik, S. (2024). Aggrescan4D: A comprehensive tool for pH-dependent analysis and engineering of protein aggregation propensity. Protein Sci. *33*, e5180. https://doi.org/10.1002/pro.5180.

88. Hickman, S., Kulczycki, A., Lynch, R.G., and Kornfeld, S. (1977). Studies of the mechanism of tunicamycin in hibition of IgA and IgE secretion by plasma cells. J. Biol. Chem. *252*, 4402–4408. https://doi.org/10.1016/S0021-9258(17)40279-1.

89. Klement, E., and Medzihradszky, K.F. (2017). Extracellular Protein Phosphorylation, the Neglected Side of the Modification. Mol. Cell. Proteomics *16*, 1–7. https://doi.org/10.1074/mcp.O116.064188.

90. Souri, M., Iwata, H., Zhang, W.G., and Ichinose, A. (2009). Unique secretion mode of human protein Z: its Gla domain is responsible for inefficient, vitamin K–dependent and warfarin-sensitive secretion. Blood *113*, 3857–3864. https://doi.org/10.1182/blood-2008-07-171884.

91. Struck, D.K., Siuta, P.B., Lane, M.D., and Lennarz, W.J. (1978). Effect of tunicamycin on the secretion of serum proteins by primary cultures of rat and chick hepatocytes. Studies on transferrin, very low density lipoprotein, and serum albumin. J. Biol. Chem. *253*, 5332–5337. https://doi.org/10.1016/S0021-9258(17)30374-5.

92. Liu, Y., Nguyen, A., Wolfert, R.L., and Zhuo, S. (2009). Enhancing the secretion of recombinant proteins by engineering N-glycosylation sites. Biotechnol. Prog. *25*, 1468–1475. https://doi.org/10.1002/btpr.241.

93. Elliott, S., Lorenzini, T., Asher, S., Aoki, K., Brankow, D., Buck, L., Busse, L., Chang, D., Fuller, J., Grant, J., et al. (2003). Enhancement of therapeutic protein in vivo activities through glycoengineering. Nat. Biotechnol. *21*, 414–421. https://doi.org/10.1038/nbt799.

94. Jimenez del Val, I., Fan, Y., and Weilguny, D. (2016). Dynamics of immature mAb glycoform secretion during CHO cell culture: An integrated modelling framework. Biotechnol. J. *11*, 610–623. https://doi.org/10.1002/biot.201400663.

95. Kotidis, P., Donini, R., Arnsdorf, J., Hansen, A.H., Voldborg, B.G.R., Chiang, A.W.T., Haslam, S.M., Betenbaugh, M., Jimenez Del Val, I., Lewis, N.E., et al. (2023). CHOGlycoNET: Comprehensive glycosylation reaction network for CHO cells. Metab. Eng. *76*, 87–96. https://doi.org/10.1016/j.ymben.2022.12.009.

96. Kotidis, P., and Kontoravdi, C. (2020). Harnessing the potential of artificial neural networks for predicting protein glycosylation. Metab. Eng. Commun. *10*, e00131. https://doi.org/10.1016/j.mec.2020.e00131.

97. Kellman, B.P., Sandoval, D., Zaytseva, O.O., Brock, K., Baboo, S., Nachmanson, D., Irvine, E.B., Armingol, E., Mih, N., Zhang, Y., et al. (2024). Protein Structure, a Genetic Encoding for Glycosylation. Preprint at bioRxiv. https://doi.org/10.1101/2024.05.15.594261.

98. Hendricks, G.G., Grigoryan, L., Navarro, M.J., Catanzaro, N.J., Hubbard, M.L., Powers, J.M., Mattocks, M., Treichel, C., Walls, A.C., Lee, J., et al. (2024). Computationally designed mRNA-launched protein nanoparticle vaccines. Preprint at bioRxiv. https://doi.org/10.1101/2024.07.22.604655.

99. Hessa, T., Meindl-Beinker, N.M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S.H., and von Heijne, G. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature *450*, 1026–1030. https://doi.org/10.1038/nature06387.

100. Sunley, K., and Butler, M. (2010). Strategies for the enhancement of recombinant protein production from mammalian cells by growth arrest. Biotechnol. Adv. *28*, 385–394. https://doi.org/10.1016/j.biotechadv.2010.02.003.

101. Kuo, C.-C., Chiang, A.W., Shamie, I., Samoudi, M., Gutierrez, J.M., and Lewis, N.E. (2018). The emerging role of systems biology for engineering protein production in CHO cells. Curr. Opin. Biotechnol. *51*, 64–69. https://doi.org/10.1016/j.copbio.2017.11.015.

