Liu *et al. Journal of Nanobiotechnology* (2025) 23:87
https://doi.org/10.1186/s12951-025-03169-5

Journal of Nanobiotechnology

REVIEW Open Access

Unveiling the new chapter in nanobody engineering: advances in traditional construction and AI-driven optimization

Jiwei Liu[1,2†], Lei Wu[1,2†], Anqi Xie[1†], Weici Liu[1,2], Zhao He[1,2], Yuan Wan[3,5*] and Wenjun Mao[1,2,4*]

Abstract

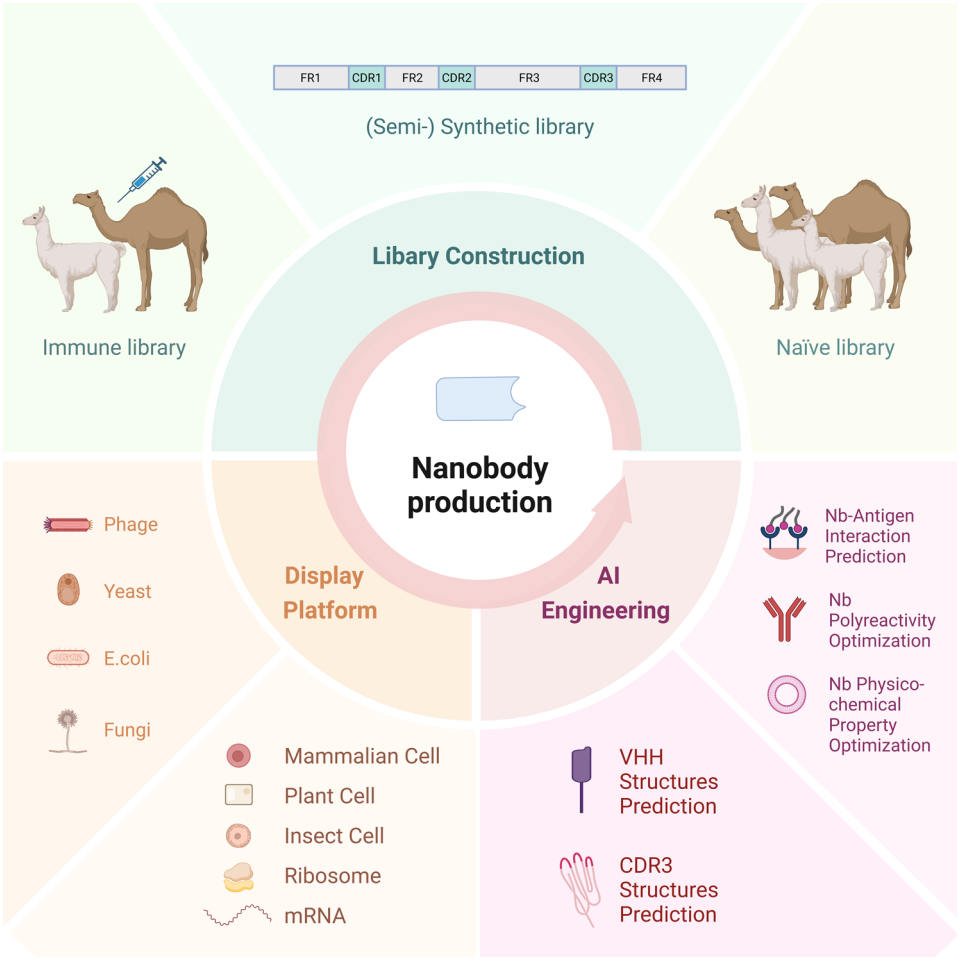
Nanobodies (Nbs), miniature antibodies consisting solely of the variable region of heavy chains, exhibit unique properties such as small size, high stability, and strong specificity, making them highly promising for disease diagnosis and treatment. The engineering production of Nbs has evolved into a mature process, involving library construction, screening, and expression purification. Different library types, including immune, naïve, and synthetic/semi-synthetic libraries, offer diverse options for various applications, while display platforms like phage display, cell surface display, and non-surface display provide efficient screening of target Nbs. Recent advancements in artificial intelligence (AI) have opened new avenues in Nb engineering. AI's exceptional performance in protein structure prediction and molecular interaction simulation has introduced novel perspectives and tools for Nb design and optimization. Integrating AI with traditional experimental methods is anticipated to enhance the efficiency and precision of Nb development, expediting the transition from basic research to clinical applications. This review comprehensively examines the latest progress in Nb engineering, emphasizing library construction strategies, display platform technologies, and AI applications. It evaluates the strengths and weaknesses of various libraries and display platforms and explores the potential and challenges of AI in predicting Nb structure, antigen-antibody interactions, and optimizing physicochemical properties.

†Jiwei Liu, Lei Wu and Anqi Xie contributed equally to this work.

*Correspondence:
Yuan Wan
ywan@binghamton.edu
Wenjun Mao
maowenjun1@njmu.edu.cn

Full list of author information is available at the end of the article

Graphical abstract



Keywords Nanobody engineering, Library construction, Screening platforms, Artificial intelligence, Design optimization

Introduction

Since their discovery, nanobodies (Nbs) have attracted extensive research interest due to their unique molecular structure and excellent physicochemical properties, demonstrating great potential for diverse applications such as disease diagnosis and treatment. Derived from camelid heavy chain-only antibodies (HCAbs), Nbs comprise only the variable heavy chain domain (VHH) and the constant CH2 and CH3 domains [1]. Interestingly, HCAbs have also been discovered in some cartilaginous fish, such as sharks, known as novel antigen receptors (NARs), which recognize antigens via a single variable domain (V-NAR) [2]. However, significant differences in sequence and structure between V-NARs and camelid VHHs or the variable domains of human heavy chains render V-NARs less frequently discussed [3]. Figure 1 illustrates the structural differences among these antibodies.

VHHs consist of three complementarity-determining regions (CDRs) and four framework regions (FRs). Structural analyses reveal that VHH crystals, with dimensions of 2.5 nm × 4 nm and a molecular mass of approximately 15 kDa, are the smallest known antigen-binding units [4, 5]. Despite sharing a scaffold formed by two β-sheets with the VH of conventional antibodies, VHHs exhibit distinctive features [6]. Variations in the length of the hypervariable regions enable VHHs to possess enhanced antigen-binding capabilities despite their smaller size. Furthermore, disulfide bonds in CDR1 and CDR3 increase the stability of VHHs [7]. Additionally, VHHs' CDR3 can form a finger-like convex structure, enabling it to bind to epitopes that are cryptic to conventional antibodies, for instance, deep clefts [8]. Substitution of FR2 hydrophobic residues with smaller and/or hydrophilic amino acids enhances the solubility of Nbs [7].

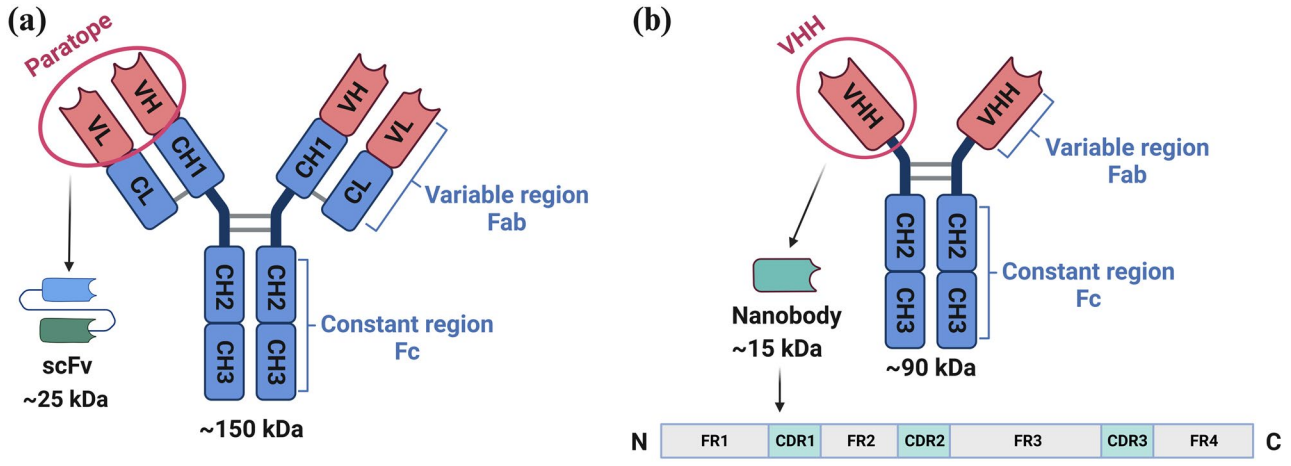


Fig. 1 Diagrams of the conventional heterotetramer antibody, the heavy chain-only antibody (HCAb), and the VHH fragment (i.e., nanobody). (**a**) The conventional heterotetrameric antibody (IgG) is shown schematically. The antigen binding fragment (Fab) comprises the complete light chain (VL and CL) and the partial heavy chain structure (VH and CH1), with the VH and VL domains forming the paratope. Such domains of IgGs can be derivatized to scFv fragments. (**b**) HCAbs are devoid of the VL and the CH1 domain, binding to the antigens through the single VHH fragment [1]. The VHH fragment contains four framework regions (FR1-4, grey) and three complementarity-determining regions (CDR1-3, green) [10]

Comparison of the amino acid sequences of VHHs with human VH sequences reveals high similarity, which simplifies the humanization of Nbs [6, 9]. Apart from these structural advantages, Nbs can be easily expressed in microorganisms, reducing production costs and facilitating widespread application [10].

The unique properties of Nbs enable their extensive applications as research or diagnostic tools, therapeutic agents. Nbs conjugated to fluorescent proteins, biosensors, or enzymes can investigate complex intracellular signaling pathways [11], while a novel human SIRPα Nb has been developed for in vivo imaging of myeloid cells, allowing the visualization of myeloid cell tumor infiltration [12]. Notably, the fast targeting, quick extravasation, rapid blood clearance, and predominantly renal excretion of Nbs make them ideal for imaging [13]. Moreover, Nbs have shown progress in treating viral infections and tumors [14–16]. As for the diagnosis and treatment of central nervous system diseases, Nbs can cross the blood-brain barrier, such as by adsorptive-mediated transcytosis, which is difficult or impossible for larger antibody fragments [17].

The advent of Nbs has revolutionized antibody engineering, providing a novel approach to target various biomolecules with unprecedented precision and affinity, and has garnered widespread attention. Structural biology has played a crucial role in the rational design of Nbs, enabling researchers to visualize the atomic interactions between Nbs and their targets. This facilitates the precise design of Nbs to improve their binding affinity, specificity, and stability [18]. Moreover, computational approaches, such as molecular dynamics simulations and machine learning algorithms, are increasingly used to predict and optimize the interactions between Nbs and their targets, expediting the design process [19].

To comprehensively understand the latest advancements in Nbs engineering, this review thoroughly compiles and discusses recent research findings related to Nbs library construction, retrieval, and production technologies. Given the rise and rapid development of artificial intelligence (AI), integrating AI with Nbs engineering has shown immense potential and significance, which will be discussed in detail in the following sections. This work will provide a clear, up-to-date understanding of the current state of Nbs engineering, which will ultimately lead to the creation of novel Nbs with enhanced properties and expanded utility in biomedicine, biotechnology, diagnostics.

The construction of Nbs libraries

Accessing libraries containing the target genomic information is crucial for generating Nbs with high specificity and affinity. Currently, three major libraries are used: immune, naïve, and synthetic/semi-synthetic [9, 10]. Figure 2 illustrates the detailed steps to build the three Nbs libraries. The generation of immune and naïve libraries requires the involvement of animals, with the immune library serving as the primary source of Nbs due to the availability of affinity-matured antibodies. However, as an attractive alternative to animal use, synthetic/semi-synthetic libraries have garnered significant attention. Table 1 outlines the differences among these three library types.

Immune library

Immune libraries are typically generated by immunizing camels (e.g., Bactrian camel, dromedary, alpaca, or llama)

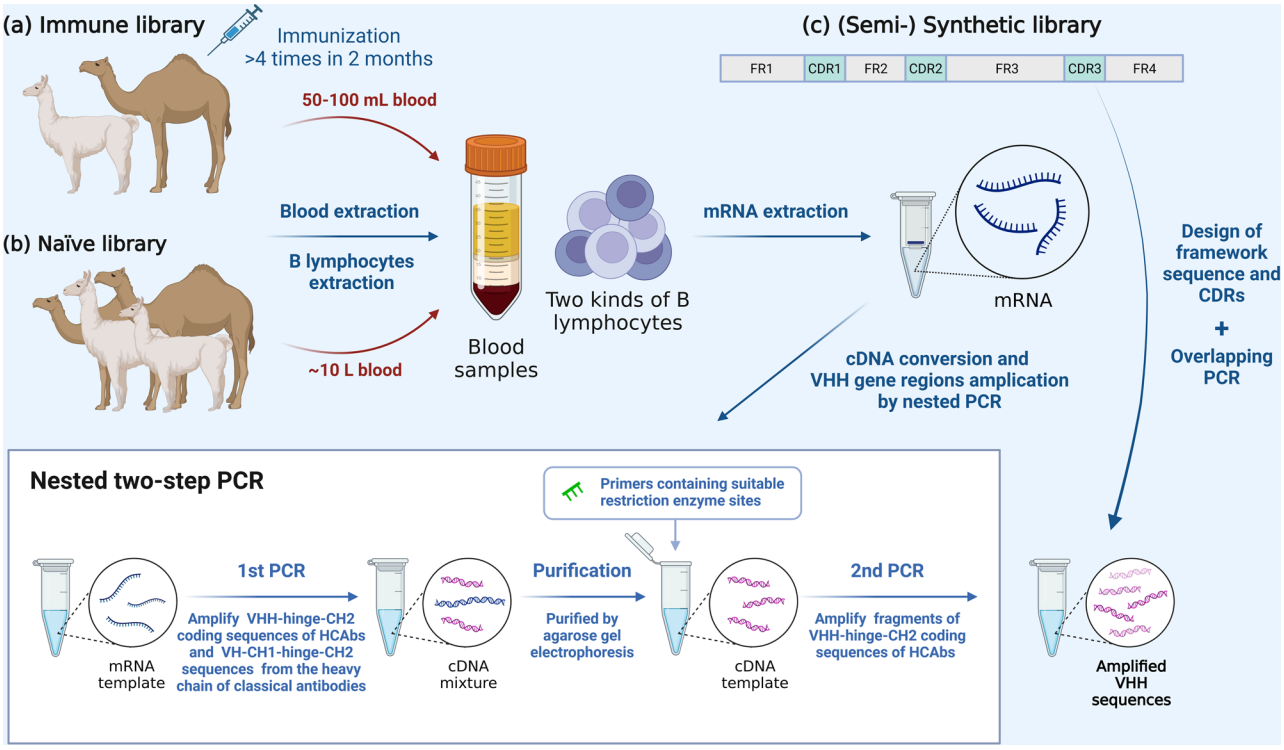


Fig. 2 General workflow of the construction of immune, naïve and (semi-) synthetic libraries. (**a**) The immune library requires immunization of camelids over 4 times in 2 months. After sampling 50–100 ml of blood and B lymphocytes isolation, mRNA is extracted [20, 21]. Following mRNA conversion, cDNA is used to amplify the VHH gene regions [23], ideally through a nested two-step PCR. The first PCR uses the mRNA template to amplify the VHH-hinge-CH2 coding sequences of HCAbs and the VH-CH1-hinge-CH2 sequences from the heavy chain of classical antibodies, yielding the cDNA mixture. The mixture was then purified by agarose gel electrophoresis to obtain the target cDNA template. Primers containing suitable restriction enzyme sites are then utilized to amplify the fragments of the VHH-hinge-CH2 coding sequences of the HCAbs, resulting in amplified VHH sequences [9]. (**b**) The naïve library is constructed without the process of animal immunization, but instead require the collection of more than 10 L of blood from 10 to 20 camelids [29]. The subsequent steps are similar to those of the immune library. (**c**) Unlike the previous two library constructions, (semi-) synthetic libraries do not require animal participation and blood extraction. After the design of framework sequence and CDRs, it is frequently performed overlapping PCR to obtain amplified VHH sequences [37]

Table 1 Comparative properties of three types of libraries

	Whether animal immunization is required	Properties of Nbs	Blood requirements	Library size	Library universality	Time span	Ref.
Immune library	Yes	Affinity-matured Target-specific	50–100 mL	10^6–10^8	Multiple libraries will be required for various projects	Relatively long Several months	[9, 21]
Naïve library	No	Works for non-immunogenic targets	More than 10 L	10^9–10^{11}	One library can be utilized for multiple projects	Relatively short Weeks to months	[29]
(Semi-) Synthetic library	No	Works for non-immunogenic targets	No demand	10^9–10^{15}	One library can be utilized for multiple projects	Rather short Weeks to months	[37]

with target immunogens, usually administered through subcutaneous injections containing 1 to 5 antigens combined with an adjuvant [20, 21]. In most cases, camels are immunized five times at weekly intervals, and 50–100 mL blood samples are collected four to five days after the final immunization [22, 23]. Transgenic mice can also be immunized to obtain B lymphocytes that exclusively produce HCAbs [24]. For mouse genes in which B cells fail to generate immunoglobulins after triple knockout, human heavy chain genes (V, D, J genes) as well as the modified murine Cγ1 gene and the intact 3' regulatory region were incorporated, at which point the B cells regained function and could yield HCAbs [25]. Lymphocytes are typically isolated from the collected blood, although some studies have extracted monocytes [26]. After mRNA conversion, cDNA is used to amplify the VHH gene regions [23], ideally through a nested PCR. A typical immune library should contain at least 10^6 unique transformations.

Naïve library

Naïve or synthetic Nbs libraries are viable alternatives when certain molecules are not immunogenic, cannot stimulate the generation of HCAbs, or are hazardous to animals. These libraries also address the time-consuming and expensive nature of immune libraries [27, 28]. To screen for high-affinity Nbs, naïve libraries typically range in size from 10^9 to 10^{11}. Due to the absence of in vivo affinity maturation, relatively large volumes of blood are required to obtain sufficient B lymphocytes. To build a naïve library with considerable scale and lower bias from individual prior immune responses, it is recommended to collect over 10 L of blood from 10 to 20 animals [9, 29, 30]. Compared to immune libraries, Nbs obtained from naïve libraries have lower affinities. Therefore, to overcome the lack of affinity maturation, isolated clones can be further improved by in vitro maturation [31, 32]. Phage display and ribosome display are the two primary methods for screening Nbs from naïve libraries [33]. The inability of bacterial cell transformation power to match the actual targeted diversity makes it challenging to have a highly diverse naïve library [9].

