# High-throughput design & production of functional antibodies enabled by large-scale computation and an integrated wet-lab validation pipeline

Grzegorz Warzecha, Sandra Skorupska, Bartosz Baranowski, Mateusz Gielata, Piotr Lassota

Genotic Inc

**Abstract**

The translation of in silico antibody designs into functionally validated molecules remains a significant bottleneck in biologics discovery, often hampered by the limited predictive power of existing computational methods and pervasive noise within public datasets. To address this challenge, Genotic has developed an integrated, High-Throughput Platform for Generating Functional Antibodies, underpinned by a large-scale Nvidia cluster and a comprehensive wet-lab validation cycle. Our platform leverages substantial computational resources, including a dedicated cluster featuring 32x NVIDIA H100 GPUs interconnected via 1.6 Tbit/s InfiniBand and over 160 additional high-performance GPUs. Multi-step computational pipeline, refined through the rigorous evaluation of at least a dozen distinct (and often conflicting) computational approaches, incorporates a critical feedback loop where experimental data directly informs and fine-tunes the AI models. This iterative refinement has yielded remarkable success, with 99% of AI-designed candidates being successfully produced in our laboratory. To date, we have designed antibody candidates for approximately 3,000 distinct targets, with antibodies for >100 targets already produced and validated, and an additional ~200 currently in the current production pipeline. Produced antibodies undergo rigorous Quality Control (QC) and functional validation using Immunofluorescence (IF), Flow Cytometry (FC), and Immunohistochemistry (IHC), the latter through collaboration with a leading hospital. While in-house Bio-Layer Interferometry (BLI) provides initial kinetic insights, we are actively pursuing high-throughput Surface Plasmon Resonance (SPR) capabilities. After three years of dedicated development, the Genotic platform demonstrably functions, enabling the scaled design of antibodies that are not only reliably producible but also exhibit high selectivity for their intended targets in laboratory assays. Our integrated approach effectively bridges the gap between computational prediction and experimental reality, delivering validated, functional antibody candidates at scale.

## 1. Introduction: navigating the confluence of antibody engineering and artificial intelligence

### 1.1. The enduring significance and expanding frontiers of antibody applications

Antibodies, exquisite molecular recognition tools shaped by evolution, represent a cornerstone of modern biological sciences and medicine. Their remarkable specificity and affinity have established them as indispensable reagents and transformative therapeutic agents. The demand for novel, high-quality antibodies continues to escalate, driven by their expanding utility across a diverse spectrum of applications. Firstly, within the realm of fundamental research (*Research Use Only*), antibodies are critical for dissecting intricate biological mechanisms, enabling scientists worldwide to probe cellular pathways, protein interactions, and subcellular structures. Providing reliable and well-characterized antibodies is paramount to ensuring the reproducibility and advancement of basic biological understanding.

Secondly, the field of diagnostics, encompassing both human and veterinary medicine, relies heavily on antibodies for the rapid and accurate detection of biomarkers associated with various physiological and pathological states. From simple Lateral Flow Assays (LFAs) and established Enzyme-Linked Immunosorbent Assays (ELISAs) to sophisticated Immunohistochemistry (IHC), Immunofluorescence (IF), and the cutting-edge spatial biology techniques offering unprecedented resolution of tissue microenvironments, high-performance antibodies are the linchpin. Accelerating the development and availability of superior diagnostic antibodies holds the potential to significantly simplify and democratize diagnostic workflows globally, aligning with our core mission at Genotic.

Thirdly, therapeutic antibodies have revolutionized the treatment landscape for numerous diseases, particularly in oncology and autoimmune disorders. Developed predominantly by large pharmaceutical companies, the success of these therapies hinges critically on meticulous engineering to optimize specificity, minimize immunogenicity, enhance stability, and fine-tune effector functions. The pursuit of next-generation therapeutic antibodies with improved efficacy and safety profiles remains a major focus of the biopharmaceutical industry.

Furthermore, the burgeoning field of cell-based therapies, exemplified by Chimeric Antigen Receptor T-cell (CAR-T) technology, leverages the targeting capabilities of antibody fragments (such as scFvs) to redirect immune cells against specific disease targets. The design and rapid generation of novel, high-affinity binding domains are crucial for expanding the applicability and efficacy of these potent therapeutic modalities.

Finally, these advancements converge towards the ultimate frontier of personalized medicine. The ability to rapidly design and produce bespoke antibodies, tailored to an individual patient's specific molecular profile or disease state, promises a paradigm shift in treatment. An automated pipeline, integrating design, production, and validation, powered by substantial computational resources, is envisioned as the key enabler for delivering truly personalized antibody-based interventions, potentially within clinically relevant timeframes.

### 1.2. The transformative potential and inherent challenges of AI in antibody discovery

The advent of Artificial Intelligence (AI) and Machine Learning (ML) has been heralded as a catalyst for a "fourth wave" of innovation in biologics discovery, promising to accelerate the design and optimization of complex molecules like antibodies. The potential is undeniable: AI offers the prospect of navigating vast sequence and structural spaces, predicting key properties, and generating novel candidates far exceeding the throughput of traditional

methods. However, the application of AI to antibody engineering is fraught with significant challenges. The profound complexity of accurately modelling antibody structure, function, binding kinetics, and developability necessitates deep neural networks often comprising tens of billions of parameters.

Crucially, the efficacy of these models is fundamentally constrained by the quality and scope of available training data. Pervasive errors, biases, and incompleteness inherent in publicly available datasets – encompassing structural information, sequence-function relationships, and developability parameters – pose a formidable obstacle. These data limitations make it exceedingly difficult, if not impossible, to train computational models that reliably predict real-world performance in the laboratory based solely on *in silico* metrics. Consequently, a recurring observation is that many promising computational approaches, while performing well on benchmark datasets, fail to yield functional, producible antibodies when subjected to experimental validation.

### 1.3. The persistent gap between *in silico* prediction and *wet lab* reality

This disparity highlights a significant chasm between computational prediction and experimental reality – the "translation gap." Standard computational benchmarks frequently falter in capturing the nuances of protein expression, folding, solubility, stability, and specific binding within a complex biological milieu. The predictive power of even the most sophisticated *in silico* tools often proves inadequate when confronted with the stringent requirements of laboratory production and functional validation. Our own extensive evaluations of numerous computational methodologies have consistently reinforced this observation: success *in silico* does not guarantee success *in vitro* or *in vivo*. This inherent limitation underscores the inescapable conclusion that empirical validation remains the ultimate arbiter of antibody functionality and viability. Without a robust feedback mechanism from the wet lab, purely computational approaches risk generating candidates that are merely theoretical constructs, lacking practical utility.

### 1.4. The Genotic vision: an integrated platform bridging computation and experimentation

It is precisely this challenge that Genotic aims to address through the development of an intimately integrated, end-to-end platform. Our core vision centres on seamlessly merging large-scale, AI-driven computational design with a complete, in-house wet-laboratory workflow encompassing protein production, purification, quality control, and multi-modal functional validation. This holistic approach is predicated on a critical **feedback loop**, whereby experimental data – confirming successful production, appropriate biophysical properties, and specific functional activity – is systematically channelled back to refine and retune our predictive AI models. This continuous cycle of design, build, test, and learn moves beyond mere prediction towards validated execution. Our ultimate ambition, driving this integration, is to enable the rapid development of personalized antibody solutions, potentially striving towards a timeframe where, leveraging significant computational power (e.g., ~200 high-performance GPUs), a tailored therapeutic or diagnostic candidate could be generated for a patient within approximately 48 hours. Such candidates could potentially be delivered not only as purified proteins but also via mRNA encoded within lipid nanoparticles (LNPs), offering further therapeutic flexibility.

### 1.5. Purpose and scope of this paper

This paper aims to elucidate the architecture and capabilities of the Genotic integrated platform. We will detail the substantial computational infrastructure underpinning our design engine and outline the systematic, multi-year effort undertaken to evaluate and refine our computational pipeline. Furthermore, we will describe the key stages of our laboratory processes for antibody production and validation, providing tangible evidence of the platform's ability to generate functional molecules at scale. We will present data demonstrating the successful production and functional characterization (including IF, FC, and IHC) of numerous antibody candidates designed using our approach, highlighting the high success rate achieved through our integrated methodology. While proprietary algorithmic details and specific protocol parameters necessarily remain confidential to protect our intellectual property, this work will transparently discuss the overall workflow (**Figure 1**), the scale of our operations, the key validation techniques employed, and the crucial role of the feedback loop in achieving reliable, functional antibody discovery. We believe this integrated approach represents a significant step towards overcoming the translation gap and accelerating the delivery of impactful antibody-based solutions.

## 2. The Genotic platform: an integrated end-to-end architecture for functional antibody discovery

### 2.1. The core concept: a unified ecosystem from computation to validation

Addressing the limitations inherent in fragmented approaches to antibody discovery requires a paradigm shift towards holistic integration. The Genotic platform embodies this shift, functioning as a unified ecosystem that seamlessly orchestrates the entire workflow from initial target selection to the delivery of functionally validated antibody candidates. This end-to-end architecture is designed not merely as a linear sequence of steps, but as a dynamic, learning system built upon three interconnected pillars: advanced Computational Design, robust Protein Production, and rigorous Functional Validation. The interplay between these pillars, particularly the crucial feedback mechanism, forms the cornerstone of our strategy for accelerating the discovery of effective antibodies. **Figure 1** presents a schematic representation of the integrated platform, illustrating the cyclical flow of information and materials that defines our process.

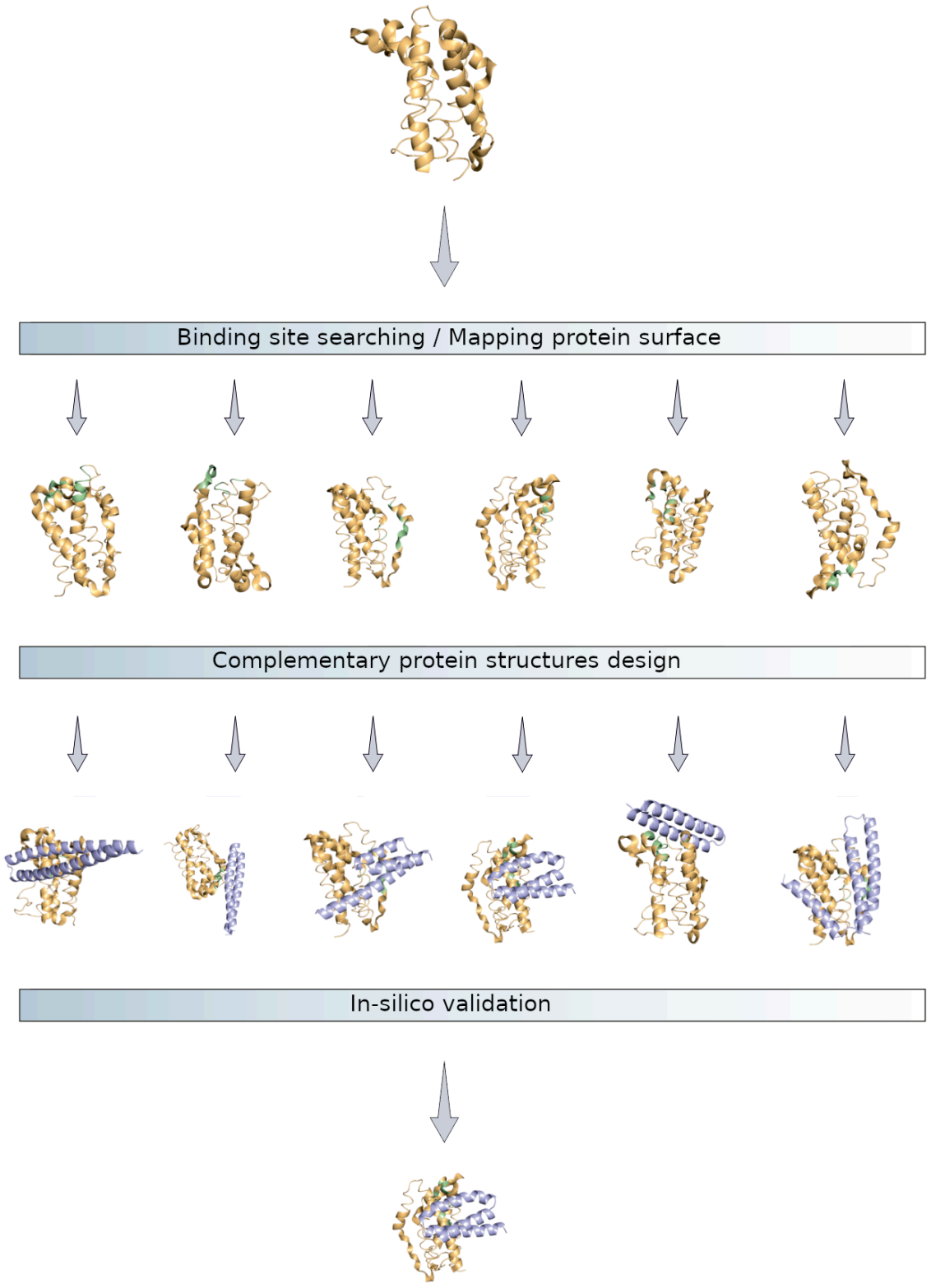



**Figure 1.** The selected target protein is first subjected to a comprehensive surface analysis to identify potential ligandable pockets; when automated binding-site detection fails to reveal discrete cavities, an exhaustive random mapping of the protein surface is performed to nominate additional interaction hotspots. For each candidate site, a diverse ensemble of complementary protein scaffolds is generated by a deep-learning model and iteratively refined via a recycle protocol that alternates between structure prediction and sequence optimization, thereby converging on designs with enhanced binding geometry and residue complementarity. Top-ranking candidates are then rigorously evaluated in silico: molecular dynamics simulations probe complex stability and

conformational resilience; redocking experiments confirm reproducibility of binding modes; cross-docking assays against a panel of off-target proteins and domain-similarity analyses assess specificity; and thermodynamic and kinetic parameters—including interaction energy, dissociation constant (K_d), Gibbs free energy change (ΔG), and overall stability energy—are calculated alongside aggregation-propensity predictions to ensure favorable biophysical behavior under physiological conditions. Finally, the most promising designs are selected for experimental validation in the wet lab.

### 2.2. Pillars of the platform: a synergistic triad

The platform's operation rests on the specialized functions and synergistic interaction of its three core components:

**Pillar I: computational design – a detailed, multi-step workflow:**
This pillar leverages substantial Nvidia cluster and deep learning models to navigate the complexities of antibody design. Far from being a monolithic black box, the process is intrinsically multi-step, demanding meticulous execution and computationally intensive analysis at each stage:

**Target selection and structural preparation:** The journey begins with the careful selection of the therapeutic or diagnostic target protein. A critical, non-trivial initial step involves obtaining a high-quality, reliable atomic structure (PDB/mmCIF format). This is challenging due to frequent errors, inconsistencies, and multiple differing structures (due to presence of other particles) often found in public databases for the same protein. Furthermore, static protein structures only partially represent the inherent dynamics of proteins in solution. Significant effort, often requiring expert curation ("data scientist approach"), is invested in selecting or refining a structural representation that best reflects the relevant conformational state(s) of the target.

**Epitope identification and strategic prioritization:** Once a suitable target structure is established, potential binding sites (epitopes) are identified across its surface **(Figure 2)**. However, simply finding potential sites is insufficient. A crucial subsequent step involves **ranking these epitopes** based on multiple criteria using several proprietary computational approaches. Key considerations include the predicted potential for achieving high-affinity binding (favoring epitopes likely to facilitate low-energy binding states) and, critically, the **uniqueness of the epitope** across the accessible human proteome (e.g., proteins present in the bloodstream or on the surface of non-target cells). Proactively selecting unique epitopes significantly mitigates the risk of off-target binding and non-specific interactions later in the development process.