102. Gutierrez, J.M., and Lewis, N.E. (2015). Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. Biotechnol. J. *10*, 939–949. https://doi.org/10.1002/biot.201400647.

103. Sha, S., Huang, Z., Wang, Z., and Yoon, S. (2018). Mechanistic modeling and applications for CHO cell culture development and production. Curr. Opin. Chem. Eng. *22*, 54–61. https://doi.org/10.1016/j.coche.2018.08.010.

104. Hefzi, H., Ang, K.S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C.A., Baycin-Hizal, D., Huang, Y., Ley, D., et al. (2016). A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary (CHO) Cell Metabolism. Cell Syst. *3*, 434–443.e8.

105. Feizi, A., Österlund, T., Petranovic, D., Bordel, S., and Nielsen, J. (2013). Genome-Scale Modeling of the Protein Secretory Machinery in Yeast. PLoS One *8*, e63284. https://doi.org/10.1371/journal.pone.0063284.

106. Masson, H.O., Samoudi, M., Robinson, C.M., Kuo, C.C., Weiss, L., Shams Ud Doha, K., Campos, A., Tejwani, V., Dahodwala, H., Menard, P., et al. (2024). Inferring secretory and metabolic pathway activity from omic data with secCellFie. Metab. Eng. *81*, 273–285. https://doi.org/10.1016/j.ymben.2023.12.006.

107. Sosa-Carrillo, S., Galez, H., Napolitano, S., Bertaux, F., and Batt, G. (2023). Maximizing protein production by keeping cells at optimal secretory stress levels using real-time control approaches. Nat. Commun. *14*, 3028. https://doi.org/10.1038/s41467-023-38807-9.

108. Fischer, K., Lulla, A., So, T.Y., Pereyra-Gerber, P., Raybould, M.I.J., Kohler, T.N., Yam-Puc, J.C., Kaminski, T.S., Hughes, R., Pyeatt, G.L., et al. (2024). Rapid discovery of monoclonal antibodies by microfluidics-enabled FACS of single pathogen-specific antibody-secreting cells. Nat. Biotechnol. *43*, 960–970. https://doi.org/10.1038/s41587-024-02346-5.

109. Petersen, B.M., Kirby, M.B., Chrispens, K.M., Irvin, O.M., Strawn, I.K., Haas, C.M., Walker, A.M., Baumer, Z.T., Ulmer, S.A., Ayala, E., et al. (2024). An integrated technology for quantitative wide mutational scanning of human antibody Fab libraries. Nat. Commun. *15*, 3974. https://doi.org/10.1038/s41467-024-48072-z.

110. Bronevetsky, Y. (2020). Directly test individual T cell function with fewer cells on the berkeley lights lightning™ platform. Cytotherapy *22*, S119. https://doi.org/10.1016/j.jcyt.2020.03.221.

111. Yu, X., Li, S., Feng, H., Liao, X., Xing, X.H., Bai, Z., Liu, X., and Zhang, C. (2023). CRISPRi-microfluidics screening enables genome-scale target identification for high-titer protein production and secretion. Metab. Eng. *75*, 192–204. https://doi.org/10.1016/j.ymben.2022.12.004.

112. Pinto, I.F., Soares, R.R.G., Mäkinen, M.E.-L., Chotteau, V., and Russom, A. (2021). Multiplexed Microfluidic Cartridge for At-Line Protein Monitoring in Mammalian Cell Culture Processes for Biopharmaceutical Production. ACS Sens. *6*, 842–851. https://doi.org/10.1021/acssensors.0c01884.

113. Diep, J., Le, H., Le, K., Zasadzinska, E., Tat, J., Yam, P., Zastrow, R., Gomez, N., and Stevens, J. (2021). Microfluidic chip-based single-cell cloning to accelerate biologic production timelines. Biotechnol. Prog. *37*, e3192. https://doi.org/10.1002/btpr.3192.

114. Aebischer-Gumy, C., Moretti, P., Ollier, R., Ries Fecourt, C., Rousseau, F., and Bertschinger, M. (2020). SPLICELECTTM: an adaptable cell surface display technology based on alternative splicing allowing the qualitative and quantitative prediction of secreted product at a single-cell level. MAbs *12*, 1709333. https://www.tandfonline.com/doi/full/10.1080/19420862.2019.1709333#d1e349.

115. Rutkauskaite, J., Berger, S., Stavrakis, S., Dressler, O., Heyman, J., Casadevall I Solvas, X., deMello, A., and Mazutis, L. (2022). High throughput single-cell antibody secretion quantification and enrichment using droplet microfluidics-based FRET assay. iScience *25*, 104515. https://doi.org/10.1016/j.isci.2022.104515.