Synthetic/semi-synthetic library

When immunization is impractical or sufficient biological samples are scarce, synthetic or semi-synthetic libraries provide a promising alternative for constructing Nbs libraries. To date, the creation of synthetic/semi-synthetic Nbs libraries has mainly referred to sequences from llama derived Nbs due to their high melting temperature (Tm), which implies that they are better heat-tolerant, with a wider range of applications, and llamas have a lower cost of care. Hence, the number of reported llamas-derived sequences is quite substantial [34, 35].

The design of Nbs in synthetic libraries involves two crucial steps. The first is the framework sequence design, which focuses on stability and universality. The second is the hypervariable loops/CDRs design, which emphasizes diversity and effectiveness [36, 37].

Designing the framework for synthetic Nbs libraries relies on two main sources. The first is previously reported universal or Nb-derived frameworks, such as the highly versatile cAbBCII10, which maintains its functional structure without disulfide bonds and has been widely used as a scaffold for new synthetic libraries [38]. Its humanized version (e.g., hs2dAb) and derivative frameworks have also been extensively studied [39–41]. The second source is consensus sequences generated from natural gene libraries [42]. For example, the MacMahon library's consensus scaffold originates from the llama IGHV1S1-S5 gene, providing a convenient and user-friendly approach for constructing robust scaffolds [43]. However, the properties may not always meet the desired requirements, necessitating comprehensive experimental characterization, such as grafting experiments [36]. Notably, selecting scaffold sequences (humanized or not) from databases is a common approach [44]. For instance, Davide Ferrari et al. used the ABVDDB database for framework selection [45], while other database options include SAbDab-nano, iCAN, etc [37].

CDRs design is crucial for synthetic Nbs libraries, with CDR3 being the most critical due to its high variability and frequent interaction with antigens. CDR1 and CDR2 are also important and are typically designed based on the amino acid variability found in natural Nbs repertoires, although some studies have randomized all three CDR positions [46]. For CDR3, common methods include complete randomization, such as the NaLi-H1 library, which randomized all amino acids except cysteine [40], and the McMahon library, which excluded cysteine and tryptophan [43]. Chen et al. used a similar approach but included cysteine [42]. Another study adjusted CDR3 randomization by altering its length, obtaining three different geometric shapes of Nb binding sites: concave, loops, and convex [47]. Mixing and matching CDR sequences from natural llama and human antibody libraries is also feasible [48]. Notably, structure-guided semi-synthetic library design, which uses known structural information of antibody-antigen complexes for targeted randomizations, can be more precise and targeted than conventional synthetic library design [49].

Amino acid variability is usually achieved using degenerate codons [50]. NNS and NNK (N = A/T/C/G, K = G/T) are cost-effective for encoding all the 20 amino acids but have drawbacks, such as generating termination codons and excess of particular amino acids due to the bias of high codon redundancy. Designing degenerate codons under certain criteria is a more efficient way than complete randomization, and designing with reference to the databases containing naturally occurring Nbs sequences mentioned above is a common option [37]. Trinucleotide DNA assembly can overcome these defects but is more expensive [40]. Its superiority lies in its ability to encode the desired amino acid at every location.

Notably, one study developed a method to generate synthetic Nbs (sybodies) targeting any protein within 3 weeks [51]. Another study introduced two novel approaches for constructing synthetic libraries from mammalian cells, reducing the uncertainty of library construction by obtaining more than one million unique sequences without plasmid transformation extraction [52]. Additionally, researchers have engineered molecules by substituting the variable regions of conventional heterotetramer antibodies with two distinct Nbs, creating a molecule clip that biparatopically binds the receptor-binding domain (RBD) with high potency [53]. Similarly,

a biparameric Nb, termed Nb1-Nb2, demonstrates even greater potency than the aforementioned antibodies [54].

Display platform

High-quality Nb libraries significantly enhance the efficiency of selecting antibodies with high affinity and specificity. However, constructing Nb libraries alone is insufficient to obtain Nbs with superior performance. Nb production involves a series of processes, including identification, affinity maturation, purification, and ultimately, mass production, which necessitates the use of various well-established display technology platforms [9]. Until now, Nbs have been expressed in several kinds of production systems, ranging from virus, prokaryotic cells(E.coli e.g.), eukaryotic cells(yeasts, fungi, mammalian cell e.g.) and many more emerging technologies [55]. Display platforms can be broadly categorized into two main groups: surface display techniques, encompassing virus and cell-surface approaches, and non-surface display techniques, such as ribosome and mRNA-based methods [56]. This section will provide an in-depth exploration of the most widely utilized display methods within each category, critically assessing their strengths and limitations, and highlighting the latest research advances and future perspectives in the field.

Virus-surface display

Virus-surface display systems primarily include phage display and eukaryotic virus display, which can infect protozoa, fungi, plants, and animals, including human cells. Among all these, the bacteriophage-based system is one of the most widely used. The concept of phage display was first introduced by George P. Smith in 1985 when he successfully incorporated a DNA sequence encoding a display peptide into the filamentous phage gene III and expressed it as a fusion to the capsid protein III [57]. Phage display is a molecular technique that involves genetic modification of phage DNA to express peptides, proteins, or antibody fragments (such as Nbs) on the phage surface by fusing them to one of the phage membrane proteins [58]. The standard phage display process has been presented in Fig. 3(a) [59]. The M13 phage is most commonly used for phage display due to its ability to accommodate long fragments of exogenous DNA into its genome [60]. The expressed Nb fragment is typically displayed fused to the N-terminal end of the protein III (pIII) coat protein on the surface of the M13 phage [56].

Screening for Nbs with high affinity is an important part of the production process, and a general phage display requires 3∼5 rounds of biopanning followed by ELISA and other means to further characterise the affinity of the Nb. Phage display technology has matured considerably, with the development of advanced antigen presentation methods and screening strategies to enhance Nb selection efficiency. In addition to traditional direct and indirect antigen presentation approaches, novel techniques such as whole-cell panning, liposomes, nanodiscs, and virus-like particles (VLPs) have been adopted [61]. These methods are particularly well-suited for membrane antigen presentation and help maintain conformational integrity.

Regarding screening strategies, the nontarget selection strategy effectively minimizes the enrichment of Nbs against off-target antigens [62]. Epitope-specific deselection can be employed to select receptors that cluster with a specific locus. In this way enrichment can be achieved through epitope masking, where Nbs bind to and occlude certain epitopes of the antigen, rendering specific epitopes inaccessible to new Nb in the selection step [63]. To investigate the influence of the in vivo microenvironment on proper protein folding and post-translational modifications, researchers have developed in vivo phage display techniques. This involves intravenous injection of phage libraries, followed by intracardiac perfusion to remove unbound phages. Phages from target tissues are then extracted for sequencing analysis. Enrichment of phages can be determined by comparing sequences from target tissues with those from input or unrelated tissues. The enriched sequences are subsequently selected for further characterization [64].

Phage display offers significant advantages, primarily its capacity to present an extensive diversity of Nbs (> 10^11 unique clones) and the ability to replicate continuously without compromising host bacterial viability [65]. Due to limitations in display size and the absence of certain eukaryotic post-translational modifications(PTM) [55, 66], Nb formats like VHH is preferred for phage display compared to full-length immunoglobulins with multiple disulfide bonds [67]. Although, as previously mentioned, Nbs theoretically do not require N-glycosylation or post-translational modifications due to the absence of the Fc region and the presence of highly conserved CH2 sequences, making them suitable for any production platform, Nbs are sometimes designed to include other protein domains that may undergo glycosylation, such as Fc tags (which can enhance therapeutic effect [68]), or structures that require PTMs [69, 70], Additionally, Nbs often undergo glycosylation during production in hosts with glycosylation modification capabilities [55]. This glycosylation can potentially impact their immunogenicity, stability, half-life, and biological activity. Therefore, it is necessary to introduce the glycosylation and PTM functions of various production platforms. A comparative analysis of phage display with alternative platforms is presented in Table 2.

Despite these challenges, phage display remains a promising approach, particularly in the development

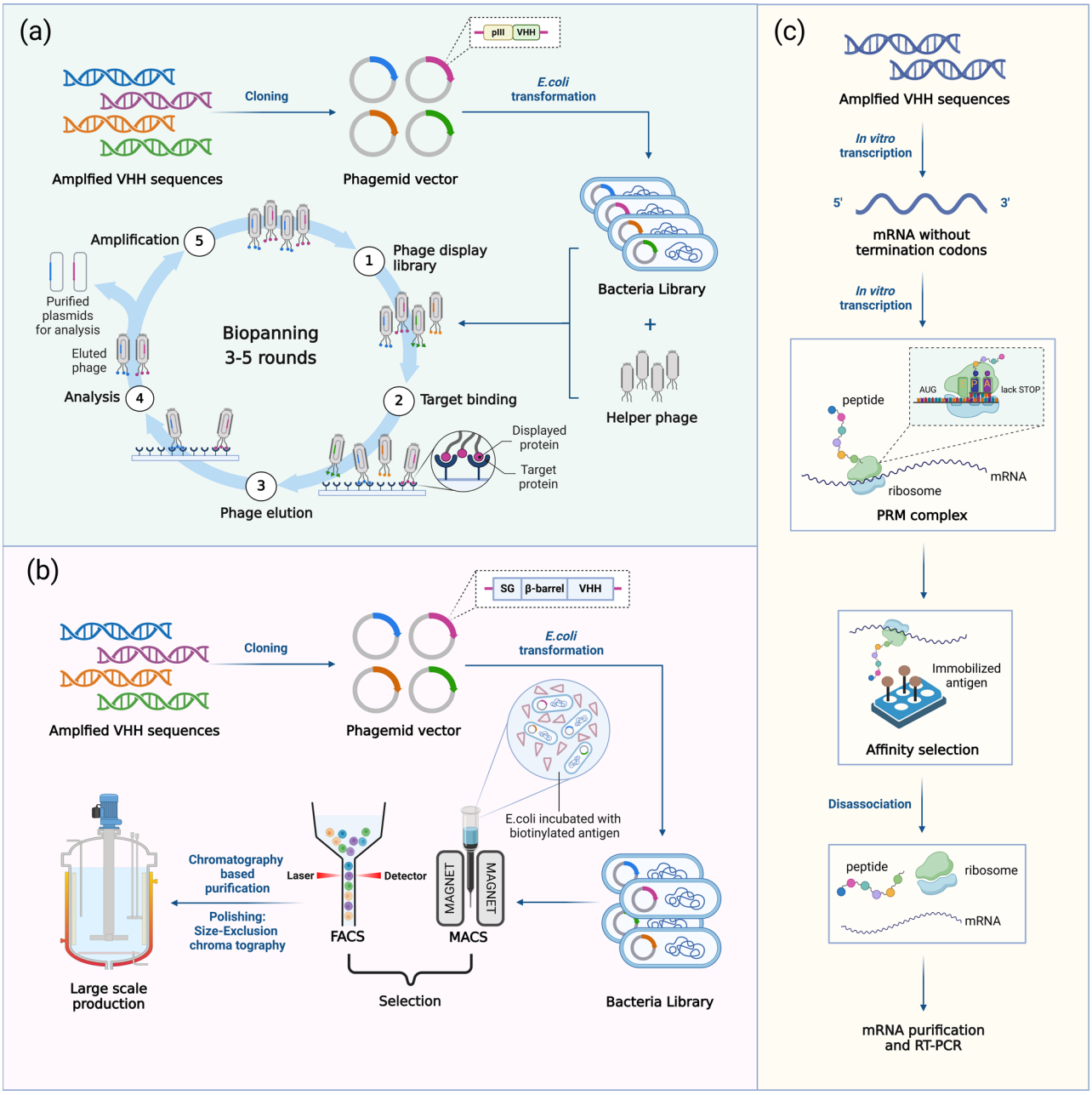


Fig. 3 Generalized protocol for three Nb display platforms: (**a**) Phage display: Nb/VHH cDNA from libraries (Immune, Naive, or Synthetic) cloned into phagemid adjacent to capsid protein gene (e.g., pIII) [58]. E. coli transformed and infected with helper phages [76]. Phage library undergoes 3–5 rounds of biopanning for antigen-specific selection [76]. Validation by ELISA [72]. (**b**) Bacterial surface display: Nb/VHH gene fragments cloned into plasmid vector and expressed on E. coli surface [80]. Screening by MACS using biotinylated antigen and antibiotic-conjugated magnetic beads [84]. FACS for further identification and characterization [84]. Selected clones undergo chromatographic purification and culture optimization [91]. (**c**) Ribosome display: mRNA without stop codon transcribed in vitro from Nb DNA library [147]. In vitro translation produces peptide-ribosome-mRNA (PRM) complexes [144]. Affinity selection against immobilized antigens [144]. Bound peptides isolated after dissociation [151]. RT-PCR of mRNA generates enriched DNA library for subsequent rounds [144]

of therapeutic antibodies(such as Nbs) [61]. Advanced strategies in library design and screening methodologies may facilitate the direct isolation of drug-like molecules [71]. The integration of phage display with emerging technologies, such as deep sequencing and bioinformatics, presents opportunities for enhanced data mining and rare clone identification [72]. Nevertheless, addressing plasmid vector instability remains a critical hurdle for further advancement in this field [73].

Table 2 Comparison of properties of several Nb Display platforms

Features	phage	Yeast	E.coli	Fungi	Mammalian Cells	Plant cell	Insect cell	Ribosome display	Ref.
Production time	7 ~ 10 days	46 ~ 72 h	30 ~ 48 h	5 ~ 7 days	2 ~ 3 weeks after pre-culture	Weeks	10 days after transfection	3 ~ 5 days	[69, 96]
Library size	10^9~10^{10}	10^6~10^7	10^9~10^{11}	10^{10}~10^{12}	10^6~10^8	10^6~10^8	10^6~10^8	10^{12}~10^{14}	[144, 159]
Expression level	High	Medium ~ High	Medium	High	Medium ~ High	High	High	High	[56, 69, 80]
N-linked glycosylation	-	High mannose No sialic acid, non-human sugars	None	Mammalian type core none human sugars	Similar to human	Complex no sialic acid non-human sugars	Complex no sialic acid non-human sugars	-	[55, 79, 97, 131]
Protein folding	Refolding required	Refolding sometimes required	Refolding required	Refolding sometimes required	Proper folding	Proper folding	Proper folding	Refolding required	[96]
Advantage	Simple Technically robust Automated [71]	Mimic natural selection [76] Scalability and cost-effective [80]	Rapid cell division, simple scale-up, high protein yields [56]	Productive Eukaryotic PTMs [101]	Human PTMs [117]	Eukaryotic PTMs Not prone to mammalian pathogens [114]	Infection Eukaryotic PTMs [138]	Bigger library size [144] In vitro induced mutagenesis [153]	
disadvantage	Limited library size [72] Plasmid instability gene deletion [65]	Limited library size [84] Transformation efficiency [80]	Limited expression levels [159]	Protease associated secretion [105] Relatively low heterologous protein production [104]	Limited diversity Elevated cost [112]	Immunogenic glycosylation pattern [114]	Glycosylation pattern [131]	Limited selection scope [72] Technically sensitive [146]	

Cell-surface display

Yeast

Yeast, a eukaryotic unicellular microorganism belonging to the fungal kingdom, utilizes its targeted secretion pathway to produce functionalized heterologous proteins [74]. The ability to undergo proper protein folding and PTM in vivo is a key advantage of yeast display [55]. Commonly used yeasts for Nb display include S.cerevisiae and P.pastoris, with S.cerevisiae being the most frequently employed species [74].

Aga1p and Aga2p are two membrane proteins of S.cereviasiae, which are interconnected by disulfide bonds on the cell wall surface [71]. The displayed Nb fragment is located at the C-terminal end of the Aga2p subunit, forming the biological basis of yeast display. The expression of both Aga2p and Nb-Aga2p products is regulated by the GAL1 promoter, so Nb display only occurs in the presence of galactose [71].

The yeast display process involves library construction, gene cloning, transformation, and selective culturing. Library construction is achieved using overlap-extension PCR or transforming cells with a linearized plasmid vector and DNA fragments [75]. The target Nb gene is cloned into a yeast display plasmid, fused to the Aga2p gene, and transformed into yeast cells by electroporation or lithium acetate treatment. Transformed cells are selectively cultured to ensure fusion Nb expression on the cell surface [76]. Magnetic-activated cell sorting (MACS) and fluorescence-activated cell sorting (FACS) play crucial roles in this process [77]. MACS mimics the screening process of phage display and can greatly reduce non-reactive background, allowing a reasonable number of yeast cells to be used for subsequent FACS screening [71]. In flow cytometry, by using lasers to individually detect cells with different morphologies, structures, and fluorescent properties within a cell stream, rapid, objective, and sensitive multi-parameter analysis of cells can be achieved [78]. Compatibility with FACS is a significant advantage of yeast display over phage display, which allows for a highly controlled and real-time selection step, enabling fine differentiation of Nbs with different properties and easy access to conjugates capable of distinguishing highly homologous antigens. After specific Nb clones are obtained, the Nbs can be purified by expanding the culture of selected clones and using techniques such as protein A/G affinity chromatography to enrich for cells that produce high-affinity, specific Nbs [76].

One of the extraordinary advantages of yeast display over phage display and prokaryotic cell display systems is that, as a unicellular eukaryote, yeast facilitates the

expression and folding of eukaryotic proteins [79]. Its highly efficient post-translational modification (PTM) mechanism allows for better functionalization of Nbs. Additionally, yeast display can generate polyclonal populations with different binding specificities in a single experiment, which is useful for diagnostic analysis of complex biological systems [43]. Main drawback lies in the potential misfolding of displayed proteins, which may lead to loss of Nb function [80]. Furthermore, in the production of Nbs, most do not contain N-glycosylation sites due to the lack of the fragment crystallizable (Fc) structural domain. However, approximately 10% of Nbs still contain potential glycosylation sites and are N-glycosylated [55], which can decrease the capacity of VHHs to bind to the antigen and increase their immunogenicity [81].

Recent optimization strategies for yeast display include improving display Nbs, promoters, signal peptides, library diversity, and Nb expression levels [76]. High-throughput methods like CRISPR/(d)Cas technologies are enhancing screening efficiency in S. cerevisiae-based systems [82]. A novel hybrid approach combining phage and yeast display has emerged, leveraging phage display for initial Nb enrichment and yeast display with FACS for fine-tuned Nb selection. This method integrates the strengths of both techniques, significantly improving overall screening efficiency [83].

Bacterial

Bacterial surface display, especially using E.coli, has shown impressive performance in producingNbs due to its advantages in cloning, amplifying, and maintaining large Nb gene libraries [80]. The process involves inserting a vector containing exogenous protein gene sequences into bacteria and expressing the Nb through various pathways [84]. E. coli is popular due to its genetic tractability and transformation efficiency. Its structure consists of two cell membranes with a periplasmic space in between, and Nb must pass through the inner membrane and periplasmic space to be anchored to the outer membrane. Various methods are used to display Nb, with inner membrane protein fusions being optimal for Nb libraries [84, 85]. Gram-positive bacteria like staphylococci, streptococci, and mycobacteria are also used for surface display.