**Binding domain design and generation:** With prioritized epitopes selected, the core design phase commences. Depending on the intended application, we employ distinct strategies: either designing or grafting Complementarity-Determining Regions (CDRs) onto suitable antibody frameworks, or generating *de novo* single-domain antibodies (e.g., VHH fragments). The objective is to generate sequences predicted to fold into structures that exhibit strong and specific binding to the chosen epitope, aiming for low dissociation constants (high affinity, low KD) and favorable binding energetics.

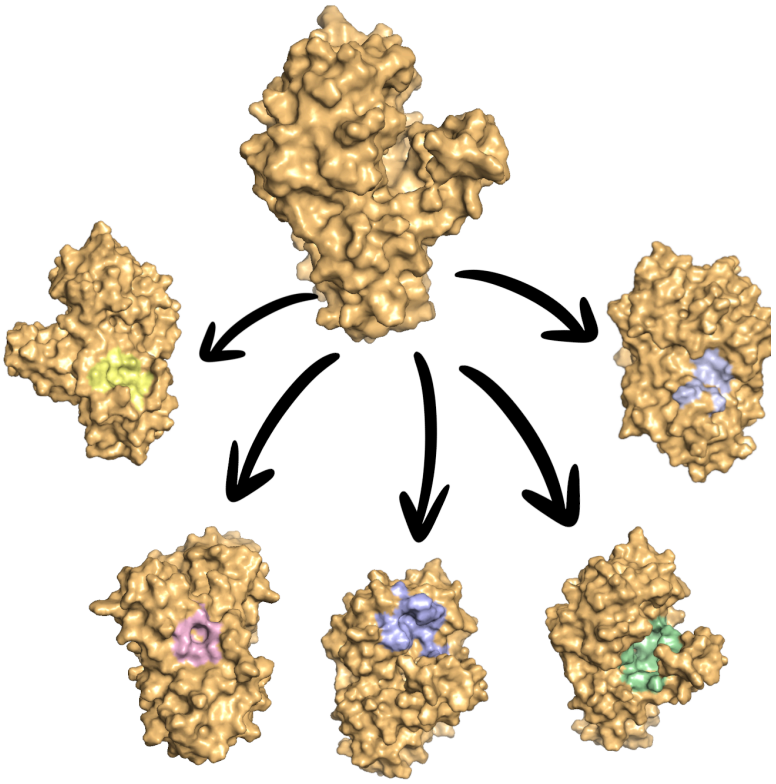


**Figure 2.** A detailed surface-mapping analysis of the target protein is performed to identify potential ligand-binding sites by integrating multiple physicochemical descriptors: solvent-accessible surface area (SASA) calculations delineate exposed cavities and grooves, while grid-based pocket detection algorithms quantify pocket volume, depth, and surface curvature. Simultaneously, hydrophobicity is mapped using normalized hydropathy indices to locate nonpolar patches capable of driving van der Waals interactions, and electrostatic potential surfaces are computed via the Poisson–Boltzmann equation to reveal regions of complementary charge for salt-bridge and hydrogen-bond formation. Hydrogen-bond donor and acceptor propensities are evaluated by scanning for suitable side-chain and backbone atoms within solvent-exposed clefts, and fragment-based energy mapping (mapping small organic probes across the surface) pinpoints energetic hotspots where favorable interaction free energies accumulate. These combined analyses yield a ranked list of high-priority binding hotspots, characterized by optimal balance of hydrophobic enclosure, electrostatic complementarity, hydrogen-bonding capacity, and geometric enclosure, thereby guiding subsequent design efforts toward the most promising interaction sites.

**Structural and *in silico* validation cascade:** The designed amino acid sequence of the binding domain is then translated into a predicted three-dimensional atom level structure, often employing techniques like inverse folding algorithms (**Figure 3**). This predicted

structure undergoes a rigorous, multi-parameter *in silico* validation cascade to assess its viability before committing to laboratory production:

**Structural integrity and fit:** Metrics such as Root Mean Square Deviation (RMSD) are calculated to evaluate how well the designed binder fits the target epitope and whether its predicted structure is energetically sound.

**Binding affinity refinement:** Advanced scoring functions and simulations are used to refine predictions of binding energy and estimate the KD.

**Developability prediction:** Crucially, proprietary models assess the likelihood of the designed sequence being successfully expressed and folded in standard laboratory systems (e.g., *E. coli* or mammalian cells). This predictive step helps eliminate candidates likely to fail during production.

**Specificity and cross-reactivity checks:** The designed binder for VHH and CDR's structure is computationally screened against databases of other human proteins, particularly membrane proteins and secreted proteins, to predict potential off-target interactions.

**Novelty and safety assessment:** Sequences are compared against known antibody databases to ensure novelty and against databases of known toxins or proteins with undesirable characteristics to flag potential safety concerns.

**Integrated fitness score:** Multiple parameters are often combined into a composite "fitness function" score, providing a holistic assessment of the candidate's potential.

**Candidate selection for production:** Only those candidates that successfully pass this stringent, multi-parameter *in silico* filtering process, demonstrating predicted high affinity, specificity, structural soundness, and good developability prospects, are selected for advancement to laboratory production.

This intricate, multi-step computational process underscores why antibody design remains challenging. It necessitates not only massive computational power to run the underlying simulations and deep learning models (often multi-billion parameter networks) but also careful orchestration and expert oversight. While our current pipeline involves these distinct stages, we are continuously working to further streamline and merge steps where feasible, enhancing overall efficiency.

**Pillar II: Protein production & purification:** Candidate designs selected from the computational pillar are seamlessly transferred to our in-house laboratory facilities for expression and purification. Utilizing optimized protocols for both prokaryotic (*E. coli*) and eukaryotic (mammalian) systems, we ensure efficient and reliable production of the designed antibody proteins. Standardized, multi-step chromatography processes yield high-purity material suitable for downstream characterization and validation.

**Pillar III: Functional validation & data feedback:** Produced antibodies undergo rigorous quality control and comprehensive functional validation. This includes confirmation of binding to the intended target using techniques such as Immunofluorescence (IF), Flow Cytometry

(FC), and Immunohistochemistry (IHC). Furthermore, biophysical characterization, including binding kinetics analysis (initially via BLI, with high-throughput SPR as a developmental goal), provides critical quantitative data. The results from this pillar are not merely endpoints but serve as vital inputs for the platform's learning mechanism.

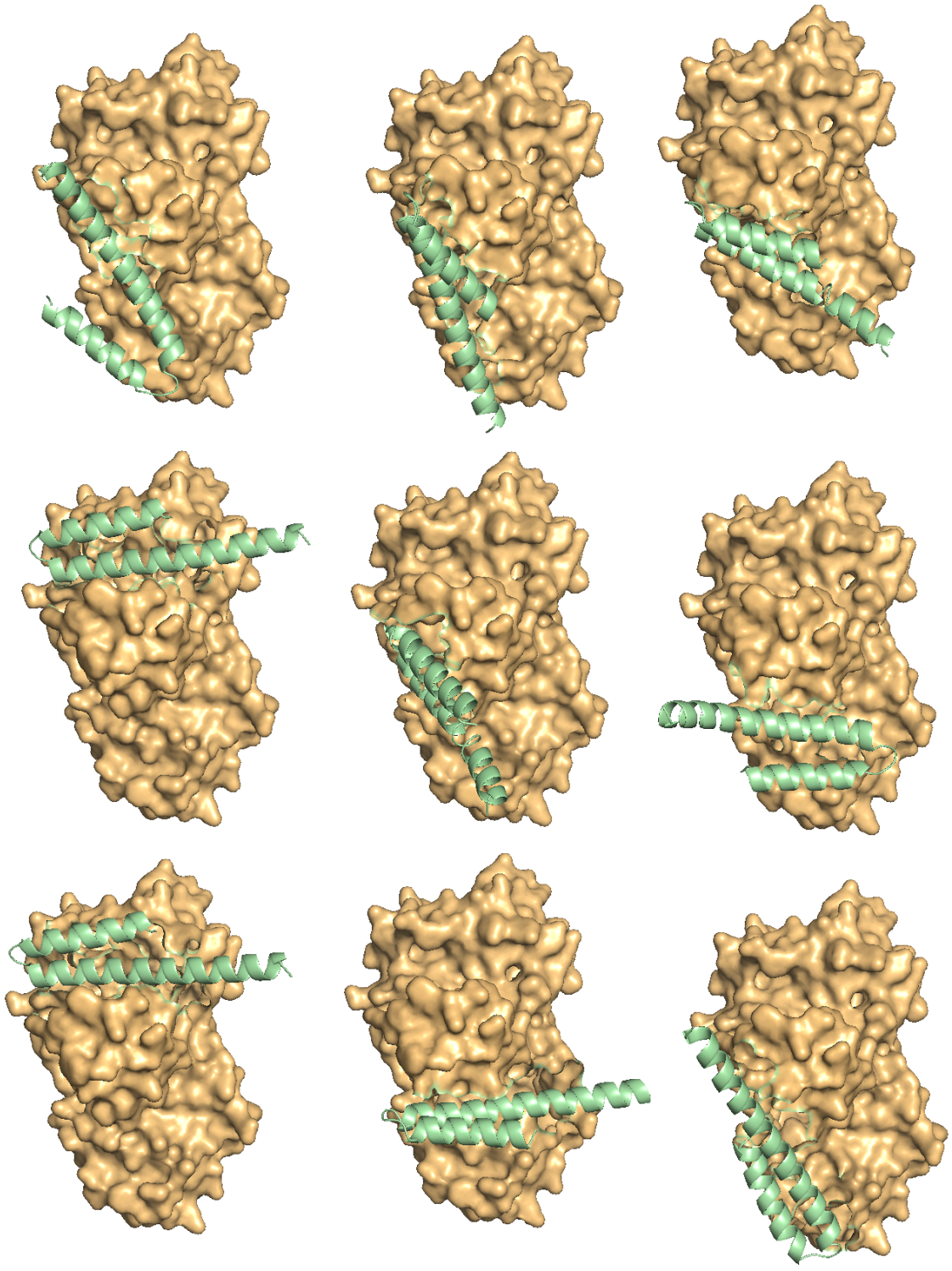


**Figure 3.** Identification of high-affinity leads against the target antigen. Screening revealed several candidates with strong and specific binding.

### 2.3. Synergy & power of an in-house approach

The deliberate integration of all three pillars within a single organization provides decisive advantages over conventional, often fragmented, discovery models that rely on outsourcing distinct stages. This co-location and tight coupling enables:

- **Unprecedented speed and agility:** The direct line of communication and data transfer between computational design, production, and validation teams drastically reduces iteration times. Designs generated computationally can often enter the production pipeline within 24 hours, facilitating rapid testing of hypotheses and acceleration of the entire discovery cycle.
- **Enhanced control and consistency:** Maintaining control over the entire process ensures methodological consistency and minimizes variability often introduced when transferring protocols and materials between different organizations. This improves the reliability and comparability of data across different projects and iterations.
- **High-fidelity feedback:** The proximity and shared infrastructure allow for the systematic capture and transfer of high-quality experimental data back into the computational pipeline. This avoids data loss or misinterpretation common in outsourced models, ensuring that the AI models learn from accurate, real-world results.
- **Cost-effectiveness and resource optimization:** While the computational demands are significant, performing intensive *in silico* pre-screening and optimization is considerably more cost-effective than pursuing numerous unpromising candidates through expensive and time-consuming laboratory experiments. Furthermore, our integrated model allows for efficient resource allocation, avoiding reliance on animal models for initial discovery phases and optimizing laboratory workflows based on computational insights (e.g., predicting potential production bottlenecks or toxicity risks based on similarity to known toxins).

### 2.4. The engine room: Nvidia as the enabler of integration and scale

The practical realization of this integrated vision is critically dependent on our substantial, state-of-the-art High-Performance Computing (HPC) infrastructure. Comprising a powerful core of **NVIDIA Hopper GPUs** interconnected via high-speed **1.6 Tbit/s InfiniBand**, infrastructure provides the raw power necessary to:

- **Train and execute transformers & diffusion models:** Run the multi-billion parameter deep learning models essential for accurately capturing the nuances of antibody structure-function relationships.
- **Enable High-Throughput Screening:** Perform vast numbers of *in silico* simulations and evaluations required for exploring diverse design possibilities and selecting promising candidates.
- **Facilitate Rapid Iteration:** Quickly process experimental feedback data, retrain or fine-tune models, and generate revised designs, thus accelerating the learning cycle.
- **Optimize Resource Allocation:** Dynamically allocate compute power, allowing for focused "bursts" on high-priority targets while maintaining continuous background processing on other projects, ensuring near-constant utilization and maximizing throughput.

Our Nvidia setup is therefore not merely a tool for computation but the fundamental engine enabling the speed, scale, and iterative learning capacity of the integrated platform.

### 2.5. The crucial feedback loop: learning from experimental reality

A defining feature of the Genotic platform is its explicit and functional **feedback loop**. This mechanism bridges the gap between computational prediction and experimental outcome, transforming the platform into an adaptive learning system. Key experimental data, particularly quantitative metrics such as **averaged Dissociation Constants (KD) obtained from multiple kinetic measurements (e.g., BLI)**, along with qualitative data on expressibility, stability, and functional assay performance (IF/FC/IHC), are systematically fed back into the computational design pillar.

This feedback directly informs subsequent design iterations by:

- **Refining predictive models:** Allowing AI models to learn correlations between sequence/structural features and actual laboratory performance, improving their predictive accuracy for affinity, expressibility, and developability.
- **Updating selection criteria:** Modifying the scoring functions and selection thresholds used to rank candidates based on empirically observed success rates.
- **Guiding design strategies:** Identifying design motifs or features associated with poor experimental outcomes, allowing the algorithms to avoid these pitfalls in the future.

Establishing this effective feedback loop has been a non-trivial undertaking, representing the culmination of **three years of intensive development, testing, and refinement**. It is this iterative learning process, grounded in experimental reality, that enables the platform to progressively improve its ability to generate antibodies that function reliably in the laboratory.

### 2.6. The Genotic workflow: a streamlined journey to validated candidates

In practice, a typical project progresses through the platform via a streamlined workflow: commencing with target identification and rigorous *in silico* characterization, followed by the intricate computational design phase generating prioritized candidates. These candidates seamlessly transition to the laboratory for production and purification, succeeded by comprehensive QC and functional validation using a suite of assays (IF, FC, IHC, kinetics). The resulting experimental data then flows back to inform and enhance the computational engine. This tightly integrated, iterative cycle is the fundamental operational principle of the Genotic platform, designed to consistently deliver functional, selective, and validated antibody candidates with significantly improved efficiency and success rates compared to traditional, siloed approaches. The subsequent sections will delve into the specifics of each pillar, providing further detail on the methodologies and capabilities underlying this integrated architecture.