116. Gérard, A., Woolfe, A., Mottet, G., Reichen, M., Castrillon, C., Menrath, V., Ellouze, S., Poitou, A., Doineau, R., Briseno-Roa, L., et al. (2020). High throughput single-cell activity-based screening and sequencing of antibodies using droplet microfluidics. Nat. Biotechnol. *38*, 715–721. https://doi.org/10.1038/s41587-020-0466-7.

117. Winters, A., McFadden, K., Bergen, J., Landas, J., Berry, K.A., Gonzalez, A., Salimi-Moosavi, H., Murawsky, C.M., Tagari, P., and King, C.T. (2019). Rapid single B cell antibody discovery using nanopens and structured light. mAbs *11*, 1025–1035. https://doi.org/10.1080/19420862.2019.1624126.

118. Jorgolli, M., Nevill, T., Winters, A., Chen, I., Chong, S., Lin, F.F., Mock, M., Chen, C., Le, K., Tan, C., et al. (2019). Nanoscale integration of single cell biologics discovery processes using optofluidic manipulation and monitoring. Biotechnol. Bioeng. *116*, 2393–2411. https://doi.org/10.1002/bit.27024.

119. Zhang, W., Li, Q., Jia, F., Hu, Z., and Wei, Z. (2021). A Microfluidic Chip for Screening and Sequencing of Monoclonal Antibody at a Single-Cell Level. Anal. Chem. *93*, 10099–10105. https://doi.org/10.1021/acs.analchem.1c00918.

120. Javanmardi, K., Chou, C.W., Terrace, C.I., Annapareddy, A., Kaoud, T.S., Guo, Q., Lutgens, J., Zorkic, H., Horton, A.P., Gardner, E.C., et al. (2021). Rapid characterization of spike variants via mammalian cell surface display. Mol. Cell *81*, 5099–5111.e8. https://doi.org/10.1016/j.molcel.2021.11.024.

121. Crook, Z.R., Sevilla, G.P., Friend, D., Brusniak, M.Y., Bandaranayake, A. D., Clarke, M., Gewe, M., Mhyre, A.J., Baker, D., Strong, R.K., et al. (2017). Mammalian display screening of diverse cystine-dense peptides for difficult to drug targets. Nat. Commun. *8*, 2244. https://doi.org/10.1038/s41467-017-02098-8.

122. Gaa, R., Mayer, H.M., Noack, D., Kumari, K., Guenther, R., Tsai, S.P., Ji, Q., and Doerner, A. (2023). Mammalian display to secretion switchable libraries for antibody preselection and high throughput functional screening. mAbs *15*, 2251190. https://doi.org/10.1080/19420862.2023.2251190.

123. Waldmeier, L., Hellmann, I., Gutknecht, C.K., Wolter, F.I., Cook, S.C., Reddy, S.T., Grawunder, U., and Beerli, R.R. (2016). Transpo-mAb display: Transposition-mediated B cell display and functional screening of full-length IgG antibody libraries. mAbs *8*, 726–740. https://doi.org/10.1080/19420862.2016.1160990.

124. Akbari, S., and Pirbodaghi, T. (2014). A droplet-based heterogeneous immunoassay for screening single cells secreting antigen-specific antibodies. Lab Chip *14*, 3275–3280. https://doi.org/10.1039/c4lc00082j.

125. Shembekar, N., Hu, H., Eustace, D., and Merten, C.A. (2018). Single-Cell Droplet Microfluidic Screening for Antibodies Specifically Binding to Target Cells. Cell Rep. *22*, 2206–2215. https://doi.org/10.1016/j.celrep.2018.01.071.

126. George, J., and Wang, J. (2016). Assay of Genome-Wide Transcriptome and Secreted Proteins on the Same Single Immune Cells by Microfluidics and RNA Sequencing. Anal. Chem. *88*, 10309–10315. https://doi.org/10.1021/acs.analchem.6b03214.

127. Chen, Z., Lu, Y., Zhang, K., Xiao, Y., Lu, J., and Fan, R. (2019). Multiplexed, Sequential Secretion Analysis of the Same Single Cells Reveals Distinct Effector Response Dynamics Dependent on the Initial Basal State. Adv. Sci. (Weinh) *6*, 1801361. https://doi.org/10.1002/advs.201801361.

128. Ma, C., Fan, R., Ahmad, H., Shi, Q., Comin-Anduix, B., Chodon, T., Koya, R.C., Liu, C.C., Kwong, G.A., Radu, C.G., et al. (2011). A clinical microchip for evaluation of single immune cells reveals high functional heterogeneity in phenotypically similar T cells. Nat. Med. *17*, 738–743. https://doi.org/10.1038/nm.2375.