Bacterial surface display starts with the construction of cloned plasmid vector with amplified Nb/VHH fragment from Nbs library. Methods like FACS, MACS, or direct screening on live tumor cells are used to identify and screen bacteria expressing surface antigens [84]. MACS and FACS are often combined to screen antigen-binding colonies, identify binders, and characterize their antigen-binding properties, such as specificity, ligand competition, and K_D assay. Flow cytometry can also be used to classify Nbs [86]. The process is illustrated in Fig. 3(b). In E. coli, Nbs can be produced through periplasmic, cytoplasmic, or extracellular expression [87]. Each cellular compartment has unique properties: (i) Periplasmic expression is well-suited for Nb production due to its oxidative environment and presence of chaperones and isomerases [88]. However, low yields may occur due to insufficient chaperone proteins, which can be addressed by using plasmids containing chaperone proteins (e.g., pTUM4) [89]. (ii) Cytoplasmic expression is necessary for large fusion proteins, but the reducing environment hinders disulfide bond formation and proper folding. This can be overcome by constructing double mutant strains with defects in both the thioredoxin and glutathione pathways [90]. (iii) Extracellular expression provides an oxidative environment and effective protein folding milieu while reducing impurities [91]. The three expression patterns are shown in Fig. 4.

Bacterial display offers advantages like high yield, productivity over other commonly used display systems such as phage, yeast, and mammalian cells. Bacterial systems like E. coli can accommodate libraries of up to 10^{11} clones, providing a 10- to 10,000-fold higher library diversity compared to the 10^10 limit for phage display and 10^7 for yeast display, enabling screening of significantly larger and more varied libraries [56, 80]. The absence of the Fc domain in Nbs makes bacterial display well-suited for their production. E. coli's simple and inexpensive medium compositionand compatibility with high-throughput screening methods like FACS further enhance its appeal [92]. Most notably, the rapid growth rate (20 min doubling time), low culturing costs of E. coli make Bacterial display a more widely used and efficient platform for library screening and Nb engineering [80]. However, prokaryotes lack human-like post-translational modifications, which can affect Nb stability and immunogenicity [56]. Another potential drawback is the presence of endotoxins [55]. Other disadvantages include inclusion body formation, codon bias, metabolic burden, and Nb degradation [91].

Researchers have made significant efforts to improve bacterial display systems by optimizing fermentation culture conditions, using genetically engineered expression vectors like pOPE101 and pETDuet-1 [92], and developing new expression methods such as the constitutive outer membrane F (OmpF) high-expression system [93]. "Bacterial glycoengineering," which utilizes the prokaryotic glycosylation system of E. coli, is an emerging approach with great potential for the N-glycosylation of Nb production [94]. This development could help overcome the constraints of bacterial display systems in generating more stable and less immunogenic Nbs.

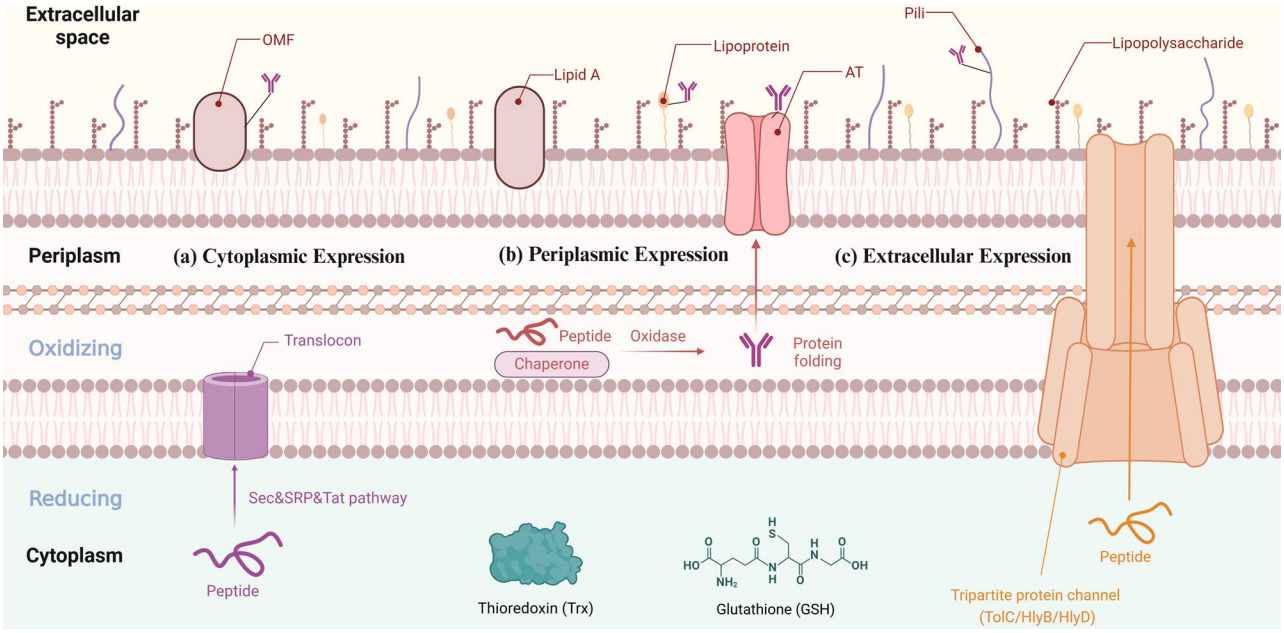


Fig. 4 Functional Nb expression in different cellular compartments of E. coli. (**a**) In the cytoplasm, Nbs are synthesized in a reducing environment maintained by thioredoxin (Trx) and glutathione (GSH) systems [55]. (**b**) For periplasmic expression, Nbs are initially synthesized in the cytoplasm and then transported across the inner membrane via Sec, SRP, or Tat pathways through the translocon [55]. In the oxidizing environment of the periplasm, chaperones and oxidases assist in protein folding [88]. Surface display can be achieved through various structures including outer membrane proteins (OMPs), pili, lipoproteins, and autotransporter (AT) proteins [80]. (**c**) For extracellular expression, Nbs can be secreted directly from the cytoplasm to the extracellular medium through the tripartite protein channel (TolC/HlyB/HlyD) that spans both inner and outer membranes [269]

Fungi

Fungi, particularly Aspergillus, Xylaria, Penicillium, Fusarium, Rhizopus, and Trichoderma, have emerged as powerful cell factories for heterologous protein production [95]. Their eukaryotic post-translational protein processing mechanisms significantly influence protein secretion, activity, stability, and immunogenicity [96, 97]. Fungi require transfection with gene expression vectors to express heterologous proteins such as Nbs, using methods like Polyethylene Glycol -mediated protoplasmic transformation, Agrobacterium-mediated transformation, electroporation [98], and underwater shockwaves, which can enhance efficiency by 5400-fold [99]. Aspergillus species are proficient in producing and secreting natural and heterologous proteins, organic acids, and secondary metabolites [100].

In terms of production yield alone, fungi are more suitable as cell factories for large-scale Nb production [96, 101]. Phage display technology, although mature and widely used in Nb antibody preparation, is better suited for library screening and analysis rather than direct Nb production [102]. Bacterial display allows for large-scale Nb production, but E. coli is best used for expressing non-glycosylated Nbs due to the lack of N-linked glycosylation [103]. However, bacteria remain the most common display platform for Nb production, owing to their outstanding cost advantage and high yields. Fungi, despite their potential, have consistently shown unsatisfactory production capacity for heterologous proteins (including Nbs) compared to homologous proteins, which is a major limitation [104], possibly due to the influence of proteases during secretion [105].

Researchers have developed strategies to augment yield, including promoter assembly, codon optimization [106], signal peptide replacement, carrier protein use, glycosylation site engineering, regulation of unfolded protein response and endoplasmic reticulum-associated protein degradation, optimization of intracellular transport, regulation of unconventional protein secretion, and construction of protease-deficient strains [101]. Mycelial cell polarity impacts protein secretion in filamentous fungi, with effective secretion primarily occurring at hyphal tips. Highly branched phenotypes typically exhibit enhanced protein secretion efficiency [107]. Strategies to promote branching hyphae formation include environmental modifications [108] and genetic engineering approaches [109].

While filamentous fungi offer advantages as cell factories, including excellent protein production efficiency, PTM capabilities, and generally recognized as safe status, further research using transcriptomics, genomics, proteomics, and metabolomics is essential to elucidate the expression, regulation, and secretion mechanisms of heterologous proteins in these organisms [110].

Mammalian cells

Mammalian cells is well-suited for the production of functional antibodies(including Nb), due to their close evolutionary relationship to humans, resulting in proteins that are more functionally intact and Nbs with reduced immunogenicity and enhanced safety profiles [55]. Traditional mammalian cell display techniques involve transfecting recombinant gene vectors [111], followed by screening to select the most productive and stable clones [112].

However, this process can be time-consuming. Novel methods like transient gene expression and recombinase-mediated cassette exchange have emerged to enhance protein production efficiency. Transient gene expression (TGE) eliminates the need for cell line cloning [113], while Recombinase-mediated cassette exchange (RMCE) enables targeted integration of the recombinant construct into a stable and transcriptionally active genomic hotspot [114]. Furthermore, lentiviral vectors have also proven to be powerful tools for gene delivery and Nb production [112].

The mammalian cell expression platform primarily includes CHO, HEK293, PER.C6, and CAP/CAP-T cell lines [112], each with distinct characteristics [111]. CHO cells are the predominant host for industrial-scale protein production due to their adaptability to serum-free conditions, while HEK293 cells are well-suited for laboratory settings due to their transient and semi-stable expression characteristics and distinct human glycan profiles. CAP cells, derived from human amniocytes, are notable for their high protein production rates and human-like glycosylation patterns. To enhance CHO productivity, researchers have developed various strategies, including advanced culture methods [115] and cellular engineering approaches [111]. However, CHO cell lines face challenges under high genomic and metabolic demands, leading to decreased productivity and product quality. Stable, high-yielding CHO clones often exhibit deletions in the telomeric region of chromosome 8 [116], and recent progress in transcriptomics has provided further insights into this phenomenon.

Precise regulation of cellular phenotypes is equally important for studying Nb production in disease contexts. Researchers employ synthetic biology tools, including gene editing techniques, regulatory expression tools [117], and bioinformatic tools for multi-omics analysis, to achieve optimal phenotypic characterization. Mammalian expression systems still present challenges in post-translational modification optimization, particularly glycosylation. Mammalian glycosylation patterns can modify the backbone sequence of proteins [103], which in turn influences the stability, half-life, and biological activity of Nbs. Cell culture conditions have been shown to influence glycan profiles [118], highlighting the complexity of optimizing mammalian cell-based protein production and the need for a holistic approach to cell line development and process optimization.

Plant cell

Plant cell display systems offer a viable alternative to mammalian systems, providing similar post-translational modifications while being generally safe and harmless to humans [119]. Heterologous protein expression in plants primarily employs three methods: stable nuclear genome display [55], transient viral transfection [120], and plastid genome display [69]. While Nb production mainly utilizes Agrobacterium-mediated stable nuclear genome integration, transient viral transfection is gaining traction due to its rapid production cycle, yielding proteins within days compared to the years required for developing transgenic plant lines.

For Nb production, N. tabacum, N. benthamiana, and A. thaliana [121] are the predominant plant hosts [121]. These plants offer versatile expression options, with antibody accumulation possible in leaves, seeds, and roots, as well as various subcellular compartments including the cytoplasm, chloroplasts, vesicles, endoplasmic reticulum, and apoplasts [121]. Although leaves account for majority of Nb production, seeds present unique advantages. They can preserve Nb stability and functionality during long-term storage and are particularly suitable for oral applications [122]. This dual benefit of seeds - long-term stability and oral delivery potential - makes them an attractive option for certain therapeutic applications, complementing the high-yield capability of leaf-based production [123].

Nb production primarily utilizes stable nuclear genome expression infiltrated by Agrobacterium strains, but transient viral transfection expression is gaining popularity due to its high production efficiency and rapid turnaround time [121]. Nbs can be produced in various plant tissues (such as leaves, roots, seeds) and accumulate in different subcellular compartments (cytoplasm, chloroplasts, vesicles, endoplasmic reticulum, apoplasts) [124]. Seeds can be stored for long periods of time to maintain antibody stability and functionality and are suitable for oral application despite approximately 70% Nbs being produced in leaves [125]. The transient viral transfection expression process typically involves constructing a recombinant vector, transforming it into E. coli for purification, introducing the purified vector into Agrobacterium, culturing in selective media, agro-infiltrating the resulting bacterial culture into targeted plant tissues, and finally, purifying and characterizing the recombinant proteins [69].

Plant expression systems offer advantages in production speed and flexibility for product and process optimization. Utilizing transient expression,

plant-manufactured VHHs for medical applications can be produced within a few weeks [69]. However, plant-specific secondary modifications to proteins can increase immunogenicity. Strategies to address this include genetic engineering to promote mammalian-type glycosylation [126] and retaining glycoproteins in the endoplasmic reticulum to block plant-specific glycan attachment [127]. Despite these challenges, plant-expressed Nbs can maintain biological activity even with differences in glycosylation [121]. Another obstacle is the high downstream purification cost [128]. Standard purification involves extraction, chromatographic techniques, and analysis using SDS-PAGE, Western blot, ELISA, and surface plasmon resonance (SPR), accounting for approximately 80% of the total process cost [69]. Strategies such as antibody recycling and producing less processed oral therapeutics are being explored to reduce costs.

Insect cell

Insect cells offer significant advantages for Nb production due to their eukaryotic nature and ability to perform post-translational modifications similar to mammalian cells [129–131]. The Baculovirus Expression System (BES) has emerged as a productive, safe, and cost-effective alternative to traditional insect cell fermentation [55], exploiting the baculovirus life cycle by inserting exogenous genes while replacing polyhedrin and P10 genes to drive target protein(Nb) expression [132]. Constructing an insect cell Nb expression system involves inserting the target Nb gene into a plasmid vector [133], generating a recombinant baculovirus through homologous recombination, and co-transfecting insect cells with the engineered plasmid and baculovirus genomic DNA. Alternatively, recombinant baculoviruses can be produced via site-specific transposition of the Tn7 transposon into a baculovirus vector (Bac-to-Bac expression system) [134], though its industrial application is limited due to inherent instability during amplification [135].

Lepidoptera cells, particularly those from undifferentiated ovarian and embryonic tissues, are preferred for Nb production. The Autographa californica multiple nuclear polyhedrovirus (AcMNPV) is the most widely used baculovirus vector [136]. Ongoing genetic modifications of the AcMNPV backbone and transfer vectors have led to improvements such as the flashBAC GOLD backbone, enhancing secretory pathway efficiency and reducing recombinant Nb degradation. Adaptive Laboratory Evolution (ALE) offers a promising approach to enhance protein production efficiency in BES by subjecting cells to non-standard culture conditions and applying continuous selection pressure [137], facilitating the emergence of more resilient cell populations adapted to the production environment.

Insect cell expression offers several advantages for Nb production, including high yields, proper folding, and post-translational modifications. While insect cells offer similar expression and processing pathways to higher eukaryotes, they produce N-chain glycosylated Nbs with simple oligomannose chains [138], contrasting the complex glycans with terminal sialic acids found in mammalian cell-produced glycoproteins, enhancing their immunogenicity. Current research focuses on obtaining high yields of secreted proteins and improved glycosylation patterns [139]. Despite this limitation, the rapid production timelines and scalability of the baculovirus-insect cell expression system make it an attractive platform for Nb display and manufacturing.

Considerable research has been conducted on optimizing insect cell culture environments for large-scale expression production, including improved media, bioreactor design, and operation strategies, coupled with investments in process optimization, production, and quality control of proteins using the Insect cell Baculovirus Expression Vector System (BEVS-IC) [138]. Recent research has also focused on the downstream process of BES, with contaminants categorized into host cell DNA, baculoviruses, host cell proteins, baculoviral proteins, media residues, and wastes [140]. A generic purification process can be designed around these contaminants, dramatically shortening development time once established. Methods include selecting appropriate elution buffers, capturing the product of interest using affinity and/or ion exchange chromatography, applying cell line or baculovirus vector engineering [136], RNA silencing [141], and DNA enzyme application [136].

Non-surface display

Cell-free display systems offer an innovative approach to Nb expression and selection, circumventing the necessity for constructing Nb vectors and transforming cell hosts. These systems utilize transcription/translation machinery extracted from ribosome-rich sources, such as wheat germ or E.coli, to establish a genotype-phenotype linkage through covalent interactions between protein, DNA, RNA, and ribosomes [142]. Researchers have developed a diverse array of cell-free in vitro display platforms, including ribosome display, mRNA display, covalent and noncovalent DNA display, and in vitro compartmentalization [56]. Among the various cell-free display platforms, two prominent methodologies have emerged as particularly noteworthy: ribosome display and mRNA display.

Ribosome

The ribosome display system couples genotypes and phenotypes by forming protein(Nb)-ribosome-mRNA (PRM) complexes, eliminating cell culture and

transfection steps, shortening the isolation and selection cycle of high-affinity functional proteins, and circumventing microbial cell transformation [143, 144]. The Ribosome Display System couples genotypes and phenotypes by forming PRM complexes, eliminating cell culture and transfection steps, shortening the isolation and selection cycle of high-affinity functional proteins, and circumventing microbial cell transformation [145]. However, RNase contamination and PRM complex instability are major challenges, requiring an RNase-free environment [146].

In ribosome Nb display, DNA encoding Nb library is transcribed into mRNA lacking a stop codon, ensuring the attachment of the newly synthesized peptide and encoding mRNA [66]. The mRNA is then transcribed and translated in vitro to form the PRM complex, which is affinity-selected by binding to immobilized antigens. The displayed Nb peptide chain is recovered, and the mRNA is reverse-transcribed and PCR-amplified into Nb-encoding DNA [144]. A typical production cycle is displayed in Fig. 3(c). Both eukaryotic and prokaryotic cell-free translation systems can be used, with no confirmed superiority [147].