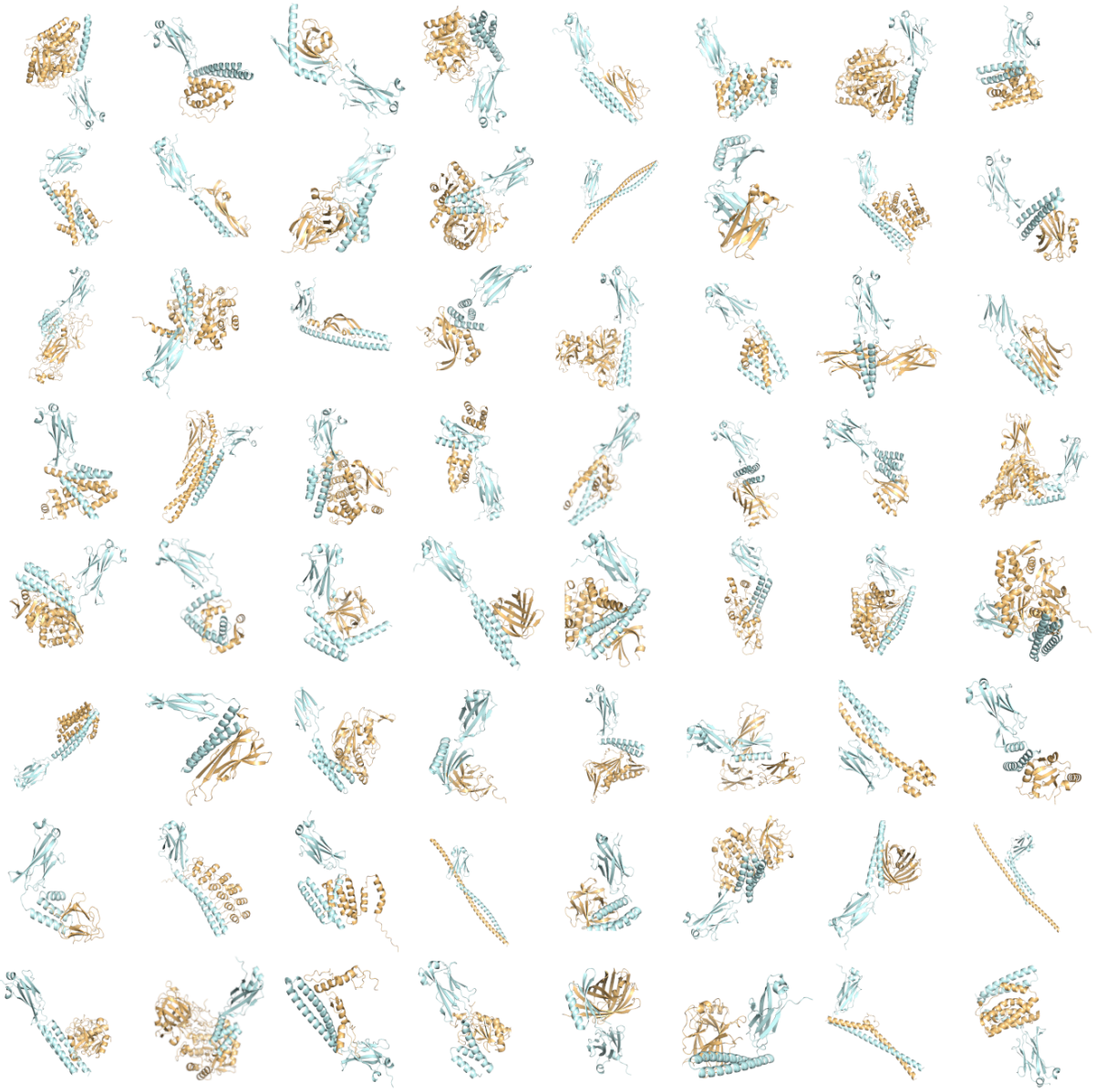


**Figure 4.** Above are presented antibody–antigen complexes available on our site; they include antibodies targeting the following antigens: ACTA2, Actin beta, Adiponectin, Aldoc, Anti-Dog minor allergen, Argonaute 2, AXL, BAX, BDNF, BRCA1, CAMK1, CATF, Collagen, CTIP, Cyclin D1, Doublecortin, EGFP, ERK2, FABP6, FGF10, FGF2, FOXP3, β-galactosidase , GAPDH, GFAP, GRP78, HIF-1α, Histone H3, human Fc, IFN-γ, IFN-γ (mouse), IκBα, IL-2, IL-6, Ki-67, Laminin, LAMP1, LC3B, MYC, Myoglobin, NRF2, OSMR, p53, PCNA, PECAM1, POU5F, Proteinase K, RFP, RUNX, S100A1, SHH, SMAD3, SQSTM1, STAT3, TET2, TGM2, TMPRSS2, TNF-α, TRAIL, TSG101, α-tubulin, βIII-tubulin, VEGFA1, Vimentin, WT1 and YNBA.

### 2.7 Current State of Computational Antibody Design Efforts

Our ongoing computational work has reached a significant milestone. The generation of the resulting substantial dataset, which involved the systematic testing and refinement of multiple *in silico* design approaches and analytical pipelines, required an estimated 25,000 GPU days of computational resources. This extensive exploration has enabled the creation and preliminary assessment of approximately 6 million unique antibody candidates. These candidates have been evaluated against a diverse set of 5,000 distinct protein targets, focusing on interactions within 14,000 characterized potential binding sites. This large-scale computational effort provides a foundational dataset and validated methodologies underpinning our developing solution for the rational design of novel antibodies tailored to engage specific, pre-selected protein targets.

## 3. From design to bench: production and validation of antibodies

### 3.1 Expression system strategy for high-throughput validation

To facilitate the rapid screening and validation of numerous AI-designed antibody variants, we employ both prokaryotic (bacterial) and eukaryotic (mammalian) expression systems. Currently, our primary focus utilizes Escherichia coli (E. coli) expression systems. This strategic choice is driven by the significantly faster production timelines and lower associated costs compared to mammalian cell culture. Given our immediate goal of validating a large pool of computationally derived candidates to identify promising leads, the efficiency and scalability of bacterial expression are paramount. This approach allows us to maximize the number of antibodies produced and tested within a feasible timeframe and budget, thereby accelerating the feedback loop for our AI design algorithms. While mammalian systems, which offer post-translational modifications potentially crucial for certain applications or antibody formats, remain an option for select candidates or later-stage development, our current high-throughput validation phase heavily relies on the bacterial platform.

### 3.2 Multi-step in-house antibody purification

The purification of recombinant antibodies from expression system lysates is conducted entirely in-house, employing a tailored, multi-step chromatographic strategy optimized for each candidate or batch. We recognize that different antibody constructs exhibit varying expression levels and possess unique biophysical properties influencing their purification behaviour. Consequently, significant effort is dedicated to optimizing purification protocols to maximize both yield and purity.

Our standard purification workflow typically initiates with affinity chromatography, leveraging specific tags engineered onto the recombinant antibodies (e.g., His-tag purification via Immobilized Metal Affinity Chromatography - IMAC). This step achieves substantial initial purification by selectively capturing the target antibody. We consistently observe variability in expression levels across different AI-designed candidates, which is reflected in the yields obtained from this initial capture step. Representative affinity chromatography profiles illustrating this range of expression yields are provided (see **Figure 5** for examples of high and low yield profiles).

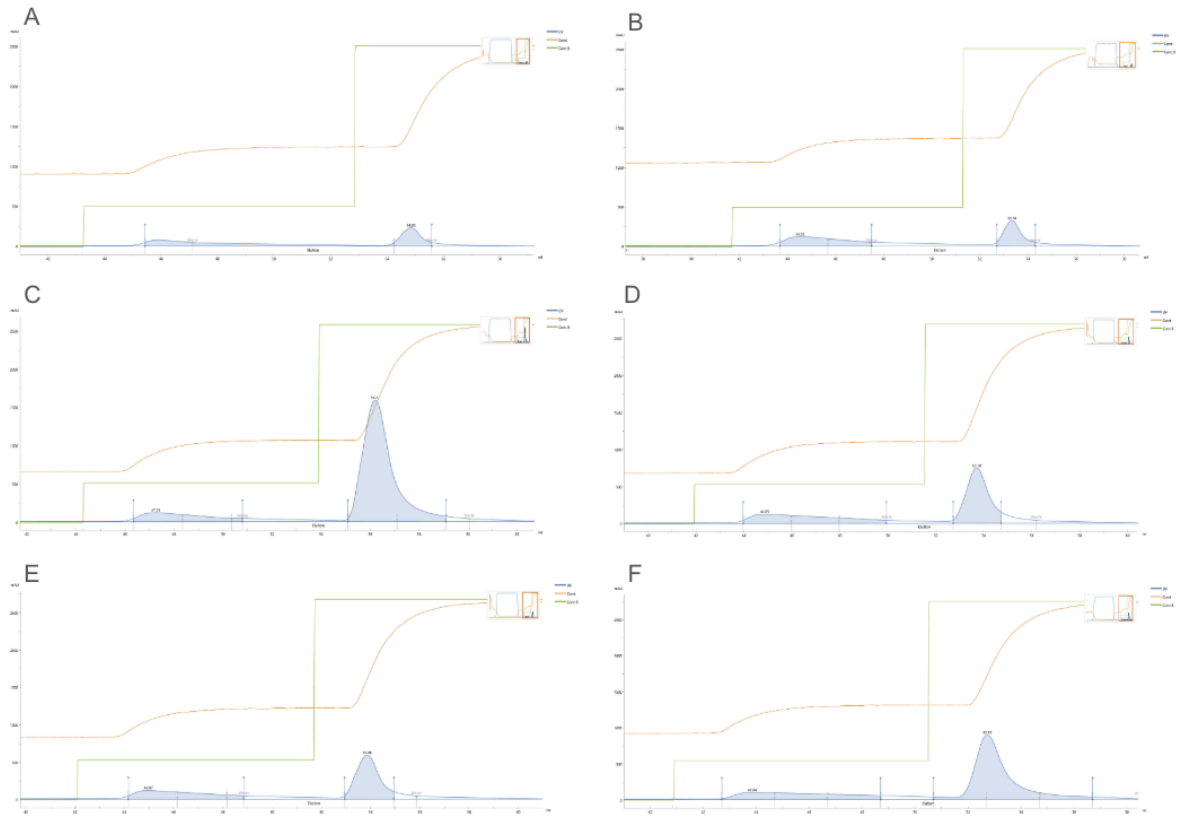



**Figure 5**. IMAC elution profiles of six recombinant antibodies. Shown are representative chromatograms (elution phase only) for (A) anti-GSK3A, (B) anti-SKP2, (C) anti-βIII Tubulin, (D) anti-Ki67, (E) anti-ERK1, and (F) anti-RFP antibodies. Each profile exhibits two peaks: a non-specific early peak eluting at ~20% elution buffer, corresponding to impurities, and a later peak at 100% elution buffer, corresponding to the target antibody.

Following affinity purification, Size Exclusion Chromatography (SEC) is routinely employed as a polishing step. SEC separates molecules based on their hydrodynamic radius, effectively removing aggregates, lower molecular weight contaminants, and buffer exchanging the antibody into a suitable final formulation. Our aim is to isolate a homogenous, monomeric antibody population, which is critical for reliable downstream applications and biophysical characterization. We consistently monitor the elution profiles from SEC to confirm the successful isolation of monomeric antibodies. Representative SEC chromatograms demonstrating the elution of pure, single peaks corresponding to the expected molecular weight of the monomeric antibody are shown (see **Figure 6)**.

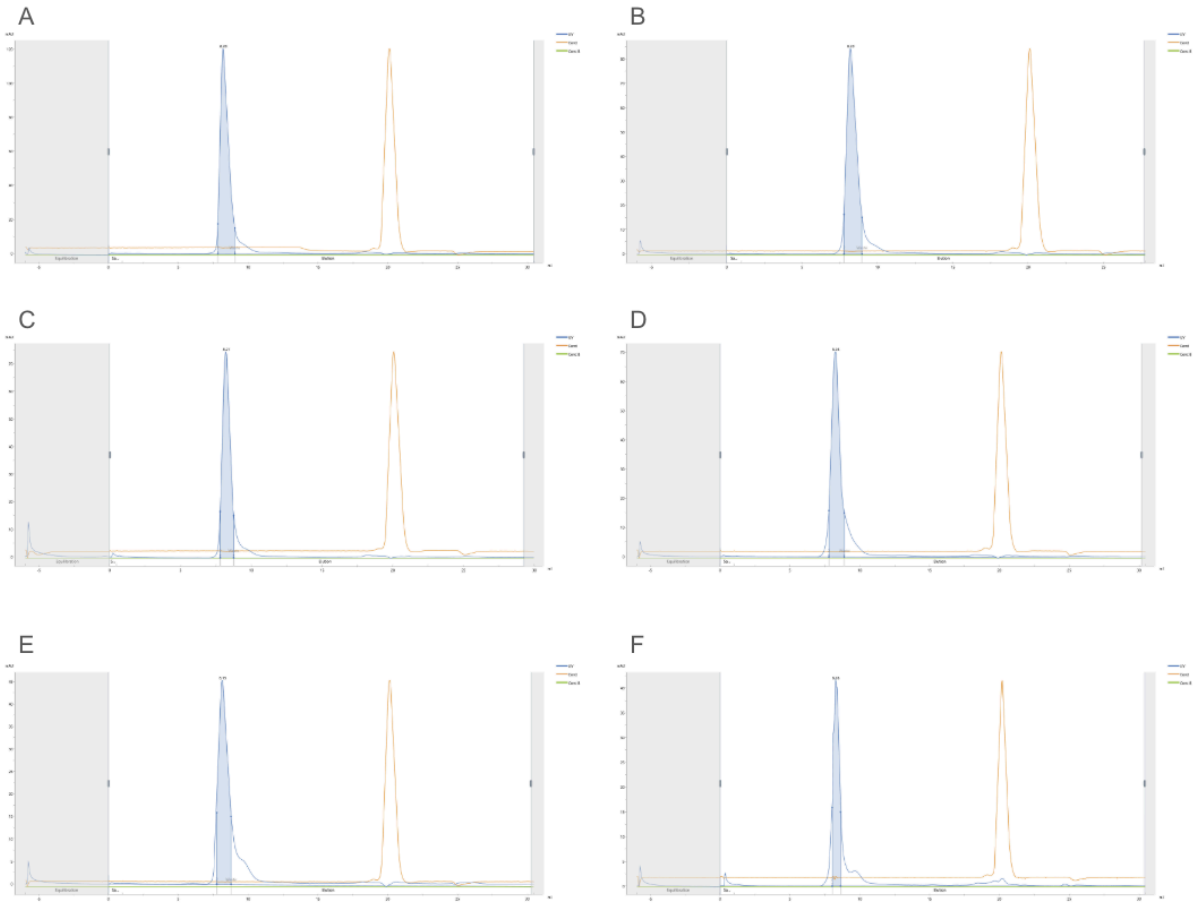



**Figure 6**. Size-exclusion chromatography (SEC) profiles of six purified antibodies. Shown are representative chromatograms for (A) anti-WT33, (B) anti-CD28, (C) anti-TGM2, (D) anti-BMP7, (E) anti-OCT4, and (F) anti-SMAD3 antibodies. Each profile displays a single sharp, symmetrical peak, indicative of a monodisperse antibody preparation and consistent with the expected molecular weight.

In instances where affinity chromatography and SEC do not yield sufficient purity, or for antibodies with particularly challenging contaminant profiles, an additional orthogonal purification step, such as Ion Exchange Chromatography (IEX), may be incorporated. The choice between anion or cation exchange depends on the calculated isoelectric point (pI) of the antibody and the buffer conditions. The entire purification strategy, from the initial capture to the final polishing steps, is subject to continuous in-house optimization to achieve the best possible outcome for each unique antibody candidate.

### 3.3 Rigorous purity assessment

Throughout the purification process, and especially for the final antibody preparation, purity is rigorously monitored using Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) under both reducing and non-reducing conditions. This allows visualization of the antibody heavy and light chains (under reducing conditions) or the intact antibody (under non-reducing conditions) and the detection of any residual host cell proteins or other contaminants. If SDS-PAGE analysis reveals unacceptable levels of impurity following the standard purification workflow, the antibody batch is subjected to further chromatographic

refinement (e.g., an additional IEX step or optimization of the SEC conditions) until the desired level of homogeneity is achieved (typically >95%).

**4. Functional validation in key immunoassays**

A critical aspect of our work is confirming that the purified, AI-designed antibodies are functional and perform effectively in relevant biological applications. We conduct extensive validation across several standard immunoassay platforms.