129. Norman, M., Gilboa, T., Ogata, A.F., Maley, A.M., Cohen, L., Busch, E.L., Lazarovits, R., Mao, C.P., Cai, Y., Zhang, J., et al. (2020). Ultrasensitive high-resolution profiling of early seroconversion in patients with COVID-19. Nat. Biomed. Eng. *4*, 1180–1187. https://doi.org/10.1038/s41551-020-00611-x.

130. Udani, S., Langerman, J., Koo, D., Baghdasarian, S., Cheng, B., Kang, S., Soemardy, C., de Rutte, J., Plath, K., and Di Carlo, D. (2024). Associating growth factor secretions and transcriptomes of single cells in nanovials using SEC-seq. Nat. Nanotechnol. *19*, 354–363. https://doi.org/10.1038/s41565-023-01560-7.

131. Wu, T., Womersley, H.J., Wang, J.R., Scolnick, J., and Cheow, L.F. (2023). Time-resolved assessment of single-cell protein secretion by sequencing. Nat. Methods *20*, 723–734. https://doi.org/10.1038/s41592-023-01841-y.

132. O'Neill, P., Mistry, R.K., Brown, A.J., and James, D.C. (2023). Protein-Specific Signal Peptides for Mammalian Vector Engineering. ACS Synth. Biol. *12*, 2339–2352. https://doi.org/10.1021/acssynbio.3c00157.

133. Czerkinsky, C.C., Nilsson, L.A., Nygren, H., Ouchterlony, O., and Tarkowski, A. (1983). A solid-phase enzyme-linked immunospot (ELISPOT) assay for enumeration of specific antibody-secreting cells. J. Immunol. Methods *65*, 109–121. https://doi.org/10.1016/0022-1759(83)90308-3.

134. Chiou, P.Y., Ohta, A.T., and Wu, M.C. (2005). Massively parallel manipulation of single cells and microparticles using optical images. Nature *436*, 370–372. https://doi.org/10.1038/nature03831.

135. Boga, E., Berger, L., De Roia, A., Zörnig, I., Embacher, R., Eichmüller, S.B., Jäger, D., Harbottle, R., and Schmidt, P. (2023). Abstract 1769: A combination approach of a cellular library and single cell microfluidics analysis for the rapid selection of CAR-T cells. Cancer Res. *83*, 1769. https://aacrjournals.org/cancerres/article/83/7_Supplement/1769/721936.

136. Slavny, P., Hegde, M., Doerner, A., Parthiban, K., McCafferty, J., Zielonka, S., and Hoet, R. (2024). Advancements in mammalian display technology for therapeutic antibody development and beyond: current landscape, challenges, and future prospects. Front. Immunol. *15*, 1469329. https://doi.org/10.3389/fimmu.2024.1469329.

137. Valldorf, B., Hinz, S.C., Russo, G., Pekar, L., Mohr, L., Klemm, J., Doerner, A., Krah, S., Hust, M., and Zielonka, S. (2022). Antibody display technologies: selecting the cream of the crop. Biol. Chem. *403*, 455–477. https://doi.org/10.1515/hsz-2020-0377.

138. Ren, H., Li, J., Zhang, N., Hu, L.A., Ma, Y., Tagari, P., Xu, J., and Zhang, M.Y. (2020). Function-based high throughput screening for antibody antagonists and agonists against G protein-coupled receptors. Commun. Biol. *3*, 146. https://doi.org/10.1038/s42003-020-0867-7.

139. Wolkowicz, R., Jager, G.C., and Nolan, G.P. (2005). A random peptide library fused to CCR5 for selection of mimetopes expressed on the mammalian cell surface via retroviral vectors. J. Biol. Chem. *280*, 15195–15201. https://doi.org/10.1074/jbc.M500254200.

140. Ho, M., Nagata, S., and Pastan, I. (2006). Isolation of anti-CD22 Fv with high affinity by Fv display on human cells. Proc. Natl. Acad. Sci. USA *103*, 9637–9642. https://doi.org/10.1073/pnas.0603653103.

141. Taube, R., Zhu, Q., Xu, C., Diaz-Griffero, F., Sui, J., Kamau, E., Dwyer, M., Aird, D., and Marasco, W.A. (2008). Lentivirus Display: Stable Expression of Human Antibodies on the Surface of Human Cells and Virus Particles. PLoS One *3*, e3181. https://doi.org/10.1371/journal.pone.0003181.