Ribosome display allows in vitro Nb evolution through mutation construction and selection, making Nb affinity and maturation its most successful application [148]. Successful biopanning requires PRM complex stability, which can be improved by high magnesium ion concentrations, appropriate temperatures, and antisense RNA [146, 149, 150]. RT-PCR generates a purified DNA library encoding the desired antibody phenotype for multiple rounds of demonstration [151]. Ribosome display has emerged as a powerful tool in Nb engineering, offering advantages over cell-based display systems [144]. It enables the generation of vast libraries (10^12 to 10^14 unique clones) [144], surpassing the capabilities of cell surface display methods like yeast display (10^6 to 10^7) and bacterial display (10^8 to 10^10), thereby enhancing the likelihood of identifying high-affinity Nbs. Furthermore, Ribosome display allows in vitro Nb evolution through mutation construction and selection, making Nb affinity and maturation its most successful application [152]. In vitro diversification strategies can be categorized into two main groups: targeted approaches and random mutagenesis [153]. Targeted methods, such as hotspot mutagenesis and parsimony mutagenesis, allow stochasticity to occur at a given location, whereas random mutagenesis methods, including error-prone PCR and shuffling, are closer to random mutagenesis in vivo [154]. Error-prone PCR utilizes low-fidelity polymerase to randomly generate point mutations in the gene sequence during PCR amplification [155], which has been used to obtain more stable and specific Nbs [156].

However, the instability of PRM complexes and ribosome-ribosome collisions due to the lack of a stop codon can hamper translation efficiency, necessitating careful optimization of the mRNA-to-ribosome ratio [152]. Future research could explore synergistic integration of ribosome display with other selection platforms (e.g., phage or yeast display) to leverage the strengths of each technology [47]. Integrating ribosome display with next-generation sequencing and AI could revolutionize the throughput and precision of antibody discovery and optimization [157].

mRNA

mRNA display, another cell-free technique akin to ribosome display, begins by transcribing a constructed DNA library into mRNA, which then forms protein-mRNA complexes via in vitro cell-free translation [158, 159]. Unlike ribosome display, mRNA display uses a short DNA-puromycin linker attached to the displayed protein to prevent peptide chain release, instead of relying on the absence of a stop codon. Puromycin mimics the amino terminus of a typical tRNA, acting as a translational repressor by entering the ribosomal A-site and binding to the resulting protein [160].

The covalent bond between mRNA and the displayed protein makes the complex more stable than in ribosome display, eliminating the need for complex-specific stabilization [161]. Furthermore, the smaller size of puromycin DNA linkers, compared to ribosomes, reduces the interaction with fixed selection targets, leading to less biased mRNA display results [162]. However, mRNA degradation remains a challenge, requiring RNase-free experimental conditions. Interestingly, a recent study showed that up to 60% of mRNA display libraries could be recovered intact after cell incubation [163], suggesting that the impact of RNase on experimental outcomes may be less significant than previously thought.

Emerging technologies

Recent advancements in antibody screening and presentation have introduced techniques that leverage computer analysis and NGS/LC-MS, such as Nestlink, Sybody, and high-throughput sequencing combined with mass spectrometry identification [9]. Nestlink is particularly suitable for complex screening, like Nb identification in organisms [9]. It overcomes the limitations of traditional biopanning and low peptide counts in LC-MS/MS due to extensive sequence homology in binding collections [164].

Nestlink uses genetically encoded barcode peptides called “flycodes,” linked to a binding Nb library through “library nesting“ [164]. NGS assigns flycodes to their corresponding binding Nb, and the nested library is expressed as a pool of flycodes from selected binders.

These flycodes are then separated and detected, enabling unambiguous identification and high-throughput characterization of thousands of binding Nbs [165]. By integrating NGS and LC-MS/MS data, Nestlink establishes a genotype-phenotype linkage independent of physical entities, allowing simultaneous monitoring of biodistribution, tissue penetration, immunogenicity, and serum half-life of numerous biopharmaceutical drug candidates in a single disease-associated model organism [165].

Sybodies are useful when binding agents cannot be generated by immune library due to high sequence conservation, toxicity, or insufficient target protein stability. Sybody generation involves a single round of ribosomal display using mRNA-encoded sybody libraries, followed by two rounds of phage display and ELISA to screen for Nbs with good binding activity [51]. SPR determines sybody affinity [166] and facilitates further purification. Compared to other display systems, sybody offers faster and more efficient screening, the ability to screen highly conserved proteins, recognition of three-dimensional (3D) epitopes, and broader target coverage [47]. Sybody can also be tailored to various applications by adding or removing tags through expression vectors [166].

High-throughput sequencing and mass spectrometry identification eliminates the need for intermediate expression systems [167]. A cDNA library from immunized alpaca bone marrow lymphocytes is amplified by PCR and subjected to high-throughput DNA sequencing. Concurrently, serum from the same alpaca undergoes affinity purification of VHH protein fragments, which are analyzed by LC-MS/MS. The resulting MS data is compared with the sequence database generated from DNA sequencing reads to determine the corresponding VHH sequence [168]. This approach eliminates the need for efficient exogenous expression, folding, and clonal presentation, enabling rapid screening of numerous Nbs targeting multiple epitopes of a specific antigen within a relatively short timeframe (10 days post-sample collection). It can generate a large library of different Nbs, although its dependence on the camel's immune system may present a potential limitation [167].

These emerging technologies are revolutionizing antibody discovery and optimization, offering increased throughput, broader applicability, and enhanced efficiency in identifying and characterizing novel binding proteins for various applications in research and therapeutics.

Artificial intelligence in Nb engineering

Currently, most Nbs are developed through animal immunization, which confers advantages in terms of antibody affinity and specificity [169]. However, advancements in in vitro techniques and ethical concerns surrounding experimental animals have propelled the emergence of novel approaches [170]. AI technologies have already achieved remarkable success in various domains. The powerful learning capabilities of deep neural networks enable them to automatically learn multifaceted features from diverse data types, constructing highly flexible and robust models. In particular, excellent models such as Graph Neural Networks (GNNs) and Transformer have demonstrated outstanding performance [171, 172]. In the biomedical field, by analyzing vast amounts of genomic or imaging data, AI can accurately identify and predict disease-associated genetic variations or clinical outcomes [173–176]. AI tools like AlphaFold have showcased exceptional predictive accuracy, precisely determining the 3D structure of proteins based on their amino acid sequences [177–179]. In the realm of drug discovery, AI has significantly accelerated the development process and reduced research costs and timelines through virtual screening and drug design [180, 181] (Fig. 5a).

The introduction of AI is imperative in the Nb design process. Conventional laboratory and computational methods suffer from several limitations, including being time-consuming, labor-intensive, and reliant on databases containing known structures [47, 178]. Moreover, the screening efficiency and functional performance of the three major libraries currently available are still restricted [182]. Therefore, integrating AI into Nb engineering will overcome the bottlenecks of traditional Nb development and design (Fig. 5b). On one hand, computer simulation and prediction can greatly reduce workload, accelerate the development process, and lower costs. On the other hand, AI can efficiently explore the vast sequence landscape of Nbs, identifying potentially superior candidate sequences while considering multiple performance criteria such as affinity, specificity, stability, and polyreactivity, which is particularly crucial for Nb discovery and performance optimization.

Moreover, the rapid advancement of high-throughput sequencing technologies has provided unprecedented volumes of Nb sequence data, which serve as a critical foundation for AI-driven optimization [183–185]. These datasets capture the inherent diversity of Nbs, including rare and unconventional sequence motifs, enabling deep learning models to identify subtle patterns linked to affinity, specificity, and stability [184, 186, 187]. For example, by analyzing sequencing data from immune and synthetic libraries, AI models can uncover sequence-function relationships, such as key residues in CDR3 loops that drive antigen specificity or structural stability [42, 188, 189]. Furthermore, these large-scale datasets enhance model generalizability by expanding the diversity of training sets, particularly for predicting novel Nb conformations or optimizing sequences for improved physicochemical properties [183, 187, 190]. To efficiently manage and

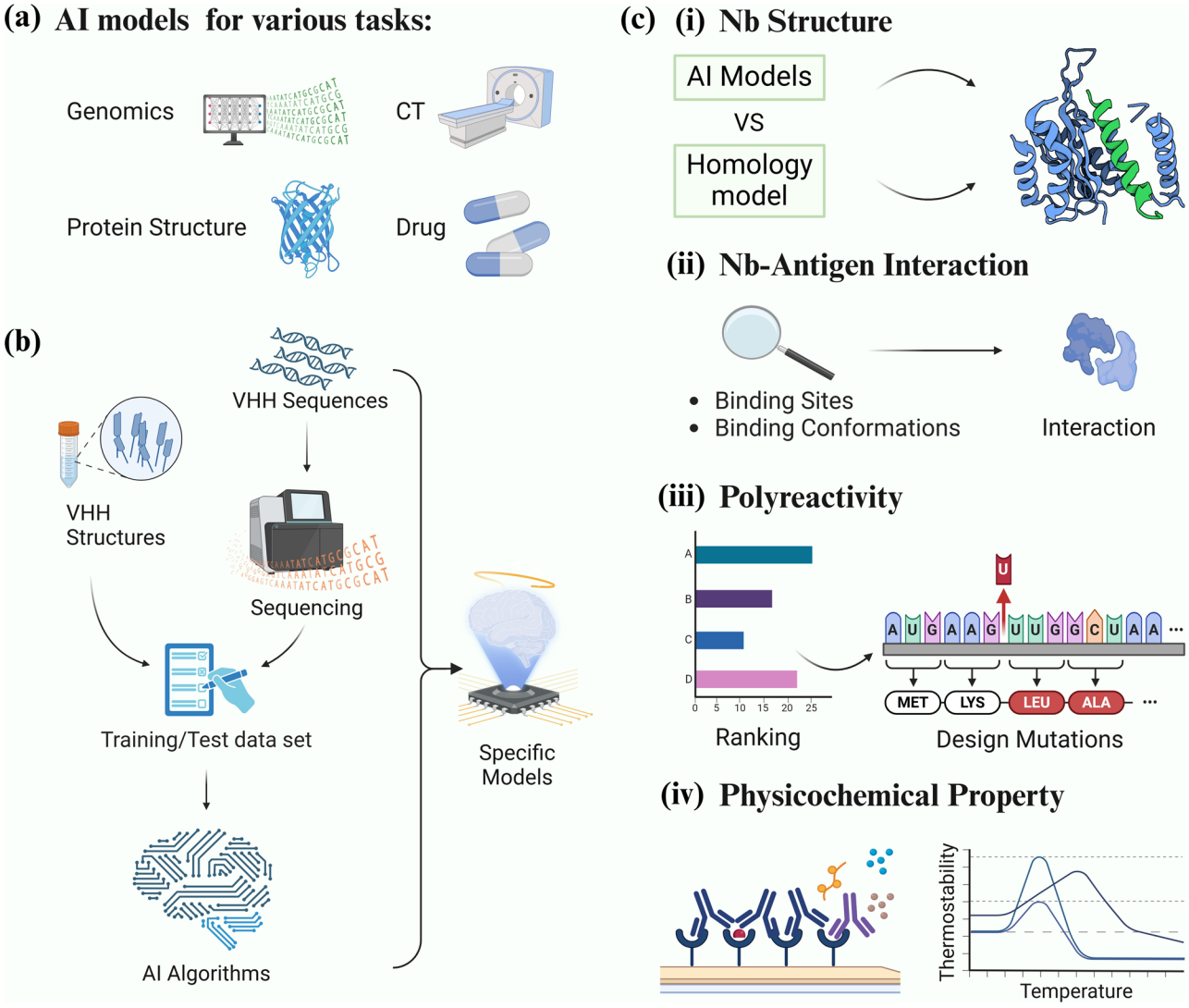


Fig. 5 Artificial Intelligence in Nb Engineering. (**a**) The integration of AI into nanobody (Nb) engineering has revolutionized the field, with AI models already being extensively applied in genomics, medical imaging, protein structure prediction, and drug discovery [176–182, 197, 198]. (**b**) AI algorithms, trained on Nb sequence and structural data, enable more efficient and accurate predictions in several key areas. (**c i**) AI models like AlphaFold and NanoNet are used to predict Nb 3D structures, focusing on the critical VHH region [178, 184, 209, 215, 221]. (**c ii**) AI is also leveraged to predict Nb-antigen interaction sites and optimize binding conformations, using algorithms like NanoBERTa-ASP and NbX to improve precision in predicting the binding sites and structural interactions [186, 234]. (**c iii**) AI-powered models have contributed to addressing Nb polyreactivity, enhancing performance by predicting the impact of amino acid mutations on polyreactivity and optimizing Nb stability [240]. (**c iv**) AI tools like nanoBERT are specifically designed to explore the effects of sequence mutations on Nb physicochemical properties, significantly improving the affinity, stability, and functional characteristics of Nbs [263]. These innovations collectively enhance the rational design of Nbs for therapeutic and diagnostic applications

utilize these vast datasets, several specialized databases have been developed, forming the backbone of AI-driven optimization in Nb engineering [191–196]. These databases not only provide structured and high-quality data but also support critical tasks, such as sequence-function relationship analysis, structure prediction, and thermal stability optimization. By integrating these resources, AI frameworks can bridge the gap between experimental discovery and computational design, significantly accelerating the identification of superior Nb candidates while maintaining high precision [42, 184–186, 188]. Table 3 provides a summary of representative nanobody databases, highlighting their primary features and roles in supporting AI-based Nb engineering.

In summary, AI can overcome the shortcomings of traditional Nb engineering, enabling more flexible, efficient, and precise antibody design. It plays a vital role in tasks such as antibody structure or antigen-binding site prediction, and physicochemical property optimization (Fig. 5c; Table 4).

Table 3 Summary of existing nanobody databases

Database Name	Data Scale	Data Content	Features	Application	Website Link
iCAN [191]	2391 nanobody entries	Includes sequences, structures, target antigens, and functional information from patents, RCSB PDB, and EMBL	First comprehensive nanobody database with a user-friendly interface and prediction tools	Facilitates nanobody research with analysis tools such as Blast and Clustal	http://ican.ils.seu.edu.cn/
INDI [192]	Over 11 million nanobody sequences	Nanobody sequences from GenBank, patents, NGS repositories, and scientific publications, with metadata	Automated data integration with extensive coverage, suitable for computational design	Supports immunoinformatics research and accelerates nanobody drug development	http://naturalantibody.com/nanobodies
NanoLAS [193]	Consolidated data from various sources	Includes structures, sequences, and target information, emphasizing applications in COVID-19 and cancer	Comprehensive data integration with a user-friendly interface, especially for SARS-CoV-2 and cancer research	Provides a unified data platform to advance biomedical applications of nanobodies	https://www.nanolas.cloud/
NbThermo [194]	564 nanobody thermostability entries	Contains melting temperature (Tm), sequences, antigens, and structures	First database focused on nanobody thermostability, with manually curated data	Provides thermostability references for nanobody engineering	https://valdes-tresanco-ms.github.io/NbThermo
SAbDab-nano [195]	Over 800 structural entries	Includes nanobody structures, resolution, CDR sequences, and antigen-binding affinities	Updated weekly with detailed annotations, suitable for structural analysis	Enables 3D structure analysis and development of biotherapeutics	http://opig.stats.ox.ac.uk/webapps/newsabdab/
sdAb-DB [196]	Over 788 nanobody sequences	Provides sequences, target information, binding affinities, and experimental data	Open-source community database, ideal for sharing and redesigning antibody sequences	Supports bioengineering and synthetic biology design	http://www.sdab-db.ca/

AI in Nb structure prediction

The quest for determining the 3D structure of proteins has navigated a winding road of advancements [197–203]. Recently, owing to the development in GPU and algorithmic innovations, deep learning-based methods have elevated the accuracy of protein structure prediction to new heights [177, 178, 204], ushering in breakthroughs for the study of Nbs 3D structures. The powerful feature extraction and learning capabilities of deep neural networks [171], coupled with the unique advantages of state-of-the-art deep learning models like GNNs [172] and Transformer [205] in processing sequence and graph-structured data, enable the automatic learning of intrinsic patterns in Nb structures from vast amounts of data.

General and specialized AI tools for VHH structure prediction

VHH, the variable region of the heavy chain of Nb, is the functional core of Nbs responsible for recognizing and binding specific antigens [182]. Elucidating their 3D structure is crucial for understanding their unique antigen recognition mechanisms and providing guidance for subsequent affinity maturation and functional optimization. Current research on it primarily focuses on their antigen-binding mechanisms [51], conformational features [206] and the dominant role of CDR3 in the binding process [207, 208]. The unique conformational characteristics of VHHs, such as the β-hairpin or helix-loop conformations [208], despite the specific features remain unclear, may play a unique role in recognizing specific antigenic epitopes. Deep learning-based tools have significantly advanced the prediction of VHH structures, providing valuable insights for rational design.

Protein and antibody structure prediction has been revolutionized by deep learning-based tools such as AlphaFold2 [178], RoseTTAFold [209], OmegaFold [210], ESMFold [211], OpenFold [212], IgFold [213] and ABodyBuilder3 [214]. These tools, while not specifically designed for Nbs, have demonstrated substantial utility in predicting VHH structures due to their ability to model general sequence-structure relationships. AlphaFold2 utilizes multiple sequence alignment (MSA), residue pairing information, and structural templates for modeling [178], while RoseTTAFold employs a three-track neural network to simultaneously refine MSAs, inter-residue contacts, and 3D structures [209]. Emerging tools like OmegaFold and ESMFold further enhance modeling capabilities, especially for unique or poorly aligned sequences, by reducing dependence on MSA data [210, 211]. OpenFold, as an open-source implementation of AlphaFold, retains the core strengths of AlphaFold while offering open-source flexibility and improved efficiency, enabling further customization for specific research needs [212]. In antibody fragment modeling, IgFold takes a step further by focusing on the accurate prediction of long variable CDR regions, such as CDR3, demonstrating significant advantages in speed and precision [213]. Complementing this, ABodyBuilder3 adopts a broader scope, modeling general antibody structures while also accommodating nanobody-like fragments due to its diverse training dataset [214]. However, despite their strengths, their lack of optimization for the unique features of nanobodies, such as elongated CDR3 loops and compact β-sheet structures, limits their performance in nanobody-specific structure prediction.