Immunofluorescence (IF): A significant focus is placed on assessing antibody performance in immunofluorescence staining of cultured cells or tissue sections. Antibodies are tested for their ability to specifically recognize and bind their target antigen in its native cellular context, producing clear and specific staining patterns. We have generated a substantial dataset of IF results across numerous antibody candidates and target antigens. A comprehensive collection of representative immunofluorescence images showcasing successful staining is provided in the Supplementary Information (see **Figure 7** and Supplementary Figures S3-S5).

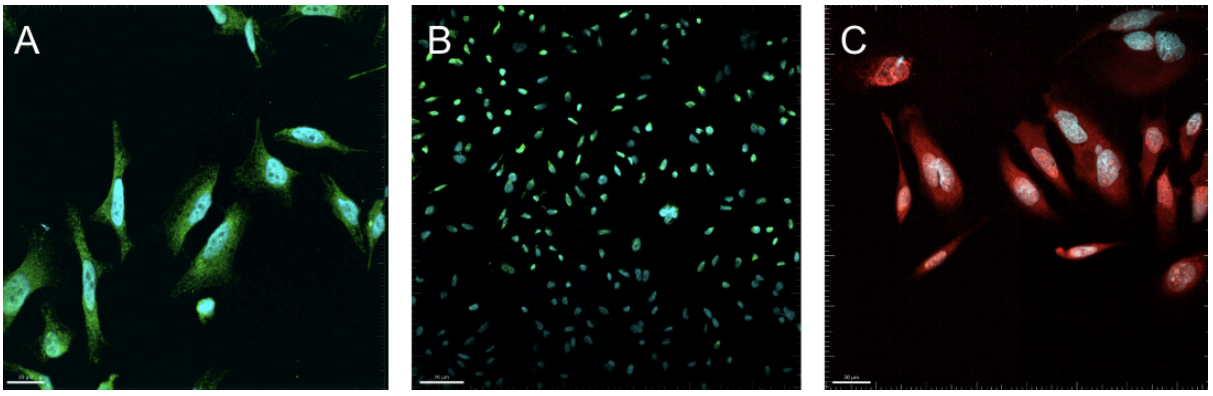



**Figure 7.** Representative confocal images of HeLa cells subjected to immunofluorescence staining using primary antibodies. a, b Staining using antibodies conjugated to Alexa Fluor 488 (green) against: a, GFAP; b, Histone H3. c, Staining for ACTA2 using an antibody conjugated to Alexa Fluor 594 (red). Cell nuclei were counterstained with DAPI (blue).

Flow Cytometry (FC): The utility of our antibodies for detecting cell surface or intracellular antigens via flow cytometry is also thoroughly evaluated. Antibodies are incubated with relevant cell populations, and binding is detected using fluorescently labeled secondary antibodies or by direct conjugation of the primary antibody. Successful candidates demonstrate specific binding to antigen-positive cells with minimal background staining on negative controls. Representative flow cytometry histograms and dot plots demonstrating specific cell population labeling are included (see **Figure 8** and Supplementary Figure S2).

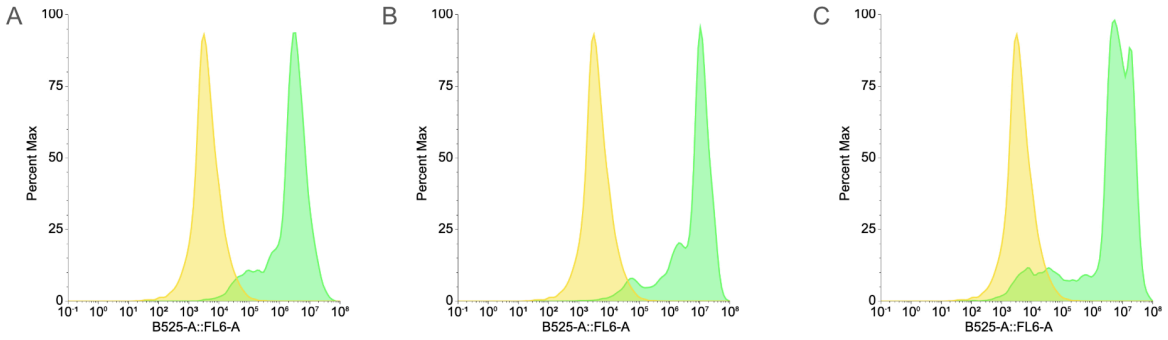



**Figure 8**. Flow cytometry analysis of antibody staining in MDA-MB-231 cells. Histograms show fluorescence intensity profiles for unstained control cells (yellow) and cells stained with antibodies against (A) FGF2, (B) PCNA, and (C) BRCA1 (green). Each plot displays a clear shift in fluorescence signal relative to the control, indicating specific binding of the antibody to its target antigen.

Immunohistochemistry (IHC): For antibodies targeting antigens relevant in tissue pathology, performance in immunohistochemistry on formalin-fixed paraffin-embedded (FFPE) or frozen tissue sections is assessed. Optimization of antigen retrieval methods and antibody dilutions is performed to achieve specific and robust staining of the target antigen within the complex tissue architecture. We have obtained promising IHC results with several AI-designed antibodies, demonstrating their potential utility in research and diagnostic contexts. Examples of specific IHC staining patterns in relevant tissue types are presented (see **Figure 9**).

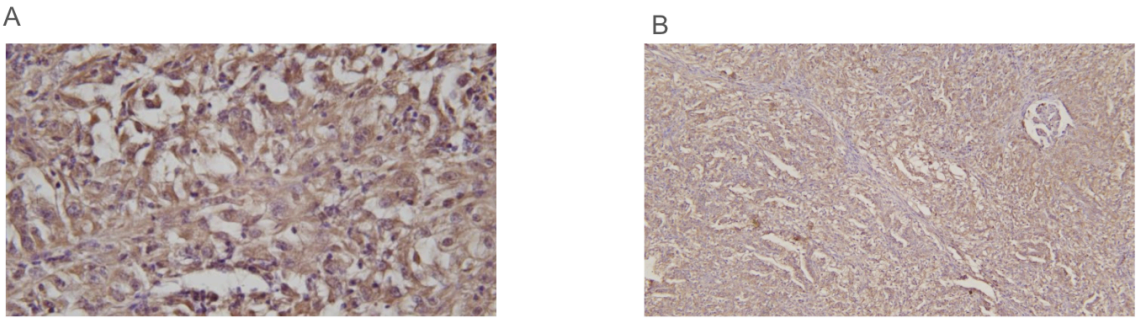



**Figure 9**. Immunohistochemical (IHC) staining of TGM2 and AXL in tissue sections. Representative images showing positive staining for (A) TGM2 and (B) AXL (brown), with nuclear counterstaining using hematoxylin (purple).

### 5. Binding kinetics analysis and ongoing optimization

To quantitatively characterize the binding interaction between our AI-designed antibodies and their target antigens, we perform binding kinetics analysis, primarily using Bio-Layer Interferometry (BLI). This technique allows for the determination of association (kon) and dissociation (koff) rate constants, and the calculation of the equilibrium dissociation constant (KD), a key measure of binding affinity.

While implementing BLI, we have encountered challenges, notably significant levels of non-specific binding (NSB) of some proteins to the sensor surfaces. This NSB can interfere with accurate kinetic measurements and necessitates careful experimental design and data interpretation. We are actively working on optimizing the BLI methodology itself, exploring different sensor types, buffer conditions, surface chemistries, and data processing strategies to mitigate these non-specific interactions and improve the reliability of our kinetic data.

Despite these ongoing technical refinements, we have successfully obtained reliable kinetic data for several of our lead candidates. These analyses demonstrate that our AI-driven design and subsequent production pipeline can yield antibodies with high affinity, achieving dissociation constants (KD) in the low nanomolar range (e.g., down to $10^{-9}$ M). Representative BLI sensorgrams and kinetic fits for selected high-affinity interactions are shown (see **Figure 10** and Supplementary Figure S6). We are committed to further refining both our antibodies and our characterization techniques to consistently achieve and reliably measure high-affinity interactions across a broader range of candidates.

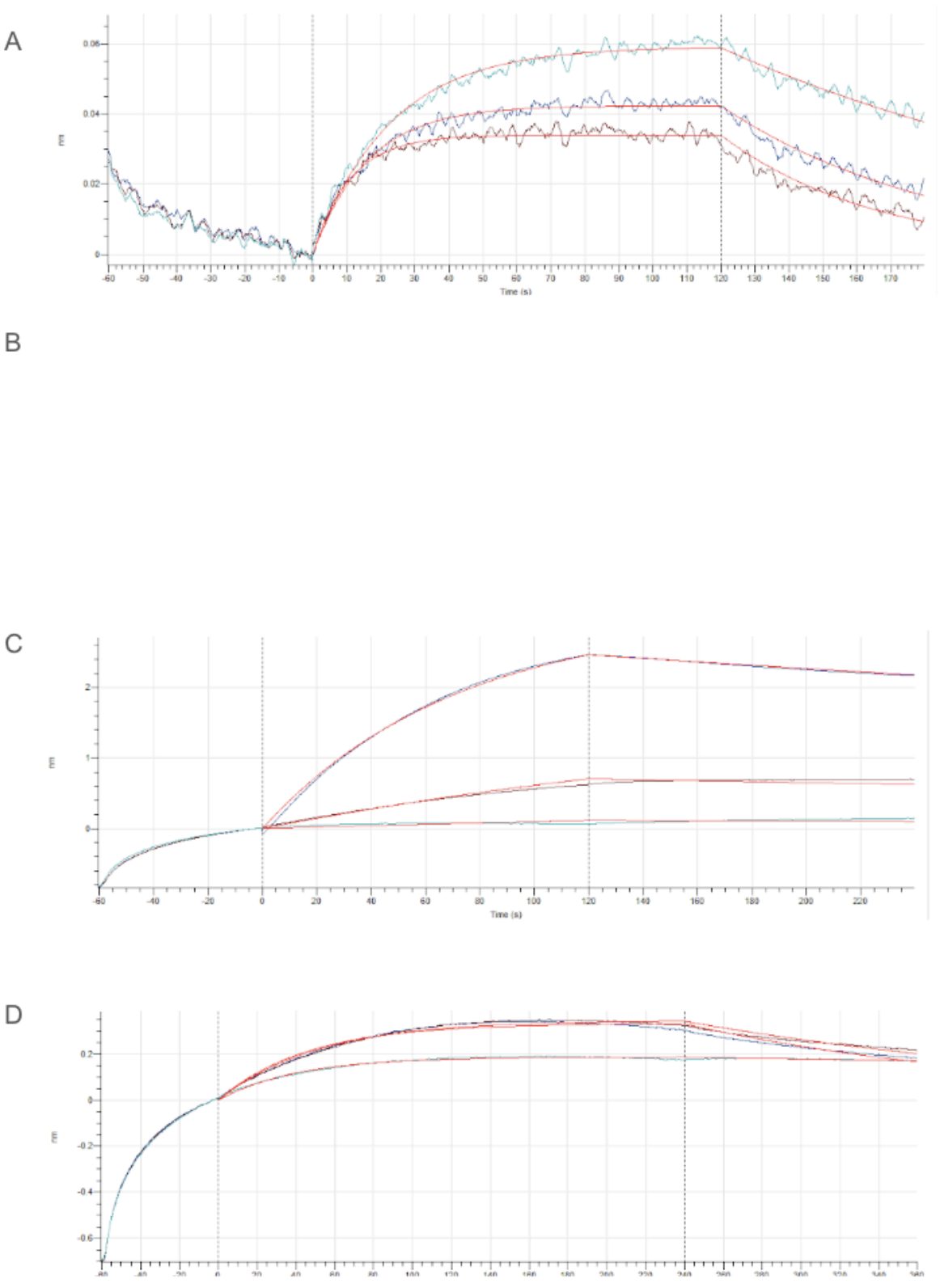



**Figure 10**. Biolayer interferometry (BLI) binding analysis of antibodies–antigen interactions. Shown are representative association and dissociation curves (black) fitted with 1:1 binding models (red) for (A) anti-GAPDH (KD ≈ 10^–6 M), (B) anti-CAMK1 (KD ≈ 10^–7 M), (C) anti-FGF2 (KD≈ 10^-8 M), and (D) anti-FGF10 (KD≈ 10^–9 M) antibodies.

## 6. Scale of operation and future directions

Our laboratory workflow is structured to support the high-throughput nature of validating AI-generated designs. To date, our experimental efforts have encompassed the execution of approximately 3000 individual bacterial cultures for antibody expression, the performance of thousands of distinct protein purification runs, and the utilization of dozens of SDS-PAGE gels for essential quality control. This intensive experimental campaign underscores our commitment to rigorously validating our computational approach.

While we continuously strive to enhance antibody properties, including affinity and developability, through iterative cycles of AI design and experimental testing, our current results already demonstrate significant success. We have successfully produced a portfolio of AI-designed antibodies that exhibit robust and specific performance in widely used and demanding applications such as IF, IHC, and FC. Our ongoing work focuses on further expanding the repertoire of validated antibodies, improving binding affinities and biophysical properties, and refining our high-throughput experimental pipeline to keep pace with the advancements in our AI design capabilities.

## 7. Conclusions

The path from in silico antibody design to functionally validated biologics remains constrained by the predictive limitations of computational models and the inherent noise in experimental systems. To overcome this critical bottleneck, we developed the Genotic integrated platform, culminating three years of focused engineering. This platform uniquely fuses large-scale deep learning design, powered by substantial NVIDIA GPU infrastructure, with an end-to-end, high-throughput wet-laboratory pipeline encompassing production, purification, and multi-modal functional assessment.

A cornerstone of our approach is the rigorously implemented closed feedback loop, where empirical outcomes—quantifying producibility, purity, binding kinetics (low nM affinities achieved), and performance in key assays (IF, FC, IHC)—directly inform and iteratively refine our AI models. This continuous learning cycle, grounded in experimental reality, fundamentally enhances predictive accuracy and design efficacy. The platform's success is evidenced by a 99% production success rate for AI-derived candidates and the validation of functional antibodies against over 100 distinct targets selected from a pool of ~3,000 computationally explored targets.

Our work demonstrates a significant step towards closing the in silico-in vitro gap, proving that a deeply integrated, learning-driven system can deliver functionally validated antibody candidates reliably and at scale. By moving beyond prediction to systematic execution and empirical learning, the Genotic platform substantially accelerates the discovery cycle. While ongoing efforts focus on scaling kinetic analysis (towards high-throughput SPR) and further advancing our AI capabilities, our platform already provides a robust and scalable engine for generating novel antibody candidates. This integrated paradigm paves the way for the rapid development of tailored antibodies for research, diagnostics, and potentially therapeutic applications, ultimately advancing the frontier towards personalized antibody solutions.