142. Parthiban, K., Perera, R.L., Sattar, M., Huang, Y., Mayle, S., Masters, E., Griffiths, D., Surade, S., Leah, R., Dyson, M.R., et al. (2019). A comprehensive search of functional sequence space using large mammalian display libraries created by gene editing. mAbs *11*, 884–898. https://doi.org/10.1080/19420862.2019.1618673.

143. Huhtinen, O., Salbo, R., Lamminmäki, U., and Prince, S. (2023). Selection of biophysically favorable antibody variants using a modified Flp-In CHO mammalian display platform. Front. Bioeng. Biotechnol. *11*, 1170081. https://doi.org/10.3389/fbioe.2023.1170081.

144. Popp, N.A., Powell, R.L., Wheelock, M.K., Holmes, K.J., Zapp, B.D., Sheldon, K.M., Fletcher, S.N., Wu, X., Fayer, S., Rubin, A.F., et al. (2025). Multiplex, multimodal mapping of variant effects in secreted proteins. Preprint at bioRxiv. https://doi.org/10.1101/2024.04.01.587474.

145. Nan, L., Zhang, H., Weitz, D.A., and Shum, H.C. (2024). Development and future of droplet microfluidics. Lab Chip *24*, 1135–1153. https://doi.org/10.1039/d3lc00729d.

146. Liu, Y., Chen, Z., and Xu, J. (2024). Recent advances in the microfluidic generation of shape-controllable hydrogel microparticles and their applications. Green Chem. Eng. *5*, 16–30. https://doi.org/10.1016/j.gce.2023.02.002.

147. Liu, Z., Chen, S., and Wu, J. (2023). Advances in ultrahigh throughput screening technologies for protein evolution. Trends Biotechnol. *41*, 1168–1181. https://doi.org/10.1016/j.tibtech.2023.03.010.

148. Dimatteo, R., and Di Carlo, D.D. (2022). IL-2 secretion-based sorting of single T cells using high throughput microfluidic on-cell cytokine capture. Lab Chip *22*, 1576–1583. https://doi.org/10.1039/d1lc01098k.

149. Wang, Y., Jin, R., Shen, B., Li, N., Zhou, H., Wang, W., Zhao, Y., Huang, M., Fang, P., Wang, S., et al. (2021). High throughput functional screening for next-generation cancer immunotherapy using droplet-based microfluidics. Sci. Adv. *7*, eabe3839. https://doi.org/10.1126/sciadv.abe3839.

150. Moreira, A., Carneiro, J., Campos, J.B.L.M., and Miranda, J.M. (2021). Production of hydrogel microparticles in microfluidic devices: a review. Microfluid. Nanofluid. *25*, 10. https://doi.org/10.1007/s10404-020-02413-8.

151. Yang, Y., Vagin, S.I., Rieger, B., and Destgeer, G. (2024). Fabrication of Crescent Shaped Microparticles for Particle Templated Droplet Formation. Macromol. Rapid Commun. *45*, e2300721. https://doi.org/10.1002/marc.202300721.

152. Lee, S., de Rutte, J., Dimatteo, R., Koo, D., and Di Carlo, D. (2022). Scalable Fabrication and Use of 3D Structured Microparticles Spatially Functionalized with Biomolecules. ACS Nano *16*, 38–49. https://doi.org/10.1021/acsnano.1c05857.

153. Wang, Y., Zhang, K., Zhao, Y., Li, Y., Su, W., and Li, S. (2023). Construction and Applications of Mammalian Cell-Based DNA-Encoded Peptide/Protein Libraries. ACS Synth. Biol. *12*, 1874–1888. https://doi.org/10.1021/acssynbio.3c00043.

154. Vlahos, A.E., Kang, J., Aldrete, C.A., Zhu, R., Chong, L.S., Elowitz, M.B., and Gao, X.J. (2022). Protease-controlled secretion and display of intercellular signals. Nat. Commun. *13*, 912. https://doi.org/10.1038/s41467-022-28623-y.

155. Praznik, A., Fink, T., Franko, N., Lonzarić, J., Benčina, M., Jerala, N., Plaper, T., Roškar, S., and Jerala, R. (2022). Regulation of protein secretion through chemical regulation of endoplasmic reticulum retention signal cleavage. Nat. Commun. *13*, 1323. https://doi.org/10.1038/s41467-022-28971-9.

156. Ranganathan, S., Suthers, P.F., and Maranas, C.D. (2010). OptForce: An Optimization Procedure for Identifying All Genetic Manipulations Leading to Targeted Overproductions. PLoS Comput. Biol. *6*, e1000744. https://doi.org/10.1371/journal.pcbi.1000744.