Table 4 Comparison of Advanced AI models for Nb

Algorithm Name	Basic Principles/Methods	Basic Architecture	Main Uses	Prediction Accuracy/ Performance	Calculation Speed	Comprehensive Evaluation
AlphaFold2 [177]	Deep learning using multiple sequence alignment (MSA), residue pairing information and structural templates	Evoformer module processes inputs; Structure module predicts 3D coordinates	General protein structure prediction	Very high, approaching experimental structure accuracy	Slow	Extremely high accuracy Wide applicability High computational resource requirements
RoseTTAFold [209]	Three-track (1D, 2D, 3D) neural network processing MSAs, contacts and structure refinement	1D and 2D neural networks process sequence and contact information; 3D refines structure	General protein structure prediction	High, slightly lower than AlphaFold2	Medium	Faster than AlphaFold 2 Lower accuracy
NanoNet [183]	Deep learning using CNN and residual neural networks (ResNet)	Two 1D ResNets for frameworks/CDRs and inter-residue interactions	Specifically for VHH structure prediction	Very high for VHHs	Fast	Specifically optimized for VHHs High accuracy Fast
NbBuilder2 [215]	Deep learning based on AlphaFold-Multimer optimized for VHHs	Similar to AlphaFold-Multimer	Specifically for VHH structure prediction	Very high for VHHs	Fast	Specifically optimized for VHHs High accuracy GPU-accelerated
H3-OPT [221]	AlphaFold2 combined with pre-trained protein language model	Template module PLM-based structure prediction module	Nb structures prediction, especially CDR3 loops	Outperforming AlphaFold2 and IgFold	Slow	High CDR3 prediction accuracy Lower computational efficiency Difficulty with very long CDR-H3 loops
NanoBERTa-ASP [185]	Pretrained language model based on RoBERTa	Transformer architecture based on RoBERTa	Predict Nb antigen-binding sites from amino acid sequences	Outperforms existing methods	-	Direct prediction from sequence High interpretability Requires large labeled datasets for training
NbX [234]	Decision tree classifier for native-like poses	XGBoost decision tree	Re-rank Nb-antigen docking poses	Improved median ranking of native-like poses by 8-fold	Fast	Significantly improves ranking of native-like poses Computationally efficient Struggles to differentiate very similar poses
Nanobody Polyreactivity Predictor & Optimizer [240]	Data acquisition, model training, mutation design	Machine learning on yeast-displayed Nb sequences	Predict/reduce Nb polyreactivity	AUC > 0.8; Spearman $\rho \approx 0.77–0.79$	-	High accuracy Effective optimization User-friendly Limited to Nbs Lack of broader validation
Nanobody Affinity Classifier [223]	CNN learns features from Nb sequences	Typical CNN	Classify high/low affinity Nbs Identify key sequence features affecting affinity	92% accuracy for affinity classification	-	High accuracy Automatic feature extraction Lack of comparison with other methods
nanoBERT [263]	BERT-based transformer model	Transformer with bidirectional attention mechanism	Nb sequence infilling prediction/nativeness assessment/ Fine-tuning for downstream tasks	Outperforms human antibody-specific models and general protein models	-	Superior performance in Nb sequence prediction Applicable to various downstream tasks Lack of broader validation
ModiBodies [186]	Monte Carlo-based mutation analysis and molecular dynamics (MD) simulation for energy optimization	Energy optimization module with iterative Monte Carlo and MD simulation	Affinity and specificity optimization for nanobody-antigen complexes	Improves KD by up to fivefold in benchmarks	Medium	High efficiency in improving affinity and specificity; High-quality structural data and computational resource dependency

To address this, several specialized algorithms have been developed for nanobody modeling, offering higher precision and efficiency tailored to the unique features of Nbs. As the first deep learning structure prediction method tailored for VHHs, NanoNet employs a convolutional neural network (CNN) and two one-dimensional residual neural networks (ResNets) to directly generate 3D coordinates of the backbone and Cβ atoms from antibody sequences within milliseconds, greatly enhancing computational efficiency [184]. This makes it highly suitable for high-throughput analysis of large antibody libraries, providing valuable structural information for optimizing Nb stability, specificity, and rational design [184]. However, it still faces limitations in predicting longer CDR3 loops. In contrast, NanoBodyBuilder2 has achieved significant improvements in accuracy and reliability. As part of the ImmuneBuilder toolkit, NanoBodyBuilder2 utilizes a deep learning architecture tailored specifically for nanobodies, excelling particularly in the prediction of CDR3 loops [215]. Moreover, it introduces uncertainty estimation for each residue, providing crucial support for the interpretability and reliability of predictions. Most impressively, NanoBodyBuilder2 is 100 times faster than AlphaFold2, offering a tremendous advantage in scenarios requiring rapid generation of high-accuracy structural models, such as large-scale sequence screening and design [215].

Overall, NanoNet and NanoBodyBuilder2 demonstrate complementary strengths in efficient modeling and precise predictions, providing robust technical support for researchers in the field of nanobody engineering. In the future, further optimization and integration of these tools' features may drive innovative developments in nanobody design and application.

High-precision CDR3 structure prediction

CDR3 is the key region for Nbs to perform antigen recognition functions [182]. Predicting the structure of CDR3 remains a significant challenge. Compared to conventional antibodies, the CDR3 of Nbs possesses unique sequence and structural features, with lengths reaching up to 24 amino acid residues [216] and special conformations like loops, hairpins, and arms [47, 216], enabling CDR3 to penetrate the concave regions on the antigen surface and recognize hidden epitopes [217]. Moreover, the amino acid composition of CDR3 is highly variable, and certain specific sequence fragments may confer Nbs the ability to recognize special antigens [9]. Hence, it's of great significance to explore the unique sequence-structure-function relationships of CDR3 to provide guidance for subsequent Nb engineering [169, 218]. The performance of AlphaFold2, IgFold, ImmuneBuilder, and NanoNet was compared [219]. By analyzing the interaction between the predicted CDR3 and the Nbs framework and the orientation of CDR3 residues, it was found that the CDR3 conformations obtained by homology modeling, IgFold, and ImmuneBuilder were relatively similar and consistent with the stretched-twisted or convex conformations inferred from the length [219]. In contrast, the CDR3 conformations predicted by AlphaFold2 and NanoNet deviated more from the typical structures [219], suggesting that the two may not be well-suited for the task. The latest data shows that the median root mean square deviation (RMSD) for modeling CDR1, CDR2, and CDR3 by various models is 1.4–2.1Å, 0.8–1.5Å, and 2.5–4.7Å, respectively, with CDR3 having the highest RMSD [220]. As CDR3 length increases, RMSDs between AI-predicted CDR3 conformations and experimental structures generally increase, indicating decreased prediction accuracy [220]. It is necessary to focus on refining the training data for CDR3 conformations and improving deep learning algorithms. Fortunately, H3-OPT cleverly combines existing models' strengths, achieving a 2.24Å average RMSD on a high-quality Nb test set, outperforming other general or Nbs-specific models, such as AlphaFold2 (3.79Å), RoseTTAFold (3.75Å), NanoBodyBuilder2 (3.44Å), and NanoNet (4.37Å) [221]. The uniqueness of H3-OPT lies in its comprehensive utilization of template alignment, AlphaFold2 structural features, and the ESM2 language model, maximizing the advantages of different methods [221]. H3-OPT's exceptional performance in high-precision CDR3 structure prediction brings new hope for computational Nb design and optimization. Reliable CDR3 structures will deepen our understanding of Nbs' unique molecular structures, facilitating analysis of key interactions and epitopes with antigens, and providing important clues for subsequent antibody-antigen docking and property optimization.

Nb-antigen interaction prediction challenges

Predicting Nb-antigen binding sites and modes is challenging but crucial for understanding the specificity and antigen recognition mechanisms, which will guide vaccine design, drug development [222]. Nb-antigen interactions can be considered a special classification of traditional antibody-antigen interactions, and molecular docking approaches have been used to obtain structural models of a large number of Nb-antigen complexes [223]. Other representative algorithms include ClusPro [224] and HADDOCK [225], but high-throughput characterization of Nb-antigen interactions still faces challenges [184]. Computational methods often have issues such as high false-positive rates and difficulty in obtaining unique solutions when modeling the 3D structure of antibody-antigen complexes [184]. Although AlphaFold-Multimer has advanced protein complex modeling, it still struggles with critical challenges such as CDR modeling and glycosylated antigens, limiting its accuracy in

capturing Nb-antigen binding modes [226]. While Yin et al. specifically focused on Nb-antigen complexes and optimized AlphaFold-Multimer to achieve near-native accuracy in 50% of cases, this still leaves a substantial gap in reliability and precision for many Nb-antigen interactions [227]. Despite leveraging large-scale sampling strategies and incorporating Nb-specific features, such as the diversity of the CDR3 loop, their approach highlights both the progress made and the challenges that remain in creating a universally reliable predictive framework for Nb-antigen interactions [226, 227]. Another concern is that Nbs differ significantly from traditional antibodies in both sequence and structure [228], such as the longer and more diverse amino acid composition of the CDR3 loop [206, 207, 229].

To address those above challenges, a new tool called NanoBERTa-ASP has been proposed [186]. It is a sequence-based method aimed at directly predicting the antigen-binding sites of Nbs from amino acid sequences [230, 231]. NanoBERTa-ASP is based on the RoBERTa model architecture and introduces an attention mechanism that can focus on key regions in the sequence, such as the highly variable CDR3 region, capturing the structural features of Nbs [232]. By treating binding site prediction as a binary classification task, NanoBERTa-ASP can accurately predict whether each residue in the Nb sequence is a binding site [186]. It is worth noting that the realization of antibody-antigen interactions depends not only on the sequence features of the binding epitopes but also on the formation of suitable conformations between the antibody and antigen in 3D space [233]. Therefore, it is also necessary to consider the structural features and interaction patterns of Nb-antigen complexes [207]. NbX has been specifically designed to improve the reordering of docking conformations of Nb-antigen complexes [234]. Unlike NanoBERTa-ASP, which focuses on sequence features, NbX employs a set of structural features including energy, contact, and interface property. It learns the feature patterns of conformations similar to the native state by training a decision tree classifier, significantly improving the ranking performance of docking conformations [234]. On the test set, NbX increased the median ranking of native-like conformations by 8-fold, greatly outperforming existing docking algorithms such as ClusPro and deep learning-based protein interaction prediction methods like DOVENbX [224, 235]. Interestingly, by analyzing SHAP values, important features that contribute significantly to NbX's predictive performance were revealed. The proportion of CDR3 residues in the antigen-binding site was considered the most important feature by NbX, consistent with previous reports [206, 229]. Interface energy density and the hydrophobicity of the epitope and antigen-binding site were also considered important features [234]. These findings not only validate known Nb characteristics but also provide new design ideas.

In summary, NanoBERTa-ASP and NbX form a good complementary pair, with NanoBERTa-ASP utilizing sequence information to predict antigen-binding sites while NbX optimizing binding conformations at the structural level. The combination enables a comprehensive understanding from sequence to structure, guiding the computational design and optimization of Nbs at multiple levels and perspectives.

AI-guided nb polyreactivity optimization

Polyreactivity, the ability of a single antibody to bind multiple different antigens, contrasts with specificity and poses a troublesome issue for developers, leading to unfavorable pharmacokinetics and clinical application limitations [236–238] and increased development costs [239]. Currently, various experimental methods are available to assess antibody polyreactivity, but they have low throughput and require antibody purification [240]. AI provides a new perspective for understanding Nb polyreactivity. A set of supervised machine learning models, including logistic regression, CNNs and recurrent neural networks (RNNs), have been developed to predict the polyreactivity of Nbs from sequence data and quantitatively predict the impact of amino acid mutations on polyreactivity [240]. The specific contributions of different amino acids to polyreactivity at each position in the CDRs vary. For example, acidic amino acids in CDR2 and CDR3 are generally positively correlated with low polyreactivity, while arginine shows the opposite trend. However, at specific positions in CDR1, arginine also reduces polyreactivity, suggesting the existence of position-dependent effects [240]. Based on these findings, which provide a more detailed and comprehensive understanding of polyreactivity, researchers can design targeted amino acid substitutions to optimize antibodies. Taking the AT118i4h32 Nb as an example, which is a Nb antagonist of the angiotensin II type receptor (AT1R)37, researchers performed amino acid substitutions, reducing polyreactivity while maintaining high affinity and functionality for the target receptor [240]. This demonstrates the feasibility of optimizing polyreactivity through site-directed mutagenesis, removing a major obstacle for the rational design and clinical application of Nbs.

AI-guided nb physicochemical property optimization

Despite Nbs' advantageous low immunogenicity, high stability, and affinity for practical applications, performance optimization to further adapt to application purposes remains crucial before implementation [182, 241]. Various experimental strategies, such as gene fusion, enzyme conjugation, and chemical crosslinking, have been formulated to improve the application

characteristics of Nbs [9]. Specifically, traditional methods including chimerization [242], CDR grafting [243, 244], specificity determining region grafting [245, 246], scaffold screening [247], and surface replacement [248, 249] have been used to enhance antibody humanization. Nbs have been developed into multivalent and biparatopic forms through covalent linking [250–254], and stability mutations targeting the CDR framework junctions have been designed [255]. Affinity maturation has been achieved by substituting amino acids in the VHH-CDR3 region [256]. Multivalent binding forms [257, 258] or the introduction of albumin-nucleic acid structures [259] have also led to improved biostability and serum half-life.

Technological advancements have led to new computer and AI-based strategies to address the inefficiency of traditional experimental methods. Protein electrostatics play a crucial role in regulating antibody stability, dimerization, complex formation, and interactions with other molecules [260]. By modifying the charged amino acid residues on Nbs, their affinity and selectivity can be modulated [261]. Cheng et al. used computational affinity maturation to optimize an anti-CD47 Nb, yielding four mutations with improved binding and thermal stability [262]. This demonstrates that electrostatic engineering combined with other design strategies can achieve better protein performance. However, when introducing charged residues through site-directed mutagenesis to enhance the binding affinity of Nbs, it is important to balance their impact on structural stability [260]. Furthermore, deep learning has advanced this field. Xiang et al. established a deep learning model capable of accurately learning key amino acid combination features associated with binding affinity from massive antigen-specific Nb omics data [223], providing a new perspective for understanding the high-affinity binding mechanism between Nbs and antigens. The length and charge distribution of the CDR3 region, as well as the charge and position of amino acids and other structural bases, significantly influence affinity [223]. It is worth noting that for infilling Nb sequences, the nanoBERT model has been proposed [263], which is a transformer model specifically designed for Nbs. By learning the evolutionary diversity of Nbs, this model can predict the feasibility of amino acid mutations, enabling the exploration and evaluation of the impact of these mutations on the structure and function of Nbs to improve stability and reduce immunogenicity. In this context, the computational model ModiBodies has demonstrated its effectiveness in enhancing binding affinity by systematically identifying and mutating hypervariable residues within CDR regions [187]. Through molecular dynamics simulations and iterative energy optimization, it achieves improved antigen-binding properties while maintaining structural stability [187]. Additionally, using combinatorial algorithms can also achieve substantial improvements in the affinity and stability of Nbs [264]. Melting temperature (Tm) is another critical parameter in Nb engineering, reflecting their structural integrity under thermal stress and directly influencing stability and functionality in diverse applications [182]. Building on databases like The NbThermo [194] which contains Tm data for 564 nanobodies, Alvarez et al. developed the TEMPRO tool, which employs protein embedding techniques and deep learning algorithms to achieve high-accuracy Tm predictions, significantly reducing the time and cost of experimental measurements [265]. These advancements enable the rapid identification of thermally stable nanobody candidates and lay the groundwork for integrating tools to optimize multiple traits, such as solubility and binding affinity, thereby broadening the scope of nanobody engineering.

Is AI-assisted nb engineering ready?

Despite the significant progress made by computational methods in Nb structure prediction and design, we must remain clear-headed because those cannot replace the entire Nb development process. Currently, the vast majority of Nb development is still laboratory-based, including animal immunization and mature in vitro display technologies [169]. However, we must acknowledge the tremendous value brought by AI technologies, which guide rational Nbs design and optimization while maximizing time and cost savings [266]. Structure prediction, utilizing AI, offers profound insights into Nb, elucidates antigen-antibody binding mechanisms, and facilitates optimization of binding affinities and physicochemical properties. This approach provides crucial guidance for experimental and clinical translation, positioning it as an indispensable component in future Nb design and development processes. At the same time, it is gratifying to see that public data related to Nbs is growing rapidly and becoming more diverse [266], and the construction of a more comprehensive Nb sequence and structure database is being realized. This will serve as a solid cornerstone, providing higher-quality training data for deep learning models. Based on this data, more accurate AI models can be trained, further improving prediction accuracy. Moreover, multimodal learning methods that integrate various data are expected to reveal new insights into the relationship between sequence, structure and function [267], promoting the maturity of computational antibody design methods.

However, this will inevitably bring new challenges. On one hand, computationally designed proteins have potential immunogenicity issues [221], which require more comprehensive preclinical and clinical studies. On the other hand, the standardization, normalization, and determinability of data collection and management are also worth attention. To achieve effective management

and utilization of data from different sources and formats, unified data standards must be established to improve data interoperability [268]. Moreover, with the rapid development of AI technologies, exploring the application of novel machine learning architectures in Nb design and enabling researchers without a background in computer programming to understand and skillfully use AI models are also worth broader attention.

Conclusion and perspective

Nbs, derived from camelid HCAbs or cartilaginous fish NARs, have emerged as potent tools for disease diagnosis and treatment due to their unique properties, such as small size, high stability, and excellent antigen-binding specificity [1, 2, 8]. Currently, three main types of Nb libraries - immune, naïve, and synthetic/semi-synthetic - are constructed through various strategies, each offering distinct advantages [9, 10]. Immune libraries provide affinity-matured, target-specific Nbs but require animal immunization, while naïve and synthetic libraries circumvent this need and are suitable for non-immunogenic targets, with synthetic libraries offering the greatest diversity [28, 29, 34].

To effectively screen and optimize Nbs, efficient display platforms are crucial. Phage display, being the most widely used biological platform, can present a wide range of protein diversity [58, 59]. Yeast and bacterial display systems offer the advantages of eukaryotic post-translational modifications and compatibility with fluorescence-activated cell sorting (FACS) screening [71, 74, 84]. Mammalian, plant, and insect cell expression systems can provide more complex modifications and human-like glycosylation patterns [111, 119, 129]. Moreover, Cell-free display methods, such as ribosome display and mRNA display, enable rapid selection cycles and in vitro protein evolution [142, 143]. We are also excited to witness the emergence of novel technologies like Nestlink, Sybody, and the combination of high-throughput sequencing and mass spectrometry, which signify the ongoing revolution in Nb discovery and characterization methods [164, 166, 167].

As display platforms advance, AI holds promise in overcoming the limitations of traditional Nb engineering approaches. Deep learning models have achieved remarkable success in protein structure prediction [178, 209], with Nb-specific algorithms like NanoNet and NbBuilder2 further optimizing VHH structure prediction [184, 215]. Although predicting Nb-antigen interactions remains challenging, tools like NanoBERTa-ASP and NbX are making significant progress [186, 234]. Furthermore, AI is being applied to optimize the physicochemical properties of Nbs, such as polyreactivity, affinity, and stability [223, 240, 262]. As AI continues to progress, it will play an increasingly important role in Nb engineering, spanning from structure prediction and antigen interaction modeling to property optimization and de novo design.

Looking to the future, the continuous development and optimization of Nb libraries, display platforms, and AI-assisted design will be key to driving Nb applications. Synthetic and semi-synthetic libraries are expected to become more complex by incorporating novel scaffolds and design strategies to enhance their diversity and functionality, such as integrating computationally optimized frameworks and using advanced mutagenesis techniques like error-prone PCR and DNA shuffling [155, 156]. Moreover, the integration of multiple display methods and the development of hybrid approaches may further improve screening efficiency and enable the identification of rare, high-performance clones [83].