**References**

[1] Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature vol. 630 493–500 (2024).
[2] Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science vol. 373 871–876 (2021).
[3] Baek, M. et al. Efficient and accurate prediction of protein structure using RoseTTAFold2. Preprint at bioRxiv (2023).
[4] Bárcenas, O. et al. Aggrescan4D: structure-informed analysis of pH-dependent protein aggregation. Nucleic Acids Research vol. 52 W170–W175 (2024).
[5] Bennett, N. R. et al. Atomically accurate de novo design of antibodies with RFdiffusion. Preprint at bioRxiv (2024).
[6] Bhat, S. et al. De NovoDesign of Peptide Binders to Conformationally Diverse Targets with Contrastive Language Modeling. Preprint at bioRxiv (2023).
[7] Bikias, T., Stamkopoulos, E. & Reddy, Sai. T. PLMFit : Benchmarking Transfer Learning with Protein Language Models for Protein Engineering. Preprint at bioRxiv (2025) .
[8] Chu, L.-S., Sarma, S. & Gray, J. J. Unified Sampling and Ranking for Protein Docking with DFMDock. Preprint at bioRxiv (2024).
[9] Cutting, D., Dreyer, F. A., Errington, D., Schneider, C. & Deane, C. M. De novo antibody design with SE(3) diffusion. Preprint in arxiv (Cornell) (2024).
[10] Dauparas, J. et al. Atomic context-conditioned protein sequence design using LigandMPNN. Nature Methods vol. 22 717–723 (2025).
[11] Delgado, J., Radusky, L. G., Cianferoni, D. & Serrano, L. FoldX 5.0: working with RNA, small molecules and a new graphical interface. Bioinformatics vol. 35 4168–4169 (2019).
[12] Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. Preprint at bioRxiv (2021).
[13] Feng, S. et al. ColabDock: inverting AlphaFold structure prediction model for protein-protein docking with experimental restraints. Preprint at bioRxiv (2023).
[14] Gainza, P. et al. Deciphering interaction fingerprints from protein molecular surfaces. Preprint at bioRxiv (2019).
[15] Gowers, R. et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proceedings of the Python in Science Conference 98–105 (2016).
[16] Jiang, B., Li, X., Guo, A., Wei, M. & Wu, J. RFdiffusion Exhibits Low Success Rate in De Novo Design of Functional Protein Binders for Biochemical Detection. Preprint at bioRxiv (2025).
[17] Jiao, X. et al. Boltzmann-Aligned Inverse Folding Model as a Predictor of Mutational Effects on Protein-Protein Interactions. Preprint in arxiv (Cornell) (2024).
[18] Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature vol. 596 583–589 (2021).
[19] Kaushik, R. & Launey, T. Decoding Protein Aggregation through Computational Approach: Identification and Scoring of Aggregation-Prone Regions in Protein Sequences. Preprint at bioRxiv (2024).
[20] Kim, N., Kim, M., Ahn, S. & Park, J. Decoupled Sequence and Structure Generation for Realistic Antibody Design. Preprint in arxiv (Cornell) (2024).
[21] Kim, S., Seok, C. & Park, H. Deep learning molecular interaction motifs from Receptor structure alone. Preprint at bioRxiv (2025).
[22] Krishna, R. et al. Generalized biomolecular modeling and design with RoseTTAFold All-Atom. Science vol. 384 (2024).

[23] Lecinski, S., Howard, J. A. L., MacDonald, C. & Leake, M. C. iPAR: a new reporter for eukaryotic cytoplasmic protein aggregation. Preprint at bioRxiv (2024).
[24] Li, Q., Vlachos, E. N. & Bryant, P. Design of linear and cyclic peptide binders of different lengths from protein sequence information. Preprint at bioRxiv (2024).
[25] Li, Y. et al. HSSPPI: hierarchical and spatial-sequential modeling for PPIs prediction. Briefings in Bioinformatics vol. 26 (2025).
[26] Liang, T. et al. AI-based IsAb2.0 for antibody design. Briefings in Bioinformatics vol. 25 (2024).
[27] Lin, Z. et al. Evolutionary-scale prediction of atomic level protein structure with a language model. Preprint at bioRxiv (2022).
[28] Lindahl, E., Hess, B. & van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. Journal of Molecular Modeling vol. 7 306–317 (2001).
[29] Ma, J. et al. An adaptive autoregressive diffusion approach to design active humanized antibody and nanobody. Preprint at bioRxiv (2024).
[30] Madani, M., Lin, K. & Tarakanova, A. DSResSol: A sequence-based solubility predictor created with dilated squeeze excitation residual networks. Preprint at bioRxiv (2021).
[31] Mahdizadeh, S. J. & Eriksson, L. A. iScore: A ML-Based Scoring Function for De Novo Drug Discovery. Journal of Chemical Information and Modeling vol. 65 2759–2772 (2025).
[32] McNutt, A. et al. GNINA 1.0: Molecular Docking with Deep Learning. (2021) doi:10.26434/chemrxiv.13578140.v1.
[33] Mokhtari, O., Grudinin, S., Karami, Y. & Khakzad, H. DynamicGT: a dynamic-aware geometric transformer model to predict protein binding interfaces in flexible and disordered regions. Preprint at bioRxiv (2025).
[34] Mou, M., Zhang, Z., Pan, Z. & Zhu, F. Deep Learning for Predicting Biomolecular Binding Sites of Proteins. Research vol. 8 (2025).
[35] Pacesa, M. et al. BindCraft: one-shot design of functional protein binders. Preprint at bioRxiv (2024).
[36] Pourmirzaei, M. et al. Using Autoregressive-Transformer Model for Protein-Ligand Binding Site Prediction. Preprint at bioRxiv (2025).
[37] Pourmirzaei, M. et al. Extending Prot2Token: Aligning Protein Language Models for Unified and Diverse Protein Prediction Tasks. Preprint at bioRxiv (2025).
[38] Samantray, S. et al. Molecular dynamics simulations of protein aggregation: protocols for simulation setup and analysis with Markov state models and transition networks. Preprint at bioRxiv (2020).
[39] Scalone, E. et al. Multi-eGO: anin-silicolens to look into protein aggregation kinetics at atomic resolution. Preprint at bioRxiv (2022).
[40] Shirali, A., Stebliankin, V., Shi, J. & Narasimhan, G. A Comprehensive Survey of Scoring Functions for Protein Docking Models. Preprint at bioRxiv (2023).
[41] Sun, J., Ru, J., Cribbs, A. P. & Xiong, D. PyPropel: a Python-based tool for efficiently processing and characterising protein data. BMC Bioinformatics vol. 26 (2025).
[42] Ullanat, V., Jing, B., Sledzieski, S. & Berger, B. Learning the language of protein-protein interactions. Preprint at bioRxiv (2025).
[43] Vangone, A. & Bonvin, A. M. Contacts-based prediction of binding affinity in protein–protein complexes. eLife vol. 4 (2015).
[44] Wang, F. et al. Target-based de-novo design of cyclic peptide binders. Preprint at bioRxiv (2025).
[45] Wang, F., Wang, Y., Feng, L., Zhang, C. & Lai, L. Target-Specific De Novo Peptide Binder Design with DiffPepBuilder. Preprint in arxiv (Cornell) (2024).

[46] Wang, R. et al. IgGM: A Generative Model for Functional Antibody and Nanobody Design. Preprint at bioRxiv (2024).
[47] Wang, T. et al. Ab initio characterization of protein molecular dynamics with AI2BMD. Nature vol. 635 1019–1027 (2024).
[48] Wang, Y. et al. Large Language Model is Secretly a Protein Sequence Optimizer. Preprint in arxiv (Cornell) (2025).
[49] Wu, J. et al. FlowDesign: Improved Design of Antibody CDRs Through Flow Matching and Better Prior Distributions. Preprint at bioRxiv (2024).
[50] Yang, W. et al. Design of high-affinity binders to immune modulating receptors for cancer immunotherapy. Nature Communications vol. 16 (2025).
[51] Yao, S. et al. De novo design and directed folding of disulfide-bridged peptide heterodimers. Nature Communications vol. 13 (2022).
[52] Zhang, Z. et al. In silicodesign of stable single-domain antibodies with high affinity. Preprint at bioRxiv (2024).

**Supplementary materials**

Experimental methods

**Expression and purification**

1. Recombinant protein expression and purification

1.1. Plasmid construction

Expression vectors were prepared by restriction enzyme digestion. The specific enzymes used were chosen based on the cloning strategy for the insert of interest. Digestions were performed by incubating the vector DNA with the appropriate restriction enzymes according to the manufacturer's recommendations for 40 minutes at 37°C. Following digestion, the linearized vector DNA was purified using Qiagen DNA purification columns (QIAquick PCR Purification Kit or similar) and eluted in a final volume of 30 µl nuclease-free water. DNA concentration was quantified using a DeNovix spectrophotometer.

Gene fragments (GFs) encoding the desired protein inserts were obtained commercially. Lyophilized GFs were reconstituted in nuclease-free water and allowed to stand for 30 minutes for complete dissolution. These inserts were then digested with restriction enzymes compatible with the prepared vector ends. Digestion was carried out for 15 minutes at 37°C. The digested insert DNA was purified using Qiagen DNA purification columns and eluted in nuclease-free water, followed by concentration measurement on a spectrophotometer.

Ligation reactions were set up to insert the digested gene fragments into the corresponding digested vectors. Molar ratios of insert to vector were calculated to optimize ligation efficiency, potentially utilizing resources such as the NEB Ligation Calculator (New England Biolabs). The specific ligation conditions (e.g., enzyme: T4 DNA Ligase; temperature: room temperature or 16°C; incubation time: 10 minutes to overnight) were determined based on the calculator's recommendations and the nature of the DNA ends (blunt or cohesive). The use of optimized ratios and conditions favors the formation of the desired vector-insert construct over undesired products like self-ligated vectors.

1.2. Transformation and verification in E. coli DH5-alpha

The ligation reaction mixtures (1-5 µl, corresponding to approximately 1 pg - 100 ng of plasmid DNA) were used to transform chemically competent E. coli DH5-alpha cells. This strain is commonly employed for cloning purposes due to genetic modifications (e.g., recA1, endA1) that enhance plasmid stability by reducing recombination and endonuclease activity.

The transformation procedure commenced with thawing a vial of competent DH5-alpha cells on ice for 10 minutes. The ligation mixture was then added to the cells, and the contents

were gently mixed by flicking the tube 4-5 times, explicitly avoiding vortexing to maintain cell integrity. The cell-DNA mixture was incubated on ice for 30 minutes to allow DNA adsorption to the cell surface. Subsequently, a heat shock was applied by placing the tube in a water bath precisely at 42°C for exactly 45 seconds, facilitating DNA uptake through transient membrane permeabilization. The tube was immediately returned to ice for 5 minutes to stabilize the cell membranes. Following this, 950 µl of room temperature SOC medium was added to the cells, providing necessary nutrients for recovery and expression of the antibiotic resistance marker encoded on the plasmid. The cells were incubated at 37°C for 60 minutes with vigorous shaking (250 rpm) or rotation. Each step, particularly the precise timing and temperature of the heat shock, is crucial for achieving efficient transformation.

After the recovery period, selection plates (presumably LB agar containing the appropriate antibiotic, e.g., carbenicillin) were pre-warmed to 37°C. The cell suspension was mixed thoroughly by flicking and inversion, and 100 µl was spread onto the pre-warmed plates. The plates were incubated at 37°C for approximately 24 hours to allow the growth of colonies containing the recombinant plasmid.

1.3. Colony PCR screening

To rapidly identify clones containing the desired insert, colony PCR screening was performed on individual colonies from each transformation plate. This method bypasses the need for plasmid isolation from every potential clone. A PCR master mix was prepared for each reaction, containing Fast PCR Master Mix, forward primer, reverse primer, and nuclease-free water. The primers used were designed to flank the cloning site within the vector.

A single bacterial colony was picked using a sterile pipette tip and briefly immersed in the corresponding PCR mixture. The tip was pipetted up and down approximately 20 times within the mixture to dislodge sufficient cells to serve as the DNA template. The initial denaturation step of the PCR cycle lyses these cells, releasing the plasmid DNA. Colonies transformed with a known control plasmid were included as controls to validate the PCR conditions.

PCR amplification was carried out using a thermal cycler. The resulting PCR products were analyzed by electrophoresis on a 1% agarose gel, alongside a DNA molecular weight marker. Clones were considered positive if they yielded a PCR product of the expected size, corresponding to the vector sequence between the primer binding sites plus the inserted gene fragment.

1.4. Plasmid isolation

A single colony confirmed positive by colony PCR was selected and used to inoculate a 5 ml culture of LB broth supplemented with the appropriate antibiotic (e.g., 100 µg/ml ampicillin). The culture was incubated overnight at 37°C with shaking. Plasmid DNA was subsequently isolated from this overnight culture using a Qiagen Plasmid MiniPrep kit (or equivalent), strictly following the manufacturer's instructions.

1.5. Transformation into E. coli BL21

For recombinant protein expression, the purified plasmid DNA (1-5 µl, 1 pg - 100 ng) isolated from verified DH5-alpha clones was transformed into chemically competent E. coli BL21 cells. BL21 strains are preferred for protein expression due to deficiencies in proteases like Lon and OmpT, which minimizes degradation of the target protein.

The transformation protocol employed for BL21 cells was identical to that used for DH5-alpha cells. Briefly, cells were thawed on ice (10 min), mixed gently with plasmid DNA, incubated on ice (30 min), subjected to heat shock (42°C, 45 s), recovered on ice (5 min), incubated in SOC medium (950 µl) at 37°C for 60 minutes with shaking (250 rpm), and plated (50-100 µl) onto pre-warmed selective LB agar plates containing the appropriate antibiotic. Plates were incubated at 37°C for approximately 24 hours. The use of an identical, standardized transformation protocol for both cloning (DH5-alpha) and expression (BL21) strains streamlines the workflow.

1.6. Recombinant protein expression and induction

A single colony from the BL21 transformation plate was used to inoculate a starter culture of LB medium containing the selective antibiotic (e.g., 100 µg/ml ampicillin). This starter culture was incubated overnight at 37°C with shaking.

The following day, the overnight starter culture was used to inoculate a larger volume of LB medium supplemented with the same antibiotic. This main culture was incubated at 37°C with shaking. Cell growth was monitored by measuring the optical density at 600 nm (OD600). When the OD600 reached approximately 0.8, indicating mid-logarithmic growth phase where cells are metabolically active and optimal for protein production, expression of the target protein was induced.

Induction was achieved by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to the culture, in a final IPTG concentration of approximately 0.4 mM. IPTG induces the expression of T7 RNA polymerase from the lacUV5 promoter integrated into the BL21 genome, which in

turn drives transcription of the target gene cloned under the T7 promoter in the pET vector. Following induction, the culture flask was transferred to a shaker incubating at 16°C and incubated overnight (typically 16-20 hours). Conducting the induction at a reduced temperature (16°C) over an extended period is a common strategy employed to slow down protein synthesis. This approach can significantly enhance the likelihood of proper protein folding and increase the yield of soluble, active protein, particularly for proteins prone to aggregation or forming inclusion bodies when expressed rapidly at higher temperatures (e.g., 37°C).

1.7. Bacterial cell lysis

After overnight induction at 16°C, the bacterial cells were harvested from the culture medium by centrifugation at 6000 RPM for 20 minutes at 4°C. The supernatant (spent medium) was carefully discarded.

The resulting cell pellet was resuspended in an appropriate lysis buffer with protease inhibitors. The resuspended cell mixture was incubated on ice for 10 minutes.

Cell disruption was achieved by sonication. The sample was kept in an ice bath throughout the process to prevent overheating and potential denaturation of the target protein. The crude lysate was then clarified to remove insoluble cell debris, inclusion bodies, and other particulate matter. This was achieved by centrifugation at 10,000 x g for 20 minutes at 4°C. The resulting supernatant, containing the soluble protein fraction, was carefully collected and filtered through a 0.45 µm syringe filter into clean tubes. Filtration ensures the removal of any remaining fine particulates that could clog chromatography columns.

1.8. Protein purification by chromatography

A sequential two-step chromatography strategy, employing Immobilized Metal Affinity Chromatography (IMAC) followed by Size Exclusion Chromatography (SEC), was used to purify the target recombinant protein (His-tagged, as IMAC is utilized). This combination leverages the high selectivity of IMAC for the initial capture and enrichment, followed by SEC as a polishing step to remove remaining contaminants and protein aggregates, achieving high purity.

For IMAC, a standard automated chromatography system protocol was followed. The specific IMAC column was equilibrated with an equilibration buffer. The filtered cell lysate was loaded onto the column using a sample pump, with the tubing properly primed beforehand. A predefined method was executed, typically involving washing the column with

binding buffer to remove non-specifically bound proteins, followed by elution of the His-tagged target protein using an elution buffer containing a higher concentration of imidazole. Fractions were collected throughout the run, typically based on UV absorbance (A 280). After the run, rigorous column and system cleaning procedures were performed.

Fractions from the IMAC elution containing the target protein (identified typically by SDS-PAGE analysis) were pooled. If necessary, the pooled sample might be concentrated using methods like ultrafiltration. This sample was then subjected to SEC. The SEC procedure also followed standard chromatography system operation: setting pressure limits appropriate for the SEC column, priming the system, and thoroughly equilibrating the column with the final desired buffer. The sample was injected into a sample loop of known volume and then loaded onto the column. Proteins were eluted isocratically with the SEC running buffer at a defined flow rate. Separation occurs based on the hydrodynamic volume of the molecules, with larger molecules eluting earlier. Fractions corresponding to the expected elution volume/peak for the monomeric target protein were collected based on UV absorbance (A280). The column and system were cleaned post-run.

1.9. Protein analysis by SDS-PAGE

Sodium Dodecyl Sulfate - Polyacrylamide Gel Electrophoresis (SDS-PAGE) was used to monitor protein expression and assess purity throughout the purification process (e.g., analysis of crude lysate, flow-through, wash fractions, IMAC eluate, and SEC fractions).

Samples for analysis were prepared by mixing a volume of the test sample with an appropriate volume of concentrated SDS-PAGE sample loading buffer. The mixture was vortexed briefly and then heated at 95°C for 5 minutes to ensure complete protein denaturation and reduction of disulfide bonds.

Electrophoresis was performed using precast polyacrylamide gels (e.g., 4-12% Bis-Tris gradient gels, offering good resolution over a broad molecular weight range). Prepared samples and a protein molecular weight marker were loaded into the wells.

The electrophoresis apparatus lid was secured, and power was applied using a constant voltage setting of 80V. Electrophoresis was continued until the bromophenol blue dye front migrated to near the bottom of the gel, with migration monitored relative to the prestained marker proteins. Running gels at a relatively low constant voltage (80V) typically results in slower migration but often yields sharper bands and better resolution by minimizing thermal

effects that can cause band distortion ("smiling"), which may be particularly beneficial for gradient gels or when resolving proteins of similar sizes.

Following electrophoresis, the gel was removed from the cassette, and proteins were visualized by staining with Coomassie Brilliant Blue R-250 using a standard staining and destaining protocol.

**Applications**

a. Immunofluorescence

Cells cultured on coverslips were prepared for immunofluorescence analysis according to standard procedures. First, the culture medium was removed, and cells were gently washed twice with Phosphate-Buffered Saline (PBS). Fixation was performed by incubating the cells with 4% paraformaldehyde (PFA) in PBS for 15-30 minutes at room temperature (RT). Following fixation, the PFA solution was removed, and the cells were washed three times with PBS.

For the detection of intracellular targets, cells were permeabilized by incubation with 0.1% Triton X-100 in PBS for 10-30 minutes at RT. After permeabilization, cells were washed three times with PBS. This step was omitted for surface protein detection.

To minimize non-specific antibody binding, cells were incubated in a blocking solution consisting of either 1% Bovine Serum Albumin (BSA) or 5% normal serum in PBS for 30-60 minutes at RT.

The primary antibody conjugated with fluorophore was diluted in PBS containing 1% BSA to a final concentration of 5 µg/ml. Cells were incubated with the diluted primary antibody solution overnight (ON) at 4°C. After primary antibody incubation, cells were washed three times with PBS, with each wash lasting 5 minutes, to remove unbound antibodies.

b. Flow cytometry

Cells grown in culture were prepared for flow cytometry analysis. Adherent cells were detached by first removing the culture medium, washing once with PBS, and then incubating with Trypsin-EDTA solution at 37°C until cells detached. The cell suspension was collected, and trypsin activity was neutralized. Cells were pelleted by centrifugation at 1500 RPM for 5 minutes.

The supernatant was discarded, and the cell pellet was washed with cold PBS. Cells were counted, and the concentration was adjusted to 1 x 10^6 cells/ml in cold PBS.

Cells were fixed by adding 4% PFA in PBS and incubating for 20 minutes at RT. After fixation, cells were pelleted by centrifugation at 500 x g for 5 minutes and washed once with PBS. For intracellular targets, permeabilization was performed by incubating cells in 0.2% Triton X-100 in PBS for 15 minutes at RT, followed by a wash with PBS. This step was omitted for surface staining. Non-specific binding sites were blocked by incubating the cells in 1% BSA in PBS for 30 minutes at RT.

Cells were stained by incubating with the primary antibody conjugated with fluorophore, typically at a concentration of approximately 2 µg per million cells, diluted in PBS (or PBS + 1% BSA), typically 60 minutes at RT. After incubation, cells were washed to remove unbound antibodies.

Stained cells were analyzed using a flow cytometer equipped with appropriate lasers and filters for the fluorophore conjugated to the antibody. Appropriate instrument settings (voltages, compensation) were applied. Controls included unstained cells (to assess background fluorescence) and cells stained with an isotype control antibody (matched to the primary antibody's species, isotype, and fluorophore, used at the same concentration) to control for non-specific antibody binding.

c. Immunohistochemistry (IHC)

To visualize specific proteins within tissue sections, immunohistochemistry was performed on formalin-fixed, paraffin-embedded samples. Sections were first deparaffinized using xylene and rehydrated through a graded series of ethanol washes. Antigen retrieval was carried out using heat-induced epitope retrieval (HIER) in an appropriate buffer (e.g., citrate buffer, pH 6.0) to unmask target epitopes. Endogenous peroxidase activity was quenched by incubation with hydrogen peroxide. Sections were then incubated with a specific primary antibody, followed by incubation with an HRP-conjugated secondary antibody for detection. Subsequent steps involving chromogen development (e.g., DAB) and counterstaining allowed for microscopic visualization of the target protein's location.

**Binding affinity - BLI**

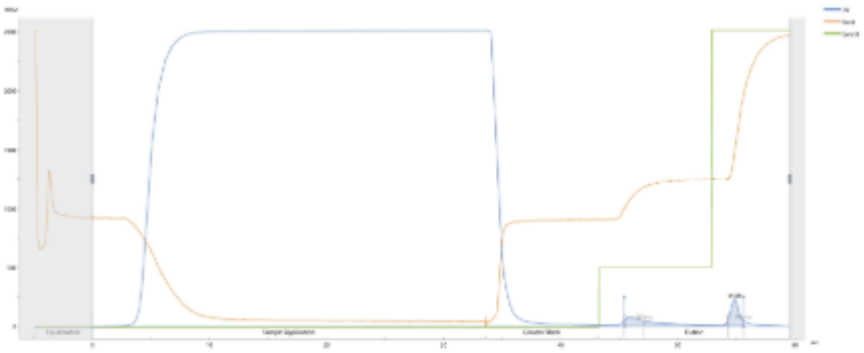
Binding kinetics and affinity were assessed using Bio-Layer Interferometry (BLI) on a suitable instrument platform Sartorius Octet. Appropriate biosensors (e.g., Protein A sensors for capturing IgG antibodies) were first hydrated in a kinetic buffer for at least 10 minutes.

The ligand (e.g., antibody) was prepared at an experimentally optimized concentration in the kinetic buffer. The analyte (e.g., antigen) was prepared at a series of concentrations, typically centered around the estimated equilibrium dissociation constant (KD), also diluted in kinetic buffer.

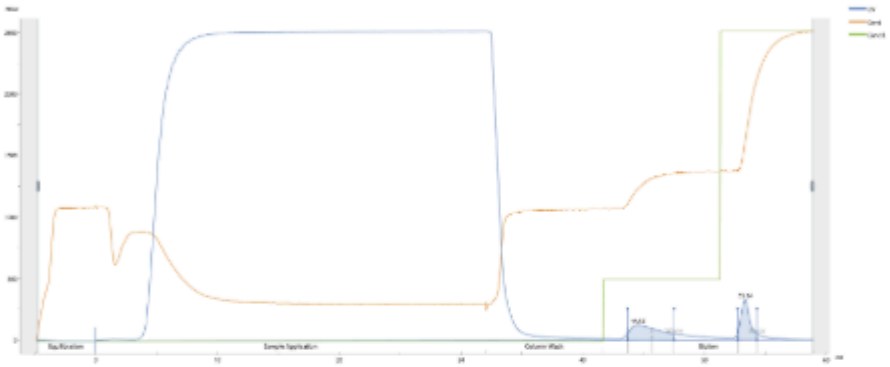
A standard kinetic assay format was employed within a microplate. This typically involved the following steps: (1) Sensor equilibration in kinetic buffer to establish an initial baseline (e.g., 60 seconds). (2) Loading of the ligand (antibody) onto the hydrated sensors (immobilization step, e.g., 120 seconds). (3) A second baseline step in the kinetic buffer to stabilize the signal after loading (e.g., 60 seconds). (4) Association step, where the ligand-loaded sensors were immersed in wells containing the analyte (antigen) at various concentrations (e.g., 240 seconds). (5) Dissociation step, where the sensors were transferred back into wells containing only a kinetic buffer to monitor the dissociation of the analyte (e.g., 240 seconds). The specific durations for each step were optimized depending on the interaction characteristics of the proteins being studied.

Raw sensorgram data obtained from the instrument were processed and analyzed using the dedicated Sartorius analysis software. Kinetic parameters, including association rate constants (kon), dissociation rate constants (koff), and the equilibrium dissociation constant (KD), were determined by fitting the experimental data to appropriate binding models (e.g., 1:1 Langmuir binding model).