Notably, AI will play an increasingly important role in Nb engineering, from structure prediction and antigen interaction modeling to property optimization and de novo design. The growing availability of Nb sequence and structural data will fuel the development of more accurate and versatile AI models. However, challenges such as data standardization, immunogenicity concerns, and the need for more user-friendly AI tools will need to be addressed [268]. Establishing unified data standards and improving interoperability will be essential for the effective utilization of the ever-expanding Nb data [268]. Immunogenicity concerns associated with computationally designed Nbs will require comprehensive preclinical and clinical studies to ensure their safety and efficacy. The development of intuitive, user-friendly AI tools will be crucial for enabling researchers without a background in computer programming to leverage these powerful technologies.

In summary, the synergistic combination of advanced library construction, display technologies, and AI-guided design will accelerate the development of Nbs as powerful tools for research, diagnostics, and therapeutics. As these three key areas continue to evolve and integrate, Nbs are poised to usher in a new era of precision medicine and biotechnology, offering unprecedented opportunities for targeted disease diagnosis, treatment, and beyond.

Abbreviations

Nb	Nanobody
VHH	Variable domain of the Heavy chain of Heavy-chain antibodies
ScFv	single-chain fragment variable
HCAbs	Heavy Chain-only Antibodies
NARs	Novel Antigen Receptors
CDR	Complementarity-Determining Regions
FR	Framework Regions
AI	Artificial Intelligence
NGS	Next Generation Sequencing
PIII	Protein III
PVIII	Protein VIII
S. cerevisiae	Saccharomyces cerevisiae

MACS	Magnetic-activated cell sorting
FACS	Fluorescence-activated cell sorting
PTM	Post-translational modification
Fc	Fragment crystallizable
E.coli	Escherichia coli
K_D	Dissociation constant
Fab	Antigen-binding fragments
OmpF	Outer membrane F
PEG	Polyethylene Glycol
GRAS	Generally Recognized as Safe
TGE	Transient gene expression
RMCE	Recombinase-mediated cassette exchange
SDS-PAGE	Sodium dodecyl sulfate - polyacrylamide gel electrophoresis
SPR	Surface plasmon resonance
BES	Baculovirus Expression System
AcMNPV	Autographa californica multiple nuclear polyhedrovirus
BEVS-IC	Insect cell Baculovirus Expression Vector System
PRM	Protein-ribosome-mRNA
Rnase	Ribonuclease
LC-MS	Liquid Chromatograph-Mass Spectrometer
GNNs	Graph Neural Networks
MSA	Multiple sequence alignment
CNN	Convolutional neural network
ResNets	Residual neural networks
RMSD	Root mean square deviation

Acknowledgements
Not applicable.

Author contributions
J.L., L.W. and A.X. wrote the main manuscript text. A.X. prepared Figs. 1 and 2; Table 1. L.W. prepared graphic abstract, Figs. 3 and 4; Table 2. J.L. prepared Fig. 5; Table 3, and 4. All authors reviewed the manuscript.

Funding
This work was supported by the Natural Science Foundation of Jiangsu Province (BK20210068), the Mega-project of Wuxi Commission of Health (Z202216), the Top Talent Support Program for Young, Middle-aged People of Wuxi Health Commission (BJ2023014) and National Cancer Institute R01CA230339 and R37CA255948.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Thoracic Surgery, Wuxi People's Hospital, Wuxi Medical Center, The Affiliated Wuxi People's Hospital of Nanjing Medical University, Nanjing Medical University, Wuxi 214023, China
[2]Wuxi College of Clinical Medicine, Nanjing Medical University, Wuxi 214023, China
[3]The Pq Laboratory of BiomeDx/Rx, Department of Biomedical Engineering, Binghamton University, Binghamton 13850, USA
[4]Department of Thoracic Surgery, Wuxi People's Hospital, Wuxi Medical Center, The Affiliated Wuxi People's Hospital of Nanjing Medical University, Nanjing Medical University, No. 299 Qingyang Rd., Wuxi 214023, China
[5]Department of Biomedical Engineering, The Pq Laboratory of BiomeDx/Rx, Binghamton University, Binghamton, NY 13902, USA

Received: 9 October 2024 / Accepted: 27 January 2025
Published online: 06 February 2025

References

1. Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hamers C, Songa EB, et al. Naturally occurring antibodies devoid of light chains. Nature. 1993;363:446–8.
2. Dooley H, Flajnik MF. Antibody repertoire development in cartilaginous fish. Dev Comp Immunol. 2006;30:43–56.
3. Stanfield RL, Dooley H, Flajnik MF, Wilson IA. Crystal structure of a shark single-domain antibody V region in complex with lysozyme. Science. 2004;305:1770–3.
4. Salvador J-P, Vilaplana L, Marco M-P. Nanobody: outstanding features for diagnostic and therapeutic applications. Anal Bioanal Chem. 2019;411:1703–13.
5. Kunz S, Durandy M, Seguin L, Feral CC. NANOBODY® Molecule, a Giga Medical Tool in Nanodimensions. Int J Mol Sci. 2023;24:13229.
6. Muyldermans S, Cambillau C, Wyns L. Recognition of antigens by single-domain antibody fragments: the superfluous luxury of paired domains. Trends Biochem Sci. 2001;26:230–5.
7. Harmsen MM, Ruuls RC, Nijman IJ, Niewold TA, Frenken LG, de Geus B. Llama heavy-chain V regions consist of at least four distinct subfamilies revealing novel sequence features. Mol Immunol. 2000;37:579–90.
8. De Genst E, Silence K, Decanniere K, Conrath K, Loris R, Kinne J, et al. Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. Proc Natl Acad Sci U S A. 2006;103:4586–91.
9. Muyldermans S. A guide to: generation and design of nanobodies. FEBS J. 2021;288:2084–102.
10. Frecot DI, Froehlich T, Rothbauer U. 30 years of nanobodies - an ongoing success story of small binders in biological research. J Cell Sci. 2023;136:jcs261395.
11. Wagner TR, Rothbauer U. Nanobodies– Little helpers unravelling intracellular signaling. Free Radic Biol Med. 2021;176:46–61.
12. Wagner TR, Blaess S, Leske IB, Frecot DI, Gramlich M, Traenkle B, et al. Two birds with one stone: human SIRPα nanobodies for functional modulation and in vivo imaging of myeloid cells. Front Immunol. 2023;14:1264179.
13. R C, S G, W C. Nanobody: the magic bullet. Mol Imaging? Theranostics. 2014;4.
14. Yang Y, Li F, Du L. Therapeutic nanobodies against SARS-CoV-2 and other pathogenic human coronaviruses. J Nanobiotechnol. 2024;22:304.
15. Minatel VM, Prudencio CR, Barraviera B, Ferreira RS. Nanobodies: a promising approach to treatment of viral diseases. Front Immunol. 2023;14:1303353.
16. Li D, Wang R, Liang T, Ren H, Park C, Tai C-H, et al. Camel nanobody-based B7-H3 CAR-T cells show high efficacy against large solid tumours. Nat Commun. 2023;14:5920.
17. Aj ER-L. S. Transportation of single-domain antibodies through the blood-brain barrier. Biomolecules. 2021;11.
18. Custódio TF, Das H, Sheward DJ, Hanke L, Pazicky S, Pieprzyk J, et al. Selection, biophysical and structural analysis of synthetic nanobodies that effectively neutralize SARS-CoV-2. Nat Commun. 2020;11:5588.
19. Li J, Kang G, Wang J, Yuan H, Wu Y, Meng S, et al. Affinity maturation of antibody fragments: a review encompassing the development from random approaches to computational rational optimization. Int J Biol Macromol. 2023;247:125733.
20. Vincke C, Gutiérrez C, Wernery U, Devoogdt N, Hassanzadeh-Ghassabeh G, Muyldermans S. Generation of single domain antibody fragments derived from camelids and generation of manifold constructs. Methods Mol Biol Clifton NJ. 2012;907:145–76.
21. Pardon E, Laeremans T, Triest S, Rasmussen SGF, Wohlkönig A, Ruf A, et al. A general protocol for the generation of Nanobodies for structural biology. Nat Protoc. 2014;9:674–93.
22. Zupancic JM, Smith MD, Trzeciakiewicz H, Skinner ME, Ferris SP, Makowski EK, et al. Quantitative flow cytometric selection of tau conformational nanobodies specific for pathological aggregates. Front Immunol. 2023;14:1164080.
23. Güttler T, Aksu M, Dickmanns A, Stegmann KM, Gregor K, Rees R, et al. Neutralization of SARS-CoV-2 by highly potent, hyperthermostable, and mutation-tolerant nanobodies. EMBO J. 2021;40:e107985.
24. Rw RJSD, van HGPA, Jk R. S, Generation of heavy-chain-only antibodies in mice. Proc Natl Acad Sci U S A. 2006;103(41):15130–5.
25. Y T, Jl Y. B E, P H, L T, C J, Diverse human VH antibody fragments with bio-therapeutic properties from the Crescendo mouse. New Biotechnol. 2020;55:65–76.
26. Fan Q, Zhao R, Chen Y, Chi L, Huang Y, Liu M, et al. Affinity purification of mAb from serum-containing hybridoma culture supernatant through a novel nanobody that discriminates mouse IgG from bovine IgG by recognizing the mouse kappa constant region (mCK). J Chromatogr A. 2024;1724:464929.

27. Sabir JSM, Atef A, El-Domyati FM, Edris S, Hajrah N, Alzohairy AM, et al. Construction of naïve camelids VHH repertoire in phage display-based library. C R Biol. 2014;337:244–9.
28. Yan J, Wang P, Zhu M, Li G, Romão E, Xiong S, et al. Characterization and applications of nanobodies against human procalcitonin selected from a novel naïve Nanobody phage display library. J Nanobiotechnol. 2015;13:33.
29. Olichon A, de Marco A. Preparation of a naïve library of camelid single domain antibodies. Methods Mol Biol Clifton NJ. 2012;911:65–78.
30. Lee HE, Cho AH, Hwang JH, Kim JW, Yang HR, Ryu T et al. Development, high-throughput profiling, and Biopanning of a large phage Display single-domain antibody Library. Int J Mol Sci. 2024;25(9):4791.
31. Ouyang Q, Wang L, Nasser IDE, Deng G, Zhang XK, You T, et al. [Preparation of interleukin-1β-targeted nanobodies and their effects on apoptosis in hypoxic cardiomyocytes of mice]. Zhonghua Yi Xue Za Zhi. 2024;104:2066–73.
32. Hoseinpoor R, Mousavi Gargari SL, Rasooli I, Rajabibazl M, Shahi B. Functional mutations in and characterization of VHH against Helicobacter pylori urease. Appl Biochem Biotechnol. 2014;172:3079–91.
33. Lu Y, Li Q, Fan H, Liao C, Zhang J, Hu H, et al. A multivalent and thermostable Nanobody Neutralizing SARS-CoV-2 Omicron (B.1.1.529). Int J Nanomed. 2023;18:353–67.
34. Kunz P, Flock T, Soler N, Zaiss M, Vincke C, Sterckx Y, et al. Exploiting sequence and stability information for directing nanobody stability engineering. Biochim Biophys Acta Gen Subj. 2017;1861:2196–205.
35. Ms V-T, Me V-T, E M-A EM. NbThermo: a new thermostability database for nanobodies. Database J Biol Databases Curation. 2023;2023.
36. Valdés-Tresanco MS, Molina-Zapata A, Pose AG, Moreno E. Structural insights into the design of synthetic nanobody libraries. Mol Basel Switz. 2022;27:2198.
37. Liu B, Yang D. Easily established and multifunctional synthetic nanobody libraries as Research Tools. Int J Mol Sci. 2022;23:1482.
38. Saerens D, Pellis M, Loris R, Pardon E, Dumoulin M, Matagne A, et al. Identification of a universal VHH framework to graft non-canonical antigen-binding loops of camel single-domain antibodies. J Mol Biol. 2005;352:597–607.
39. Ju M-S, Min S-W, Lee SM, Kwon HS, Park JC, Lee JC, et al. A synthetic library for rapid isolation of humanized single-domain antibodies. Biotechnol Bioprocess Eng. 2017;22:239–47.
40. Moutel S, Bery N, Bernard V, Keller L, Lemesre E, de Marco A, et al. NaLi-H1: a universal synthetic library of humanized nanobodies providing highly functional antibodies and intrabodies. eLife. 2016;5:e16228.
41. Lagoutte P, Bourhis J-M, Mariano N, Gueguen-Chaignon V, Vandroux D, Moali C, et al. Mono- and bi-specific nanobodies targeting the CUB domains of PCPE-1 reduce the Proteolytic Processing of Fibrillar Procollagens. J Mol Biol. 2024;436:168667.
42. Chen X, Gentili M, Hacohen N, Regev A. A cell-free nanobody engineering platform rapidly generates SARS-CoV-2 neutralizing nanobodies. Nat Commun. 2021;12:5506.
43. McMahon C, Baier AS, Pascolutti R, Wegrecki M, Zheng S, Ong JX, et al. Yeast surface display platform for rapid discovery of conformationally selective nanobodies. Nat Struct Mol Biol. 2018;25:289–96.
44. Jang H, Wilson PG, Sau M, Chawla U, Rodgers DW, Galperin E. Single-domain antibodies for functional targeting of the signaling scaffold Shoc2. Mol Immunol. 2020;118:110–6.
45. Ferrari D, Garrapa V, Locatelli M, Bolchi A. A Novel Nanobody Scaffold optimized for bacterial expression and suitable for the construction of Ribosome Display libraries. Mol Biotechnol. 2020;62:43–55.
46. Contreras MA, Serrano-Rivero Y, González-Pose A, Salazar-Uribe J, Rubio-Carrasquilla M, Soares-Alves M, et al. Design and construction of a synthetic Nanobody Library: testing its potential with a single selection round strategy. Mol Basel Switz. 2023;28:3708.
47. Zimmermann I, Egloff P, Hutter CA, Arnold FM, Stohler P, Bocquet N, et al. Synthetic single domain antibodies for the conformational trapping of membrane proteins. eLife. 2018;7:e34317.
48. Yuan TZ, Garg P, Wang L, Willis JR, Kwan E, Hernandez AGL, et al. Rapid discovery of diverse neutralizing SARS-CoV-2 antibodies from large-scale synthetic phage libraries. mAbs. 2022;14:2002236.
49. Kilisch M, Götzke H, Gere-Becker M, Crauel A, Opazo F, Frey S. Discovery and characterization of an ALFA-Tag-Specific Affinity Resin optimized for protein purification at low temperatures in physiological buffer. Biomolecules. 2021;11:269.
50. Cornish-Bowden A. Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. Nucleic Acids Res. 1985;13:3021–30.
51. Zimmermann I, Egloff P, Hutter CAJ, Kuhn BT, Bräuer P, Newstead S, et al. Generation of synthetic nanobodies against delicate proteins. Nat Protoc. 2020;15:1707–41.
52. Zhao Y, Wang Y, Su W, Li S. Construction of Synthetic Nanobody Library in mammalian cells by dsDNA-Based Strategies*. Chembiochem Eur J Chem Biol. 2021;22:2957–65.
53. Misson Mindrebo L, Liu H, Ozorowski G, Tran Q, Woehl J, Khalek I, et al. Fully synthetic platform to rapidly generate tetravalent bispecific nanobody-based immunoglobulins. Proc Natl Acad Sci U S A. 2023;120:e2216612120.
54. Chi X, Zhang X, Pan S, Yu Y, Shi Y, Lin T, et al. An ultrapotent RBD-targeted biparatopic nanobody neutralizes broad SARS-CoV-2 variants. Signal Transduct Target Ther. 2022;7:44.
55. Liu Y, Huang H. Expression of single-domain antibody in different systems. Appl Microbiol Biotechnol. 2018;102:539–51.
56. Jaroszewicz W, Morcinek-Orłowska J, Pierzynowska K, Gaffke L, Węgrzyn G. Phage display and other peptide display technologies. FEMS Microbiol Rev. 2022;46:fuab052.
57. Smith GP. Filamentous Fusion Phage: Novel expression vectors that display cloned antigens on the Virion Surface. Science. 1985;228:1315–7.
58. Sioud M. Phage display libraries: from binders to targeted drug delivery and human therapeutics. Mol Biotechnol. 2019;61:286–303.
59. Jin B-K, Odongo S, Radwanska M, Magez S, NANOBODIES®. A review of diagnostic and therapeutic applications. Int J Mol Sci. 2023;24:5994.
60. Castel G, Chtéoui M, Heyd B, Tordo N. Phage Display of combinatorial peptide libraries: application to Antiviral Research. Molecules. 2011;16:3499–518.
61. Ledsgaard L, Ljungars A, Rimbault C, Sørensen CV, Tulika T, Wade J, et al. Advances in antibody phage display technology. Drug Discov Today. 2022;27:2151–69.
62. Yeboah M, Papagregoriou C, Jones DC, Chan HTC, Hu G, McPartlan JS, et al. LILRB3 (ILT5) is a myeloid cell checkpoint that elicits profound immunomodulation. JCI Insight. 2020;5:e141593.
63. Even-Desrumeaux K, Nevoltris D, Lavaut MN, Alim K, Borg J-P, Audebert S, et al. Masked selection: a straightforward and flexible approach for the selection of binders against specific epitopes and differentially expressed proteins by phage display. Mol Cell Proteom MCP. 2014;13:653–65.
64. Pasqualini R, Ruoslahti E. Organ targeting in vivo using phage display peptide libraries. Nature. 1996;380:364–6.
65. Hess KL, Jewell CM. Phage display as a tool for vaccine and immunotherapy development. Bioeng Transl Med. 2020;5:e10142.
66. Marintcheva B. Harnessing the power of viruses. London; San Diego, CA: Academic Press, an imprint of Elsevier; 2018.
67. Shim H. Antibody phage Display. In: Lim TS, editor. Recombinant antibodies for infectious diseases. Cham: Springer International Publishing; 2017. pp. 21–34.
68. Qasemi M, Behdani M, Shokrgozar MA, Molla-Kazemiha V, Mohseni-Kuchesfahani H, Habibi-Anbouhi M. Construction and expression of an anti-VEGFR2 Nanobody-Fc fusionbody in NS0 host cell. Protein Expr Purif. 2016;123:19–25.
69. Malaquias ADM, Marques LEC, Pereira SS, de Freitas Fernandes C, Maranhão AQ, Stabeli RG, et al. A review of plant-based expression systems as a platform for single-domain recombinant antibody production. Int J Biol Macromol. 2021;193:1130–7. Pt B.
70. Weng D, Yang L, Xie Y. Engineering and characterization of GFP-targeting nanobody: expression, purification, and post-translational modification analysis. Protein Expr Purif. 2024;221:106501.
71. Sheehan J, Marasco WA. Phage and yeast Display. Microbiol Spectr. 2015;3:AID–0028.
72. Hoogenboom HR. Selecting and screening recombinant antibody libraries. Nat Biotechnol. 2005;23:1105–16.
73. Ledsgaard L, Kilstrup M, Karatt-Vellatt A, McCafferty J, Laustsen AH. Basics of antibody phage Display Technology. Toxins. 2018;10:236.
74. Li Y, Wang X, Zhou N-Y, Ding J. Yeast surface display technology: mechanisms, applications, and perspectives. Biotechnol Adv. 2024;76:108422.
75. Gera N, Hussain M, Rao BM. Protein selection using yeast surface display. Methods San Diego Calif. 2013;60:15–26.
76. Mustafa MI, Alzebair AA, Mohammed A. Development of recombinant antibody by yeast surface Display Technology. Curr Res Pharmacol Drug Discov. 2024;6:100174.
77. Mei M, Zhou Y, Peng W, Yu C, Ma L, Zhang G, et al. Application of modified yeast surface display technologies for non-antibody protein engineering. Microbiol Res. 2017;196:118–28.
78. Orfao A, Ruiz-Arguelles A. General concepts about cell sorting techniques. Clin Biochem. 1996;29:5–9.