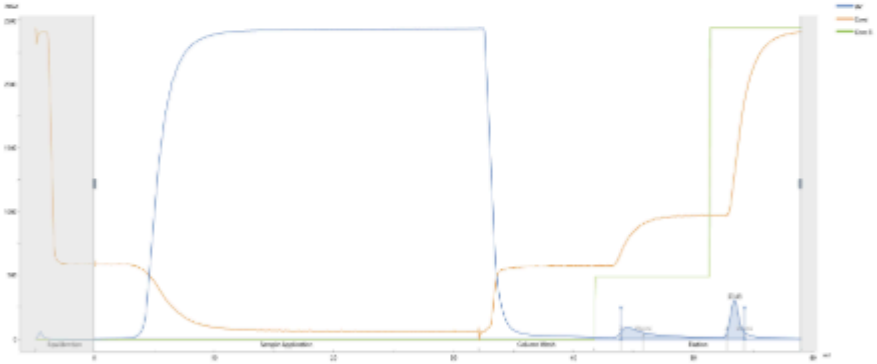
A



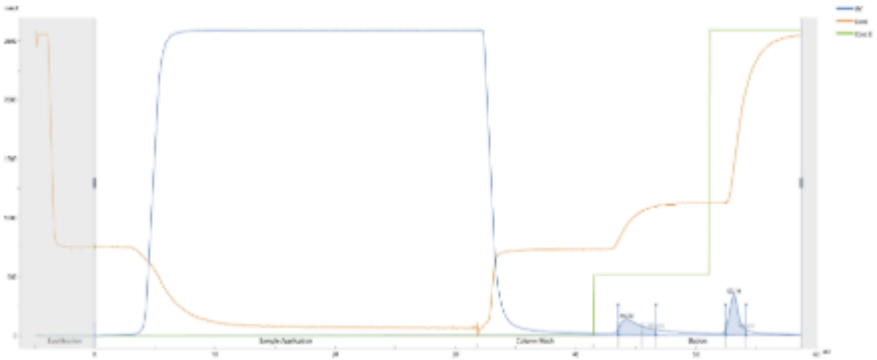
B



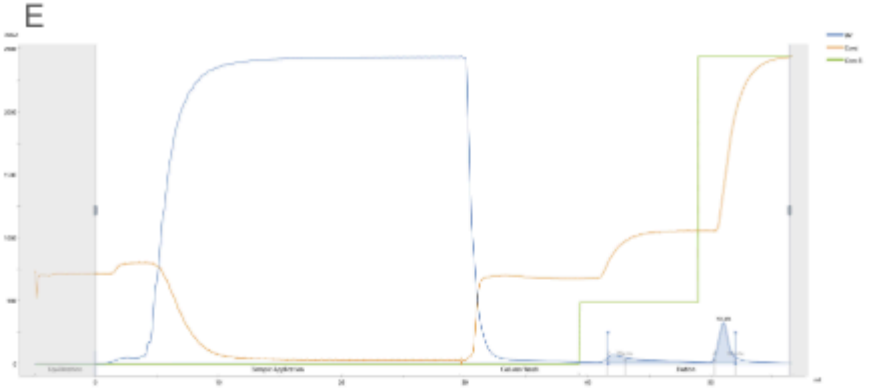
C



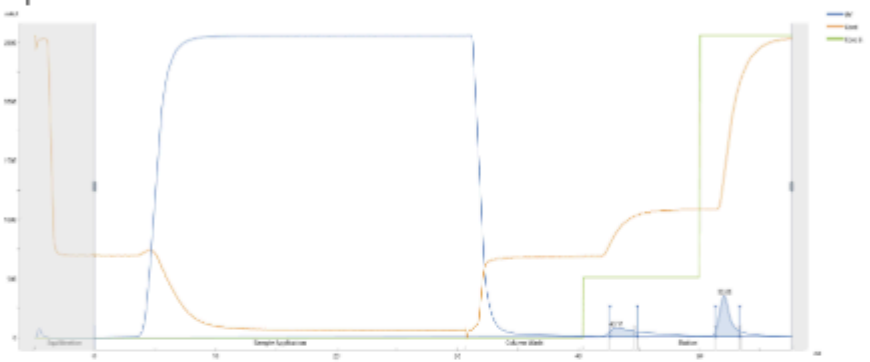
D



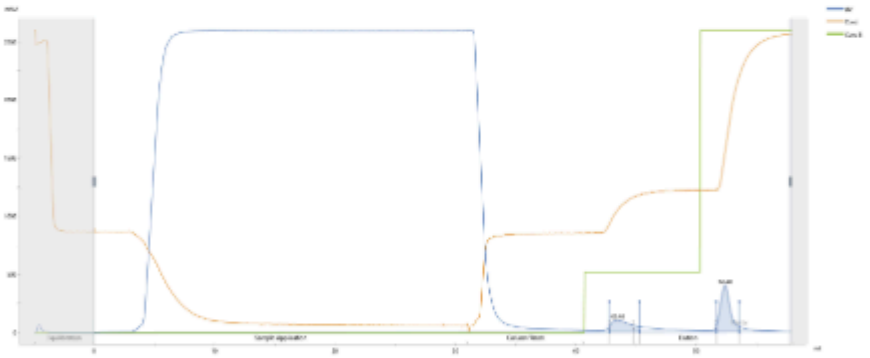
E



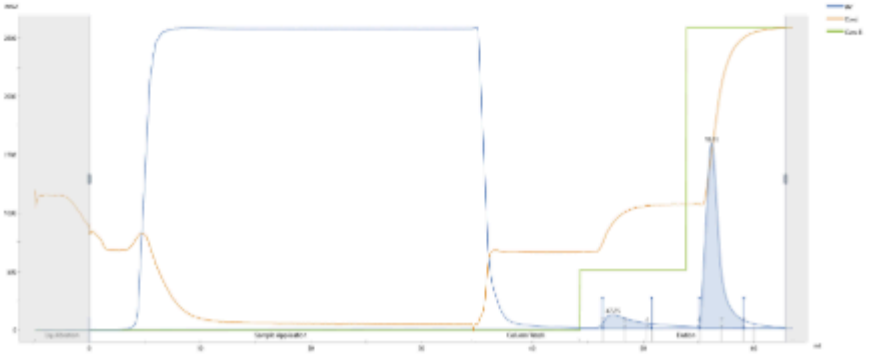
F



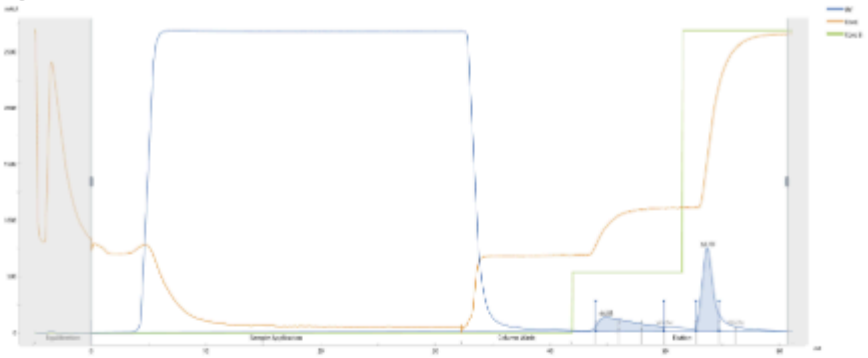
G



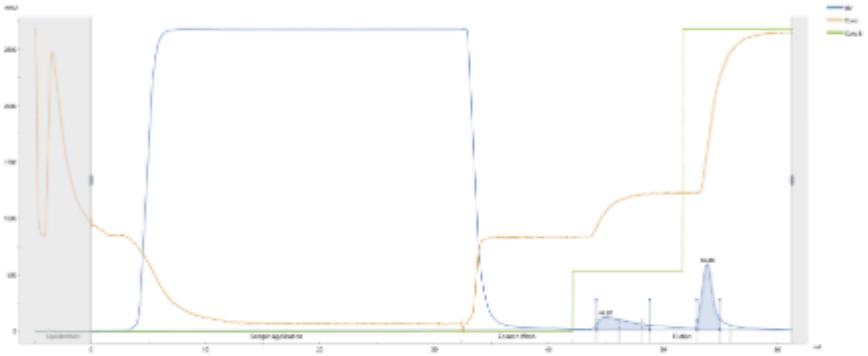
H



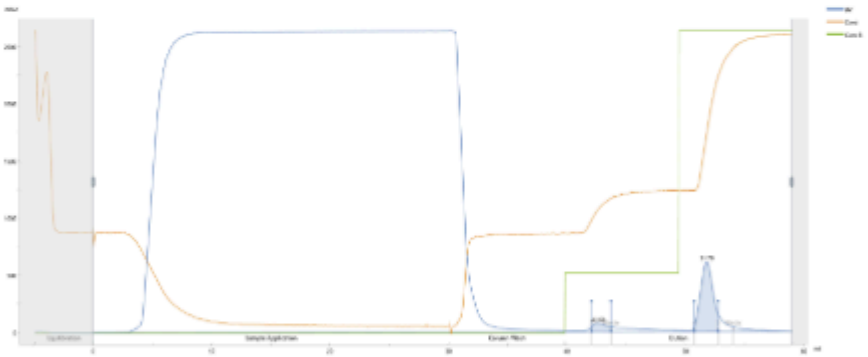
I



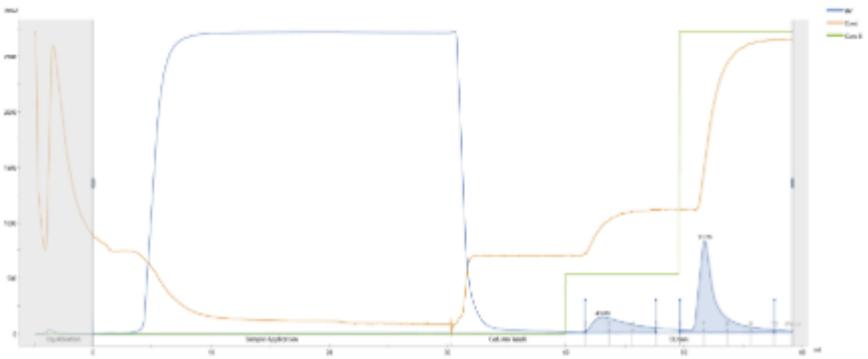
J



K



L



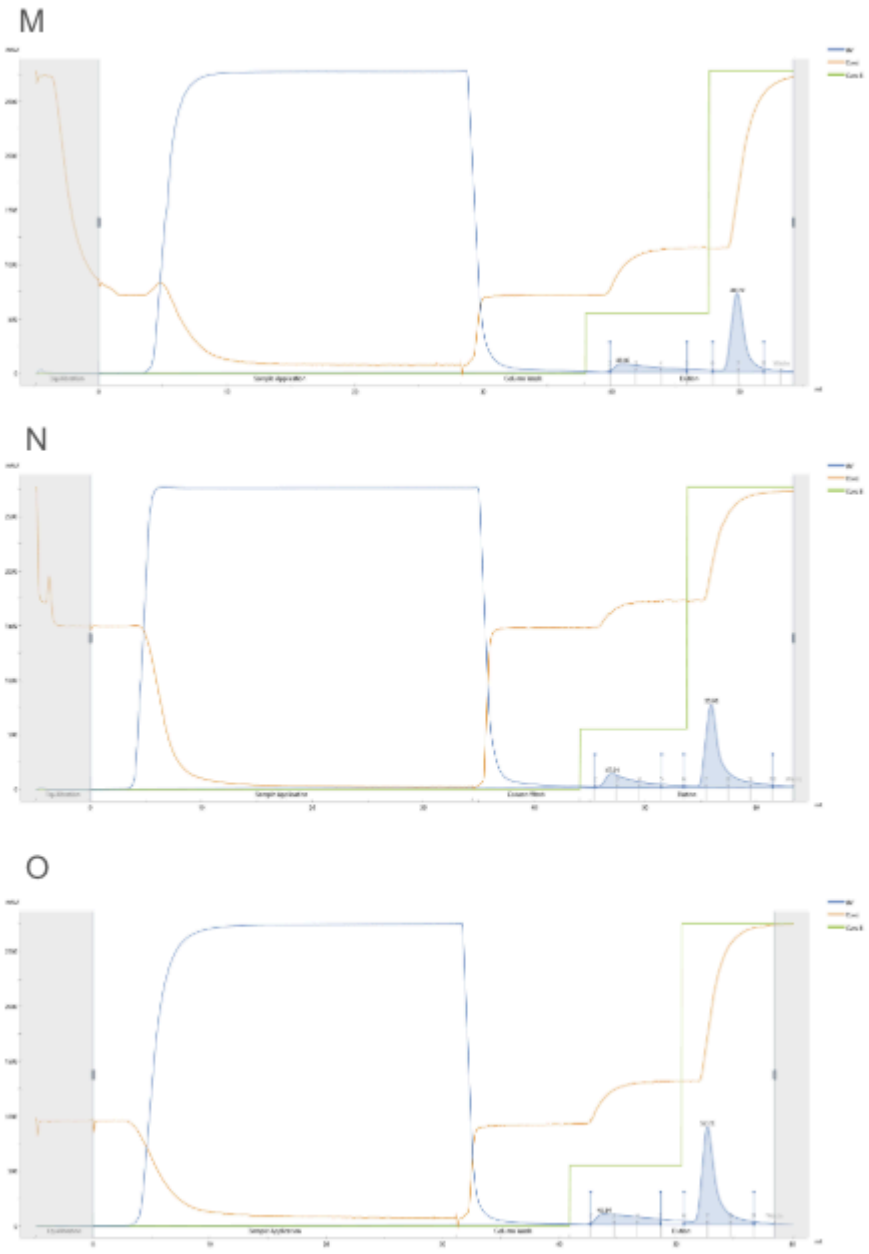



**Figure S1**. Full IMAC chromatograms of fifteen recombinant antibodies. Shown are complete chromatographic profiles, including flow-through, wash, and elution phases, for (A) anti-GSK3A, (B) anti-SKP2, (C) anti-Survivin, (D) anti-IDO1, (E) anti-Mitogen-activated protein kinase 12, (F) anti-BRD4, (G) anti-Wilms tumor protein (WT33), (H) anti-βIII Tubulin, (I) anti-Ki67, (J) anti-ERK1, (K) anti-TSG101, (L) anti-OCT4, (M) anti-TGM2, (N) anti-FGF2, and (O) anti-RFP antibodies. The elution phase typically displays two peaks: an early, non-specific peak (~20% elution buffer) and a later peak (~100% elution buffer) corresponding to the target antibody.

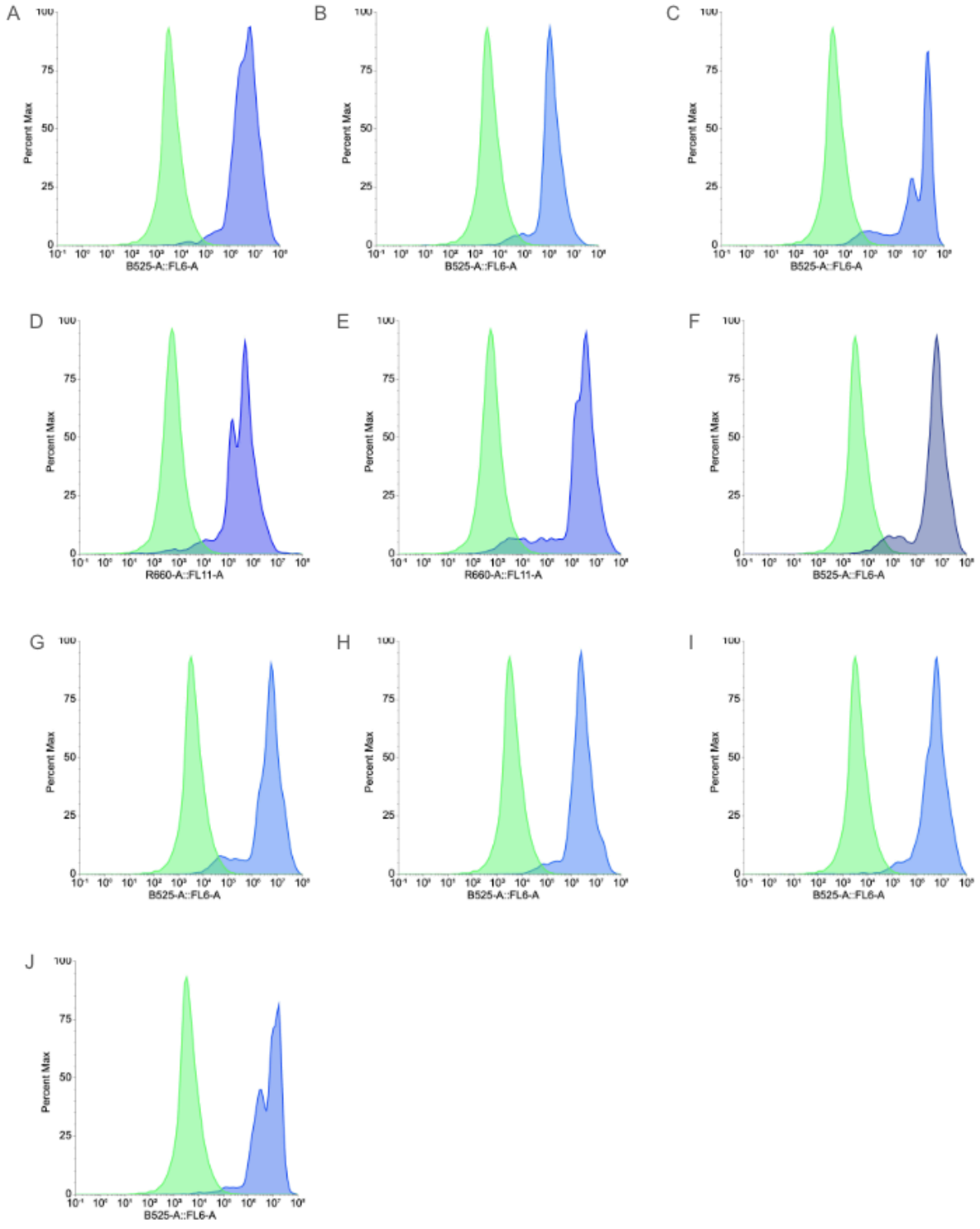



**Figure S2.** Flow cytometry analysis of antibody staining in HeLa cells. Histograms show fluorescence intensity profiles for unstained control cells (green) and cells stained with antibodies against (A) CAMK1, (B) FGF2, (C) Beta-III-tubulin, (D) Beta-tubulin, (E) Ki67, (F) PCNA, (G) p62, (H) SMAD3, (I) Cyclin D1, (J) Nrl2 (blue). Each plot displays a clear shift in fluorescence signal relative to the control, indicating specific binding of the antibody to its target antigen.

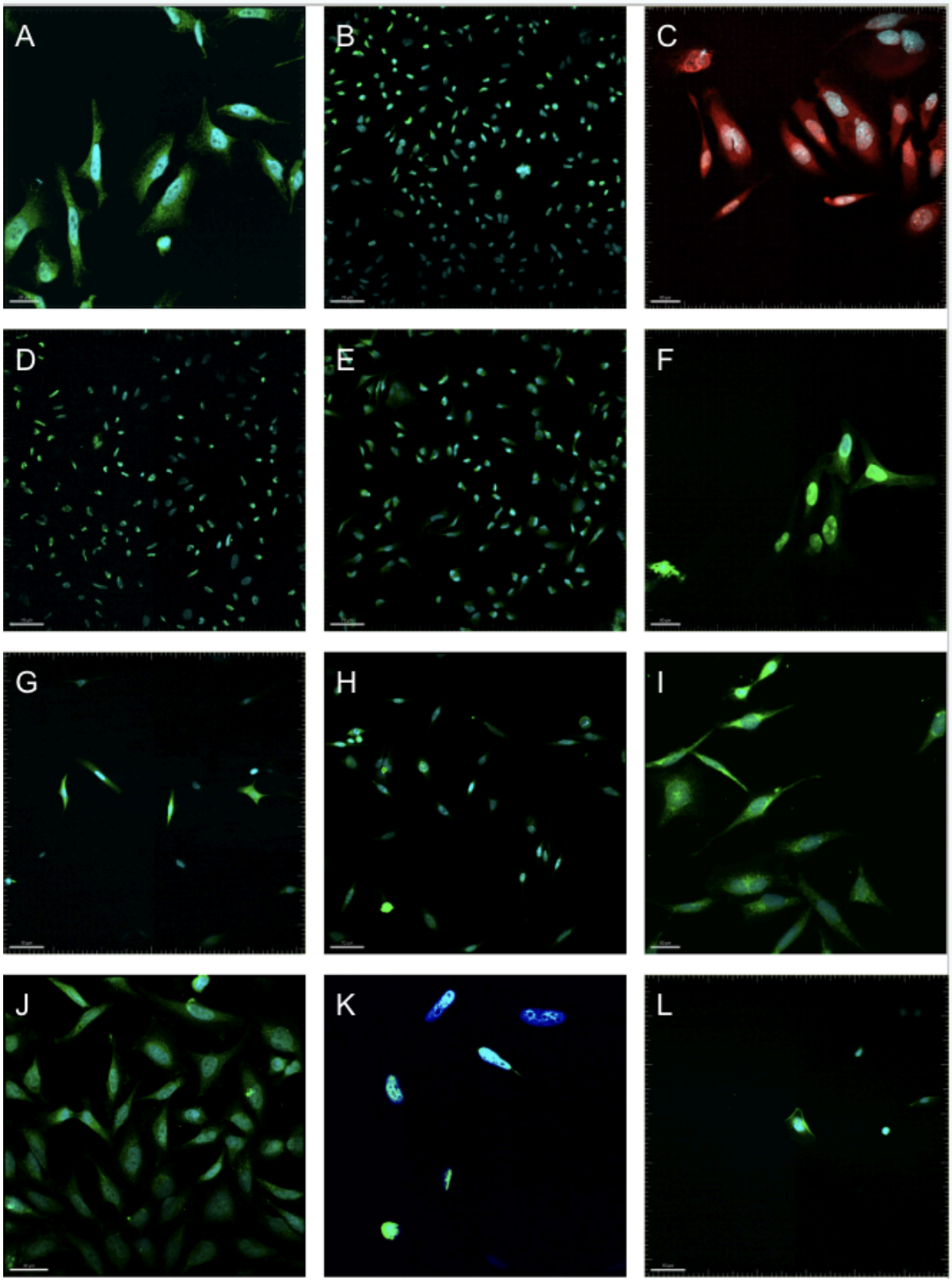



**Figure S3.** Immunofluorescence analysis of protein expression in HeLa cells.
Representative confocal images of HeLa cells subjected to immunofluorescence staining using primary antibodies. a, b, d–l, Staining using antibodies conjugated to Alexa Fluor 488 (green) against: a, GFAP; b, Histone H3; d, POU5F1; e, WT1; f, TUBB3; g, EIF2C2; h, FGF10; i, FGF2; j, GAPDH; k, MKI67; l, LAMP1. c, Staining for ACTA2 using an antibody conjugated to Alexa Fluor 594 (red). Cell nuclei were stained with DAPI (blue). A scale bar depending on the magnification was applied to all panels.

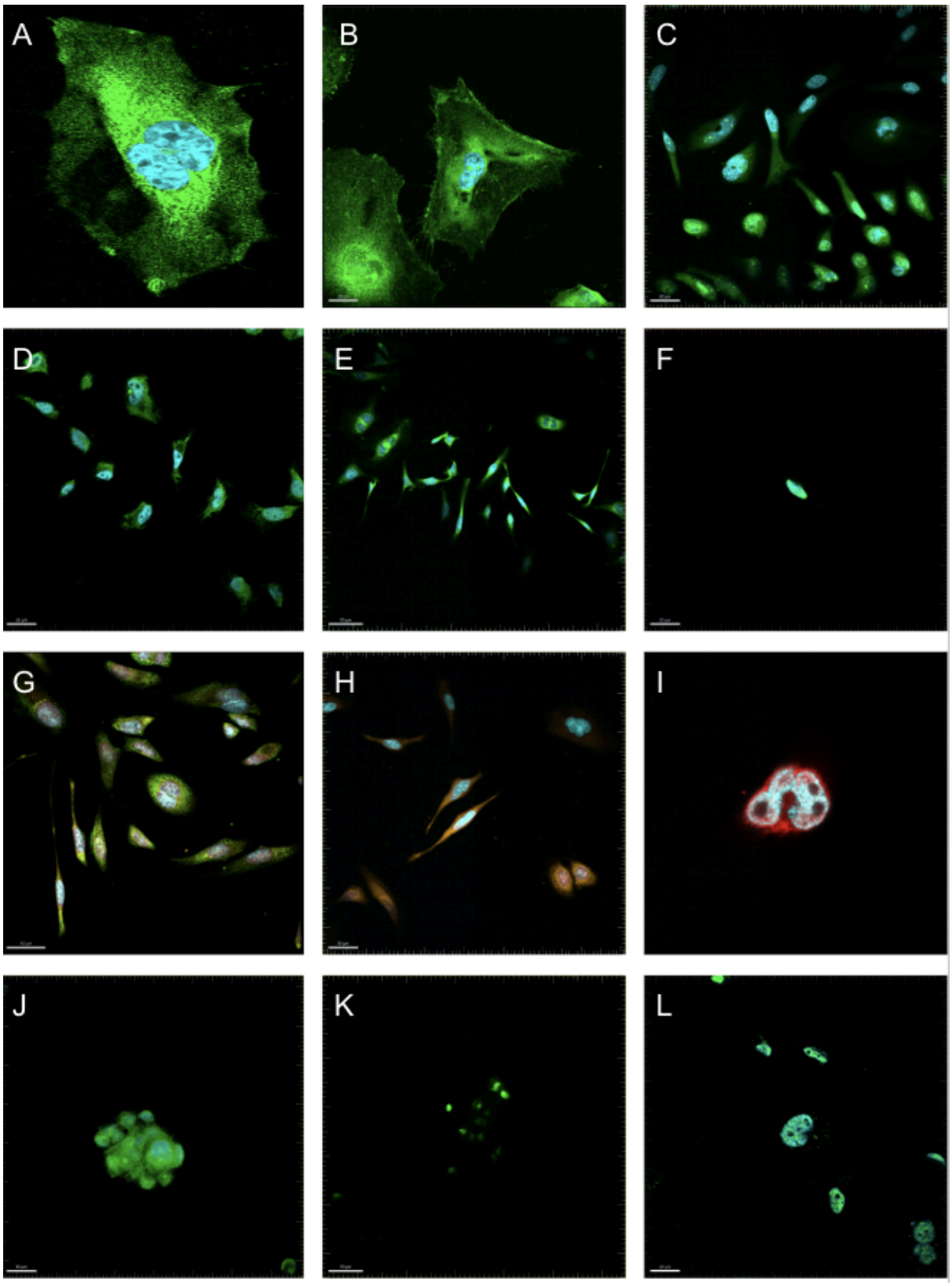



**Figure S4**. Immunofluorescence analysis of protein expression in HeLa cells. Representative confocal images of HeLa cells subjected to immunofluorescence staining using primary antibodies. a–l, Staining using antibodies conjugated to Alexa Fluor 488 (green) against: a, NFE2L2 ; b, SMAD3 ; c, CAMK2B ; d, PECAM1 ; e, MAPK1 ; f, PCNA ; g, STAT3 ; h, VIM ; i, CCND1 ; j, ALDOA ; k, BDNF ; l, RUNX2. Cell nuclei were stained with DAPI (blue). A scale bar depending on the magnification was applied to all panels.

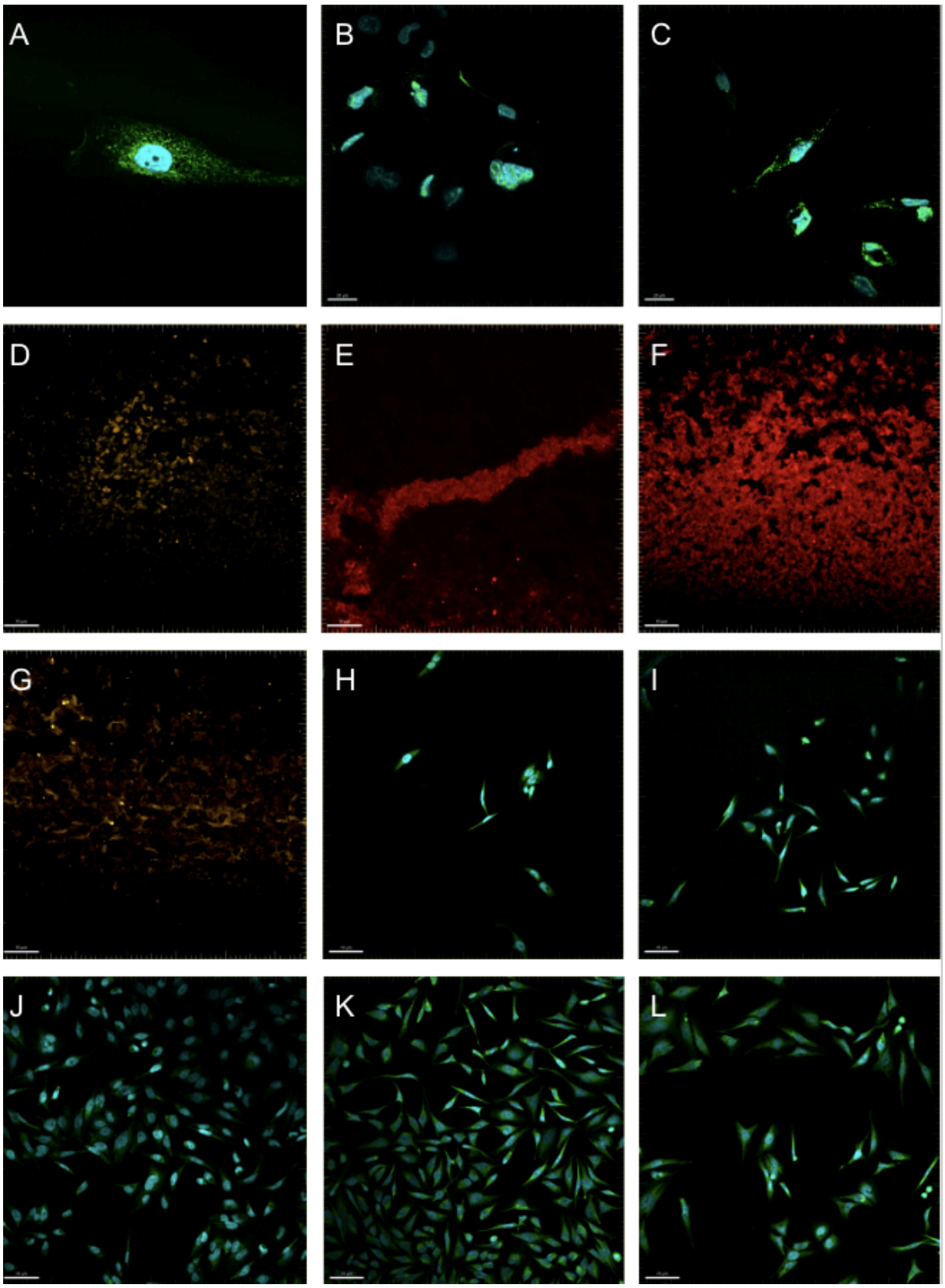
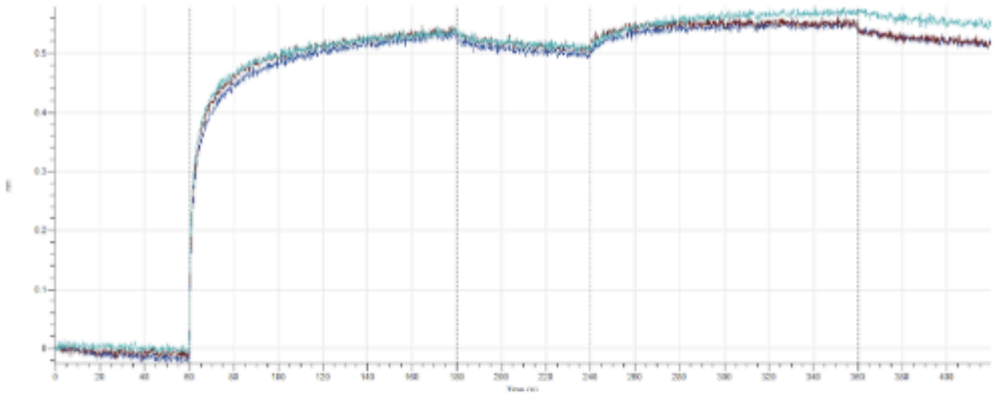
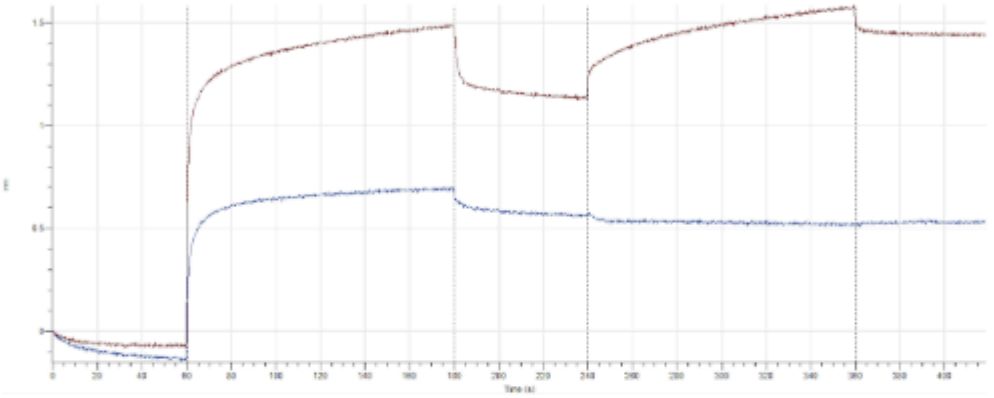



**Figure S5.** Immunofluorescence analysis of protein expression in HeLa cells and MDA-231 xenografts. Representative confocal images. a–c, h–l, HeLa cells stained using antibodies conjugated to Alexa Fluor 488 (green) against: a, BRCA1 ; b, NFKBIA ; c, SQSTM1 ; h, SHH ; i, GLB1 ; j, MYC ; k, BAX ; l, MAP1LC3B. d–g, MDA-231 xenograft tumor tissue sections stained using antibodies against: d, MKI67 (Alexa Fluor 647, orange) ; e, Laminin (Alexa Fluor 647, red) ; f, β-Tubulin (Alexa Fluor 594, red) ; g, VIM (Alexa Fluor 594, red). Cell nuclei were stained with DAPI (blue). A scale bar depending on the magnification was applied to all panels).

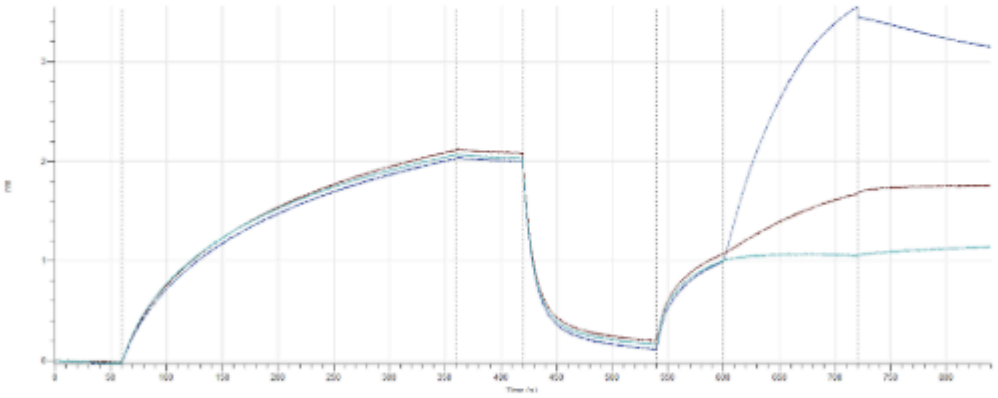
A



B



C



D

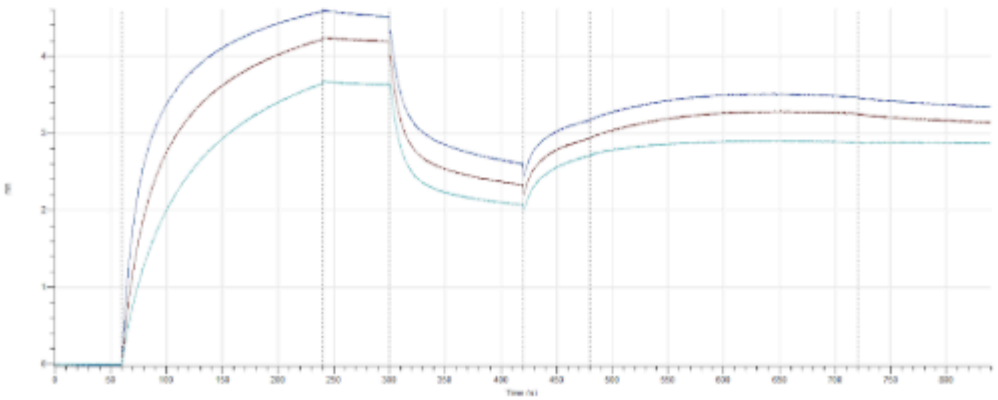


**Figure S6.** Raw biolayer interferometry (BLI) sensorgrams for antibody–antigen binding analysis. Shown are unprocessed sensorgrams for (A) anti-GAPDH, (B) anti-CAMK1, (C) anti-FGF2, and (D) anti-FGF10 antibodies, including all experimental phases: initial baseline, antibody loading, second baseline, association, and dissociation.