79. De Pourcq K, De Schutter K, Callewaert N. Engineering of glycosylation in yeast and other fungi: current state and perspectives. Appl Microbiol Biotechnol. 2010;87:1617–31.
80. Mahdavi SZB, Oroojalian F, Eyvazi S, Hejazi M, Baradaran B, Pouladi N, et al. An overview on display systems (phage, bacterial, and yeast display) for production of anticancer antibodies; advantages and disadvantages. Int J Biol Macromol. 2022;208:421–42.
81. De Meyer T, Laukens B, Nolf J, Van Lerberge E, De Rycke R, De Beuckelaer A, et al. Comparison of VHH-Fc antibody production in Arabidopsis thaliana, Nicotiana benthamiana and Pichia pastoris. Plant Biotechnol J. 2015;13:938–47.
82. Holland K, Blazeck J. High throughput mutagenesis and screening for yeast engineering. J Biol Eng. 2022;16:37.
83. Ferrara F, Soluri MF, Sblattero D. Recombinant antibody selections by combining phage and yeast Display. Methods Mol Biol Clifton NJ. 2019;1904:339–52.
84. Salema V, Fernández LÁ. Escherichia coli surface display for the selection of nanobodies. Microb Biotechnol. 2017;10:1468–84.
85. Salema V, Marín E, Martínez-Arteaga R, Ruano-Gallego D, Fraile S, Margolles Y, et al. Selection of single domain antibodies from immune libraries displayed on the surface of E. Coli cells with two β-domains of opposite topologies. PLoS ONE. 2013;8:e75126.
86. Francisco JA, Campbell R, Iverson BL, Georgiou G. Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface. Proc Natl Acad Sci U S A. 1993;90:10444–8.
87. Kang TH, Seong BL. Solubility, Stability, and avidity of recombinant antibody fragments expressed in microorganisms. Front Microbiol. 2020;11:1927.
88. Missiakas D, Betton JM, Raina S. New components of protein folding in extracytoplasmic compartments of Escherichia coli SurA, FkpA and Skp/OmpH. Mol Microbiol. 1996;21:871–84.
89. Shriver-Lake LC, Goldman ER, Zabetakis D, Anderson GP. Improved production of single domain antibodies with two disulfide bonds by co-expression of chaperone proteins in the Escherichia coli periplasm. J Immunol Methods. 2017;443:64–7.
90. Åslund F, Berndt KD, Holmgren A. Redox Potentials of Glutaredoxins and other Thiol-Disulfide oxidoreductases of the Thioredoxin Superfamily determined by Direct Protein-Protein Redox Equilibria. J Biol Chem. 1997;272:30780–6.
91. Sandomenico A, Sivaccumar JP, Ruvo M. Evolution of Escherichia coli expression system in producing antibody recombinant fragments. Int J Mol Sci. 2020;21:6324.
92. Gupta SK, Shukla P. Microbial platform technology for recombinant antibody fragment production: a review. Crit Rev Microbiol. 2017;43:31–42.
93. Chen T, Wang K, Chi X, Zhou L, Li J, Liu L, et al. Construction of a bacterial surface display system based on outer membrane protein F. Microb Cell Factories. 2019;18:70.
94. Palma JA, Bunyatov MI, Hulbert SW, Jewett MC, DeLisa MP. Bacterial glycoengineering: cell-based and cell-free routes for producing biopharmaceuticals with customized glycosylation. Curr Opin Chem Biol. 2024;81:102500.
95. Meyer V, Basenko EY, Benz JP, Braus GH, Caddick MX, Csukai M, et al. Growing a circular economy with fungal biotechnology: a white paper. Fungal Biol Biotechnol. 2020;7:5.
96. Nevalainen H, Peterson R, Curach N. Overview of Gene expression using filamentous Fungi. Curr Protoc Protein Sci. 2018;92:e55.
97. Deshpande N, Wilkins MR, Packer N, Nevalainen H. Protein glycosylation pathways in filamentous fungi. Glycobiology. 2008;18:626–37.
98. Su X, Schmitz G, Zhang M, Mackie RI, Cann IKO. Heterologous gene expression in filamentous fungi. Adv Appl Microbiol. 2012;81:1–61.
99. Magaña-Ortíz D, Coconi-Linares N, Ortiz-Vazquez E, Fernández F, Loske AM, Gómez-Lim MA. A novel and highly efficient method for genetic transformation of fungi employing shock waves. Fungal Genet Biol FG B. 2013;56:9–16.
100. Sun Z, Wu Y, Long S, Feng S, Jia X, Hu Y, et al. Aspergillus oryzae as a cell factory: research and applications in Industrial Production. J Fungi Basel Switz. 2024;10:248.
101. Liu D, Garrigues S, de Vries RP. Heterologous protein production in filamentous fungi. Appl Microbiol Biotechnol. 2023;107:5019–33.
102. Messer A, Butler DC. Optimizing intracellular antibodies (intrabodies/nanobodies) to treat neurodegenerative disorders. Neurobiol Dis. 2020;134:104619.
103. Rashid MH. Full-length recombinant antibodies from Escherichia coli: production, characterization, effector function (fc) engineering, and clinical evaluation. mAbs. 2022;14:2111748.
104. Li Q, Lu J, Zhang G, Liu S, Zhou J, Du G, et al. Recent advances in the development of aspergillus for protein production. Bioresour Technol. 2022;348:126768.
105. Van Den Hombergh JPTW, Van De Vondervoort PJI, Fraissinet-Tachet L, Visser J. Aspergillus as a host for heterologous protein production: the problem of proteases. Trends Biotechnol. 1997;15:256–63.
106. Fu H, Liang Y, Zhong X, Pan Z, Huang L, Zhang H, et al. Codon optimization with deep learning to enhance protein expression. Sci Rep. 2020;10:17617.
107. Sun X, Su X. Harnessing the knowledge of protein secretion for enhanced protein production in filamentous fungi. World J Microbiol Biotechnol. 2019;35:54.
108. Driouch H, Roth A, Dersch P, Wittmann C. Optimized bioprocess for production of fructofuranosidase by recombinant aspergillus Niger. Appl Microbiol Biotechnol. 2010;87:2011–24.
109. Aerts D, Hauer EE, Ohm RA, Arentshorst M, Teertstra WR, Phippen C, et al. The FlbA-regulated predicted transcription factor Fum21 of Aspergillus Niger is involved in fumonisin production. Antonie Van Leeuwenhoek. 2018;111:311–22.
110. Sakekar AA, Gaikwad SR, Punekar NS. Protein expression and secretion by filamentous fungi. J Biosci. 2021;46:5.
111. SECRETERS - European Union's Horizon 2020 Programme. Electronic address: l.rettenbacher@kent.ac.uk, SECRETERS - European Union's Horizon 2020 Programme. Microbial protein cell factories fight back? Trends Biotechnol. 2022;40:576–90.
112. Bandaranayake AD, Almo SC. Recent advances in mammalian protein production. FEBS Lett. 2014;588:253–60.
113. Gutiérrez-Granados S, Cervera L, Kamen AA, Gòdia F. Advancements in mammalian cell transient gene expression (TGE) technology for accelerated production of biologics. Crit Rev Biotechnol. 2018;38:918–40.
114. O'Flaherty R, Bergin A, Flampouri E, Mota LM, Obaidi I, Quigley A, et al. Mammalian cell culture for production of recombinant proteins: a review of the critical steps in their biomanufacturing. Biotechnol Adv. 2020;43:107552.
115. Xu W-J, Lin Y, Mi C-L, Pang J-Y, Wang T-Y. Progress in fed-batch culture for recombinant protein production in CHO cells. Appl Microbiol Biotechnol. 2023;107:1063–75.
116. Ritter A, Voedisch B, Wienberg J, Wilms B, Geisse S, Jostock T, et al. Deletion of a telomeric region on chromosome 8 correlates with higher productivity and stability of CHO cell lines. Biotechnol Bioeng. 2016;113:1084–93.
117. Eisenhut P, Marx N, Borsi G, Papež M, Ruggeri C, Baumann M, et al. Manipulating gene expression levels in mammalian cell factories: an outline of synthetic molecular toolboxes to achieve multiplexed control. New Biotechnol. 2024;79:1–19.
118. Tejwani V, Andersen MR, Nam JH, Sharfstein ST. Glycoengineering in CHO cells: advances in Systems Biology. Biotechnol J. 2018;13:e1700234.
119. Rattanapisit K, Phakham T, Buranapraditkun S, Siriwattananon K, Boonkrai C, Pisitkun T, et al. Structural and in Vitro functional analyses of Novel Plant-Produced Anti-human PD1 antibody. Sci Rep. 2019;9:15205.
120. Dobhal S, Chaudhary VK, Singh A, Pandey D, Kumar A, Agrawal S. Expression of recombinant antibody (single chain antibody fragment) in transgenic plant Nicotiana tabacum Cv. Xanthi Mol Biol Rep. 2013;40:7027–37.
121. Satheeshkumar PK. Expression of single chain variable fragment (scFv) molecules in plants: a Comprehensive Update. Mol Biotechnol. 2020;62:151–67.
122. Zimmermann J, Saalbach I, Jahn D, Giersberg M, Haehnel S, Wedel J, et al. Antibody expressing pea seeds as fodder for prevention of gastrointestinal parasitic infections in chickens. BMC Biotechnol. 2009;9:79.
123. Wang D, Ma J, Sun D, Li H, Jiang C, Li X. Expression of bioactive anti-CD20 antibody fragments and induction of ER stress response in Arabidopsis seeds. Appl Microbiol Biotechnol. 2015;99:6753–64.
124. Rech E, Vianna G, Murad A, Cunha N, Lacorte C, Araujo A, et al. Recombinant proteins in plants. BMC Proc. 2014;8:O1.
125. De Greve H, Virdi V, Bakshi S, Depicker A. Simplified monomeric VHH-Fc antibodies provide new opportunities for passive immunization. Curr Opin Biotechnol. 2020;61:96–101.
126. Hanania U, Ariel T, Tekoah Y, Fux L, Sheva M, Gubbay Y, et al. Establishment of a tobacco BY2 cell line devoid of plant-specific xylose and fucose as a platform for the production of biotherapeutic proteins. Plant Biotechnol J. 2017;15:1120–9.
127. Park SR, Lee J-H, Kim K, Kim TM, Lee SH, Choo Y-K, et al. Expression and in Vitro function of Anti-breast Cancer Llama-based single domain antibody VHH expressed in Tobacco plants. Int J Mol Sci. 2020;21:1354.
128. Edgue G, Twyman RM, Beiss V, Fischer R, Sack M. Antibodies from plants for bionanomaterials. WIREs Nanomed Nanobiotechnol. 2017;9:e1462.

129. Harrison RL, Jarvis DL. Protein N-glycosylation in the baculovirus-insect cell expression system and engineering of insect cells to produce mammalianized recombinant glycoproteins. Adv Virus Res. 2006;68:159–91.
130. Geisler C, Jarvis DL. Innovative use of a bacterial enzyme involved in sialic acid degradation to initiate sialic acid biosynthesis in glycoengineered insect cells. Metab Eng. 2012;14:642–52.
131. van Oers MM, Pijlman GP, Vlak JM. Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology. J Gen Virol. 2015;96:6–23.
132. Jehle JA, Blissard GW, Bonning BC, Cory JS, Herniou EA, Rohrmann GF, et al. On the classification and nomenclature of baculoviruses: a proposal for revision. Arch Virol. 2006;151:1257–66.
133. Kitts PA, Possee RD. A method for producing recombinant baculovirus expression vectors at high frequency. Biotechniques. 1993;14:810–7.
134. Pijlman GP, van den Born E, Martens DE, Vlak JM. Autographa californica baculoviruses with large genomic deletions are rapidly generated in infected insect cells. Virology. 2001;283:132–8.
135. Pijlman GP, van Schijndel JE, Vlak JM. Spontaneous excision of BAC vector sequences from bacmid-derived baculovirus expression vectors upon passage in insect cells. J Gen Virol. 2003;84:2669–78.
136. Possee RD, Hitchman RB, Richards KS, Mann SG, Siaterli E, Nixon CP, et al. Generation of baculovirus vectors for the high-throughput production of proteins in insect cells. Biotechnol Bioeng. 2008;101:1115–22.
137. Mundhada H, Seoane JM, Schneider K, Koza A, Christensen HB, Klein T, et al. Increased production of L-serine in Escherichia coli through Adaptive Laboratory Evolution. Metab Eng. 2017;39:141–50.
138. Contreras-Gómez A, Sánchez-Mirón A, García-Camacho F, Molina-Grima E, Chisti Y. Protein production using the baculovirus-insect cell expression system. Biotechnol Prog. 2014;30:1–18.
139. Aucoin MG, Mena JA, Kamen AA. Bioprocessing of baculovirus vectors: a review. Curr Gene Ther. 2010;10:174–86.
140. Cox MMJ. Innovations in the insect cell expression system for industrial recombinant vaccine Antigen Production. Vaccines. 2021;9:1504.
141. Zhao Z, Deng J, Fan D. Green biomanufacturing in recombinant collagen biosynthesis: trends and selection in various expression systems. Biomater Sci. 2023;11:5439–61.
142. Contreras-Llano LE, Tan C. High-throughput screening of biomolecules using cell-free gene expression systems. Synth Biol. 2018;3:ysy012.
143. Kunamneni A, Ogaugwu C, Bradfute S, Durvasula R. Ribosome Display Technology: applications in Disease diagnosis and control. Antibodies Basel Switz. 2020;9:28.
144. Li R, Kang G, Hu M, Huang H. Ribosome Display: a potent Display Technology used for selecting and evolving specific binders with Desired Properties. Mol Biotechnol. 2019;61:60–71.
145. Lipovsek D, Plückthun A. In-vitro protein evolution by ribosome display and mRNA display. J Immunol Methods. 2004;290:51–67.
146. Gan R, Jewett MC. Evolution of translation initiation sequences using in vitro yeast ribosome display. Biotechnol Bioeng. 2016;113:1777–86.
147. Douthwaite JA, Jackson RH, editors. Ribosome Display and Related technologies: methods and protocols. New York, NY: Springer New York; 2012.
148. Dreier B, Plückthun A. Rapid Selection of High-Affinity antibody scFv fragments using Ribosome Display. In: Nevoltris D, Chames P, editors. Antibody Engineering. New York, NY: Springer New York; 2018. pp. 235–68.
149. Cong C, Yu X, He Y, Dai Y, Zhang Y, Wang M, et al. Cell-free ribosome display and selection of antibodies on arrayed antigens. Proteomics. 2016;16:1291–6.
150. Lee K, Willi JA, Cho N, Kim I, Jewett MC, Lee J. Cell-free biosynthesis of Peptidomimetics. Biotechnol Bioprocess Eng BBE. 2023;28:905–21.
151. Chance R, Kang AS. Eukaryotic ribosome display for antibody discovery: a review. Hum Antibodies. 2024;32:107–20.
152. Zahnd C, Amstutz P, Plückthun A. Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target. Nat Methods. 2007;4:269–79.
153. Harel Inbar N, Benhar I. Selection of antibodies from synthetic antibody libraries. Arch Biochem Biophys. 2012;526:87–98.
154. Sheedy C, MacKenzie CR, Hall JC. Isolation and affinity maturation of hapten-specific antibodies. Biotechnol Adv. 2007;25:333–52.
155. Ehsasatvatan M, Kohnehrouz BB, Gholizadeh A, Ofoghi H, Shanehbandi D. The production of the first functional antibody mimetic in higher plants: the chloroplast makes the DARPin G3 for HER2 imaging in oncology. Biol Res. 2022;55:32.
156. Hanes J, Schaffitzel C, Knappik A, Plückthun A. Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. Nat Biotechnol. 2000;18:1287–92.
157. Currin A, Swainston N, Day PJ, Kell DB. Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. Chem Soc Rev. 2015;44:1172–239.
158. Takahashi TT, Austin RJ, Roberts RW. mRNA display: ligand discovery, interaction analysis and beyond. Trends Biochem Sci. 2003;28:159–65.
159. Golynskiy MV, Haugner JC, Morelli A, Morrone D, Seelig B. In vitro evolution of enzymes. Methods Mol Biol Clifton NJ. 2013;978:73–92.
160. Gold L. mRNA display: diversity matters during in vitro selection. Proc Natl Acad Sci U S A. 2001;98:4825–6.
161. Nemoto N, Miyamoto-Sato E, Husimi Y, Yanagawa H. In vitro virus: bonding of mRNA bearing puromycin at the 3′-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. FEBS Lett. 1997;414:405–8.
162. Hamzeh-Mivehroud M, Alizadeh AA, Morris MB, Church WB, Dastmalchi S. Phage display as a technology delivering on the promise of peptide drug discovery. Drug Discov Today. 2013;18:1144–57.
163. Lee J-H, Song HS, Park TH, Lee S-G, Kim B-G. Screening of cell-penetrating peptides using mRNA display. Biotechnol J. 2012;7:387–96.
164. Cotham VC, Horton AP, Lee J, Georgiou G, Brodbelt JS. Middle-Down 193-nm Ultraviolet Photodissociation for unambiguous antibody identification and its implications for immunoproteomic analysis. Anal Chem. 2017;89:6498–504.
165. Egloff P, Zimmermann I, Arnold FM, Hutter CAJ, Morger D, Opitz L, et al. Engineered peptide barcodes for in-depth analyses of binding protein libraries. Nat Methods. 2019;16:421–8.
166. Hutter CAJ, Timachi MH, Hürlimann LM, Zimmermann I, Egloff P, Göddeke H, et al. The extracellular gate shapes the energy profile of an ABC exporter. Nat Commun. 2019;10:2260.
167. Fridy PC, Li Y, Keegan S, Thompson MK, Nudelman I, Scheid JF, et al. A robust pipeline for rapid production of versatile nanobody repertoires. Nat Methods. 2014;11:1253–60.
168. Scheid JF, Mouquet H, Ueberheide B, Diskin R, Klein F, Oliveira TYK, et al. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. Science. 2011;333:1633–7.
169. Jovčevska I, Muyldermans S. The therapeutic potential of Nanobodies. BioDrugs Clin Immunother Biopharm Gene Ther. 2020;34:11–26.
170. Laustsen AH, Greiff V, Karatt-Vellatt A, Muyldermans S, Jenkins TP. Animal immunization, in Vitro Display Technologies, and machine learning for antibody Discovery. Trends Biotechnol. 2021;39:1263–73.
171. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–44.
172. Zhang Z, Cui P, Zhu W. Deep learning on Graphs: a Survey. IEEE Trans Knowl Data Eng. 2022;34:249–70.
173. Poplin R, Chang P-C, Alexander D, Schwartz S, Colthurst T, Ku A, et al. A universal SNP and small-indel variant caller using deep neural networks. Nat Biotechnol. 2018;36:983–7.
174. Zou J, Huss M, Abid A, Mohammadi P, Torkamani A, Telenti A. A primer on deep learning in genomics. Nat Genet. 2019;51:12–8.
175. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542:115–8.
176. Wainberg M, Merico D, Delong A, Frey BJ. Deep learning in biomedicine. Nat Biotechnol. 2018;36:829–38.
177. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, et al. Highly accurate protein structure prediction for the human proteome. Nature. 2021;596:590–6.
178. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–9.
179. Bagdonas H, Fogarty CA, Fadda E, Agirre J. The case for post-predictional modifications in the AlphaFold protein structure database. Nat Struct Mol Biol. 2021;28:869–70.
180. Schneider P, Walters WP, Plowright AT, Sieroka N, Listgarten J, Goodnow RA, et al. Rethinking drug design in the artificial intelligence era. Nat Rev Drug Discov. 2020;19:353–64.
181. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A Deep Learning Approach to Antibiotic Discovery. Cell. 2020;180:688–e70213.
182. Muyldermans S. Nanobodies: natural single-domain antibodies. Annu Rev Biochem. 2013;82:775–97.
183. Wu X, Zhang J, Fang L. Next-generation sequencing and Proteomics-Enabled Approach for Rapid and High-Throughput isolation of Virus-Neutralizing Nanobodies. Methods Mol Biol Clifton NJ. 2024;2824:135–46.

184. Cohen T, Halfon M, Schneidman-Duhovny D, NanoNet. Rapid and accurate end-to-end nanobody modeling by deep learning. Front Immunol. 2022;13:958584.
185. Alexander E, Leong KW. Discovery of nanobodies: a comprehensive review of their applications and potential over the past five years. J Nanobiotechnol. 2024;22:661.
186. Li S, Meng X, Li R, Huang B, Wang X. NanoBERTa-ASP: predicting nanobody paratope based on a pretrained RoBERTa model. BMC Bioinformatics. 2024;25:122.
187. Hacisuleyman A, Erman B, ModiBodies:. A computational method for modifying nanobodies in nanobody-antigen complexes to improve binding affinity and specificity. J Biol Phys. 2020;46:189–208.
188. Sardar U, Ali S, Ayub MS, Shoaib M, Bashir K, Khan IU et al. Sequence-based nanobody-antigen binding prediction. 2023.
189. De Keyser P, Kalichuk V, Zögg T, et al. A biosensor-based phage display selection method for automated, high-throughput Nanobody discovery. Biosens Bioelectron. 2025;271:116951.
190. Gallo E. The rise of big data: deep sequencing-driven computational methods are transforming the landscape of synthetic antibody design. J Biomed Sci. 2024;31:29.
191. Zuo J, Li J, Zhang R, Xu L, Chen H, Jia X, et al. Institute collection and analysis of Nanobodies (iCAN): a comprehensive database and analysis platform for nanobodies. BMC Genomics. 2017;18:797.
192. Deszyński P, Młokosiewicz J, Volanakis A, Jaszczyszyn I, Castellana N, Bonissone S, et al. INDI—integrated nanobody database for immunoinformatics. Nucleic Acids Res. 2022;50:D1273–81.
193. Xiong S, Liu Z, Yi X, Liu K, Huang B, Wang X. NanoLAS: a comprehensive nanobody database with data integration, consolidation and application. Database. 2024;2024:baae003.
194. Valdés-Tresanco MS, Valdés-Tresanco ME, Molina-Abad E, Moreno E. NbThermo: a new thermostability database for nanobodies. Database J Biol Databases Curation. 2023;2023:baad021.
195. Schneider C, Raybould MIJ, Deane CM. SAbDab in the age of biotherapeutics: updates including SAbDab-nano, the nanobody structure tracker. Nucleic Acids Res. 2022;50:D1368–72.
196. Wilton EE, Opyr MP, Kailasam S, Kothe RF, Wieden H-J. sdAb-DB: the single domain antibody database. ACS Synth Biol. 2018;7:2480–4.
197. Gło N. Statistical mechanics of protein folding, unfolding and fluctuation. Adv Biophys. 1976;7:65–113.
198. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 1993;234:779–815.
199. Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences1. J Mol Biol. 1999;287:797–815.
200. Ghouzam Y, Postic G, de Brevern AG, Gelly J-C. Improving protein fold recognition with hybrid profiles combining sequence and structure evolution. Bioinforma Oxf Engl. 2015;31:3782–9.
201. Rohl CA, Strauss CEM, Misura KMS, Baker D. Protein structure prediction using Rosetta. Methods Enzymol. 2004;383:66–93.
202. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: protein structure and function prediction. Nat Methods. 2015;12:7–8.
203. Bystroff C, Shao Y. Fully automated ab initio protein structure prediction using I-SITES, HMMSTR and ROSETTA. Bioinforma Oxf Engl. 2002;18(Suppl 1):S54–61.
204. Xu J. Distance-based protein folding powered by deep learning. Proc Natl Acad Sci U S A. 2019;116:16856–65.
205. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A et al. Transformers: state-of-the-art natural language processing. In: Liu Q, Schlangen D, editors. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Online: Association for Computational Linguistics; 2020. pp. 38–45.
206. Mitchell LS, Colwell LJ. Analysis of nanobody paratopes reveals greater diversity than classical antibodies. Protein Eng Des Sel PEDS. 2018;31:267–75.
207. Zavrtanik U, Lukan J, Loris R, Lah J, Hadži S. Structural basis of Epitope Recognition by Heavy-Chain Camelid Antibodies. J Mol Biol. 2018;430:4369–86.
208. Yamamoto K, Nagatoishi S, Nakakido M, Kuroda D, Tsumoto K. Conformational features and interaction mechanisms of VHH antibodies with β-hairpin-like CDR-H3: a case of Nb8-HigB2 interaction. 2023;32(12):e4827.
209. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science. 2021;373:871–6.
210. Wu R, Ding F, Wang R, Shen R, Zhang X, Luo S et al. High-resolution de novo structure prediction from primary sequence. bioRxiv. 2022:500999.
211. Lin Z, Akin H, Rao R, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science. 2023;379(6637):1123–30.
212. Ahdritz G, Bouatta N, Floristean C, et al. OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. Nat Methods. 2024;21(8):1514–24.
213. Ruffolo JA, Chu L-S, Mahajan SP, Gray JJ. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. Nat Commun. 2023;14:2389.
214. Kenlay H, Dreyer FA, Cutting D, Nissley D, Deane CM. ABodyBuilder3: improved and scalable antibody structure predictions. 2024.
215. Abanades B, Wong WK, Boyles F, Georges G, Bujotzek A, Deane CM. ImmuneBuilder: deep-learning models for predicting the structures of immune proteins. Commun Biol. 2023;6:575.
216. Sircar A, Sanni KA, Shi J, Gray JJ. Analysis and modeling of the Variable Region of Camelid single-domain antibodies. J Immunol. 2011;186:6357–67.
217. Li B, Qin X, Mi L-Z. Nanobodies: from structure to applications in non-injectable and bispecific biotherapeutic development. Nanoscale. 2022;14:7110–22.
218. Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC, Flajnik MF. A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature. 1995;374:168–73.
219. Eshak F, Pion L, Scholler P, Nevoltris D, Chames P, Rondard P, et al. Epitope identification of an mGlu5 receptor Nanobody using physics-based molecular modeling and deep learning techniques. J Chem Inf Model. 2024;64:4436–61.
220. Valdés-Tresanco MS, Valdés-Tresanco ME, Jiménez-Gutiérrez DE, Moreno E. Structural modeling of nanobodies: a Benchmark of State-of-the-art Artificial Intelligence Programs. Mol Basel Switz. 2023;28:3991.
221. Chen H, Fan X, Zhu S, Pei Y, Zhang X, Zhang X, et al. Accurate prediction of CDR-H3 loop structures of antibodies with deep learning. eLife. 2024;12:RP91512.
222. Chiu ML, Goulet DR, Teplyakov A, Gilliland GL. Antibody structure and function: the basis for Engineering therapeutics. Antibodies Basel Switz. 2019;8:55.
223. Xiang Y, Sang Z, Bitton L, Xu J, Liu Y, Schneidman-Duhovny D, et al. Integrative proteomics identifies thousands of distinct, multi-epitope, and high-affinity nanobodies. Cell Syst. 2021;12:220–e2349.
224. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein–protein docking. Nat Protoc. 2017;12:255–78.
225. van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis PL, Karaca E, et al. The HADDOCK2.2 web server: user-friendly integrative modeling of Biomolecular complexes. J Mol Biol. 2016;428:720–5.
226. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv. 2021:463034.
227. Yin R, Pierce BG. Evaluation of AlphaFold antibody-antigen modeling with implications for improving predictive accuracy. Protein Sci Publ Protein Soc. 2024;33:e4865.
228. Muyldermans S. Single domain camel antibodies: current status. Rev Mol Biotechnol. 2001;74:277–302.
229. Mitchell LS, Colwell LJ. Comparative analysis of nanobody sequence and structure data. Proteins. 2018;86:697–706.
230. Kovaltsuk A, Leem J, Kelm S, Snowden J, Deane CM, Krawczyk K. Observed Antibody Space: a resource for data mining next-generation sequencing of antibody repertoires. J Immunol Baltim Md. 1950. 2018;201:2502–9.
231. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D et al. RoBERTa: a robustly optimized BERT pretraining approach. 2019.
232. Moonens K, Gideonsson P, Subedi S, Bugaytsova J, Romaõ E, Mendez M, et al. Structural insights into polymorphic ABO glycan binding by Helicobacter pylori. Cell Host Microbe. 2016;19:55–66.
233. Sela-Culang I, Kunik V, Ofran Y. The structural basis of antibody-antigen recognition. Front Immunol. 2013;4:302.
234. Tam C, Kumar A, Zhang KYJ. NbX: machine learning-guided re-ranking of Nanobody-Antigen binding poses. Pharm Basel Switz. 2021;14:968.
235. Wang X, Terashi G, Christoffer CW, Zhu M, Kihara D. Protein docking model evaluation by 3D deep convolutional neural networks. Bioinforma Oxf Engl. 2020;36:2113–8.
236. Cunningham O, Scott M, Zhou ZS, Finlay WJJ. Polyreactivity and polyspecificity in therapeutic antibody development: risk factors for failure in preclinical and clinical development campaigns. mAbs. 2021;13:1999195.
237. Sigounas G, Harindranath N, Donadel G, Notkins AL. Half-life of polyreactive antibodies. J Clin Immunol. 1994;14:134–40.

238. Kelly RL, Sun T, Jain T, Caffry I, Yu Y, Cao Y, et al. High throughput cross-interaction measures for human IgG1 antibodies correlate with clearance rates in mice. mAbs. 2015;7:770–7.
239. Berglund L, Björling E, Oksvold P, Fagerberg L, Asplund A, Szigyarto CA-K, et al. A genecentric human protein atlas for expression profiles based on antibodies. Mol Cell Proteom MCP. 2008;7:2019–27.
240. Harvey EP, Shin J-E, Skiba MA, Nemeth GR, Hurley JD, Wellner A, et al. An in silico method to assess antibody fragment polyreactivity. Nat Commun. 2022;13:7554.
241. Wesolowski J, Alzogaray V, Reyelt J, Unger M, Juarez K, Urrutia M, et al. Single domain antibodies: promising experimental and therapeutic tools in infection and immunity. Med Microbiol Immunol (Berl). 2009;198:157–74.
242. Morrison SL, Johnson MJ, Herzenberg LA, Oi VT. Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. Proc Natl Acad Sci U S A. 1984;81:6851–5.
243. Baca M, Presta LG, O'Connor SJ, Wells JA. Antibody humanization using monovalent phage display. J Biol Chem. 1997;272:10678–84.
244. Jones PT, Dear PH, Foote J, Neuberger MS, Winter G. Replacing the complementarity-determining regions in a human antibody with those from a mouse. Nature. 1986;321:522–5.
245. Kashmiri SVS, De Pascalis R, Gonzales NR, Schlom J. SDR grafting–a new approach to antibody humanization. Methods San Diego Calif. 2005;36:25–34.
246. De Pascalis R, Iwahashi M, Tamura M, Padlan EA, Gonzales NR, Santos AD, et al. Grafting of abbreviated complementarity-determining regions containing specificity-determining residues essential for ligand contact to engineer a less immunogenic humanized monoclonal antibody. J Immunol Baltim Md 1950. 2002;169:3076–84.
247. Dall'Acqua WF, Damschroder MM, Zhang J, Woods RM, Widjaja L, Yu J, et al. Antibody humanization by framework shuffling. Methods San Diego Calif. 2005;36:43–60.
248. Roguska MA, Pedersen JT, Keddy CA, Henry AH, Searle SJ, Lambert JM, et al. Humanization of murine monoclonal antibodies through variable domain resurfacing. Proc Natl Acad Sci U S A. 1994;91:969–73.
249. Padlan EA. A possible procedure for reducing the immunogenicity of antibody variable domains while preserving their ligand-binding properties. Mol Immunol. 1991;28:489–98.
250. Shi Z, Li X, Wang L, Sun Z, Zhang H, Chen X, et al. Structural basis of nanobodies neutralizing SARS-CoV-2 variants. Struct Lond Engl 1993. 2022;30:707–e7205.
251. Zhao L, Meng F, Li Y, Liu S, Xu M, Chu F, et al. Multivalent Nanobody conjugate with rigid, reactive oxygen species scavenging Scaffold for Multi-target Therapy of Alzheimer's Disease. Adv Mater Deerfield Beach Fla. 2023;35:e2210879.
252. Sun Z, Wang L, Li L, Sun Y, Zhang D, Zhou S, et al. Structure basis of two nanobodies neutralizing SARS-CoV-2 Omicron variant by targeting ultra-conservative epitopes. J Struct Biol. 2023;215:107996.
253. Wang Y, Chen J, Zhang S, Jiang H, Zhu J, Jiang G, et al. Bispecific Nanobody-Aptamer conjugates for enhanced Cancer therapy in solid tumors. Small Weinh Bergstr Ger. 2024;20:e2308265.
254. Liao X, Zhang Y, Liang Y, Zhang L, Wang P, Wei J, et al. Enhanced sandwich immunoassay based on bivalent nanobody as an efficient immobilization approach for foodborne pathogens detection. Anal Chim Acta. 2024;1289:342209.
255. Tam C, Kukimoto-Niino M, Miyata-Yabuki Y, Tsuda K, Mishima-Tsumagari C, Ihara K, et al. Targeting ras-binding domain of ELMO1 by computational nanobody design. Commun Biol. 2023;6:284.
256. Heidari MM, Shirazi EA, Cheraghi SF, Shahshahani R, Rahnama T, Khatami M. CDR grafting and site-directed mutagenesis approach for the generation and affinity maturation of Anti-CD20 nanobody. Mol Biol Rep. 2024;51:751.
257. Wang J, Kang G, Lu H, de Marco A, Yuan H, Feng Z, et al. Novel bispecific nanobody mitigates experimental intestinal inflammation in mice by targeting TNF-α and IL-23p19 bioactivities. Clin Transl Med. 2024;14:e1636.
258. Liu H, Wang R, An D, Liu H, Ye F, Li B et al. An engineered IL-21 with half-life extension enhances anti-tumor immunity as a monotherapy or in combination with PD-1 or TIGIT blockade. Int Immunopharmacol. 2021;101 Pt A:108307.
259. Dinesen A, Andersen VL, Elkhashab M, Pilati D, Bech P, Fuchs E, et al. An Albumin-Holliday Junction Biomolecular Modular Design for Programmable Multifunctionality and prolonged circulation. Bioconjug Chem. 2024;35:214–22.
260. Vascon F, Gasparotto M, Giacomello M, Cendron L, Bergantino E, Filippini F, et al. Protein electrostatics: from computational and structural analysis to discovery of functional fingerprints and biotechnological design. Comput Struct Biotechnol J. 2020;18:1774–89.
261. Mahajan SP, Meksiriporn B, Waraho-Zhmayev D, Weyant KB, Kocer I, Butler DC, et al. Computational affinity maturation of camelid single-domain intrabodies against the nonamyloid component of alpha-synuclein. Sci Rep. 2018;8:17611.
262. Cheng X, Wang J, Kang G, Hu M, Yuan B, Zhang Y, et al. Homology modeling-based in Silico Affinity Maturation improves the Affinity of a Nanobody. Int J Mol Sci. 2019;20:4187.
263. Hadsund JT, Satława T, Janusz B, Shan L, Zhou L, Röttger R, et al. nanoBERT: a deep learning model for gene agnostic navigation of the nanobody mutational space. Bioinforma Adv. 2024;4:vbae033.
264. Bai Z, Wang J, Li J, Yuan H, Wang P, Zhang M, et al. Design of nanobody-based bispecific constructs by in silico affinity maturation and umbrella sampling simulations. Comput Struct Biotechnol J. 2023;21:601–13.
265. Alvarez JAE, Dean SN. TEMPRO: nanobody melting temperature estimation model using protein embeddings. Sci Rep. 2024;14:19074.
266. Norman RA, Ambrosetti F, Bonvin AMJJ, Colwell LJ, Kelm S, Kumar S, et al. Computational approaches to therapeutic antibody design: established methods and emerging trends. Brief Bioinform. 2020;21:1549–67.
267. Leem J, Dunbar J, Georges G, Shi J, Deane CM. ABodyBuilder: automated antibody structure prediction with data-driven accuracy estimation. mAbs. 2016;8:1259–68.
268. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding principles for scientific data management and stewardship. Sci Data. 2016;3:160018.
269. Fernández LA. Prokaryotic expression of antibodies and affibodies. Curr Opin Biotechnol. 2004;15:364–73.

Publisher's note