PhytoNB: AI-enabled Nanobody Design in Plants

Tianhao WU[1], Zixuan WANG[2], Anwen ZHAO[1], Fan XIA[1], Yuxuan LOU[1], Can YIN[1], Yanfen XU[2], Jianan ZHANG[2], Xiangfeng WANG[1], Qian CHENG[1]

([1]*State Key Laboratory of Maize Bio-breeding, National Maize Improvement Center, Frontiers Science Center for Molecular Design Breeding, College of Agronomy and Biotechnology, China Agricultural University, Beijing 100094, China;* [2]*Molbreeding Biotechnology Co., Ltd., Shijiazhuang 051430, China*)

Abstract: The rapid expansion of the AlphaFold Database has provided unprecedented structural coverage of plant proteomes, thereby creating new opportunities for the computational design of functional proteins tailored for agricultural and biotechnological applications. However, existing deep learning-based antibody design methods encounter substantial computational bottlenecks when scaled for high-throughput screening of plant targets. We present PhytoNB, an automated, parallel-accelerated framework for the *de novo* design of plant nanobodies. This pipeline integrates domain-level segmentation via Chainsaw, multi-view binding site prediction using GPSite and MVGNN, generative nanobody design with IgGM, and large-scale energy-based filtering through Rosetta. To address the considerable computational demands of large-scale generative models, we developed a parallel acceleration engine incorporating dynamic GPU scheduling and multi-threading optimization. This engine enables efficient task allocation across multiple GPUs and CPU cores. Starting from structural inputs, PhytoNB autonomously performs structure prediction, epitope localization, sequence-structure co-generation, and biophysical validation. The pipeline thereby identifies high-stability nanobody candidates that target key functional regions of plant proteins. Benchmarks demonstrate that the parallelized workflow achieves orders-of-magnitude acceleration in design throughput while maintaining structural fidelity and binding specificity. PhytoNB provides an efficient and scalable platform for plant nanobody discovery, with an interface that supports streamlined application of the design workflow. Extensible applications of this platform include multi-epitope targeting, multi-specific binder design, and cross-species applications.

Key words: plant nanobody; parallel acceleration; IgGM; dynamic GPU scheduling

Wu T, Wang Z, Zhao A, Xia F, Lou Y, Yin C, Xu Y, Zhang J, Wang X, Cheng Q. 2026. PhytoNB: AI-enabled nanobody design in plants. *Maize Sciences*, **1**: 100004.

INTRODUCTION

Nanobodies are single-domain antibody fragments derived from heavy-chain antibodies found in camelids, such as alpacas and camels (Arbabi-Ghahroudi, 2022). These heavy-chain antibodies are naturally devoid of light chains and the CH1 domain. Their isolated variable regions, known as VHH domains, constitute nanobodies. These molecules are also referred to as single-domain antibodies (sdAbs). Nanobodies possess several advantageous properties, including a small molecular weight of approximately 12–15 kDa, high stability, and strong binding affinity (Alexander and Leong, 2024). These characteristics have enabled their application in biomedicine. More recently, nanobodies have also been explored in plant biotechnology (Wang et al., 2021). The classical route for obtaining nanobodies involves the immunization of camelids. In this process, target antigens are first injected into animals such as alpacas or camels (Armstrong et al., 2023). This step induces the production of antigen-specific heavy-chain antibodies. Subsequently, researchers collect peripheral blood lymphocytes from the immunized animals. These lymphocytes undergo high-throughput sequencing and phage display screening. This workflow ultimately yields high-affinity nanobody sequences. However, this animal immunization process is both time-consuming and costly (Zhu and Ding, 2025). In recent years, alternative methods based on fully

Received: 23 March 2026; **Accepted:** 3 April 2026
Corresponding author: Qian Cheng (qchengray@cau.edu.cn)
©The Author(s) 2026. Published by Higher Education Press
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0)
https://doi.org/10.2738/MS.2026.0004

synthetic libraries have emerged as substitutes for immunization (Liu et al., 2025).

Recent advances in artificial intelligence have transformed protein research. Deep learning-based protein structure prediction and design have become prevailing trends in this field (Bennett et al., 2023; Jumper et al., 2021). Structure prediction tools such as AlphaFold2 have dramatically enhanced prediction accuracy. These tools have enabled the generation of large-scale structural databases that encompass plant proteins (Fleming et al., 2025). In recent years, various AI-driven protein design methodologies have been proposed. Language model-based approaches such as ProteinMPNN represent one major direction (Dauparas et al., 2022). Diffusion model-based frameworks such as RFDiffusion constitute another important category (Ahern et al., 2026; Kırboğa and Küçüksille, 2026; Watson et al., 2023). Since 2020, AI tools have been increasingly applied to nanobody design. These applications have significantly improved the structural prediction of nanobodies. They have also facilitated the rational design of multi-epitope nanobodies (Liu et al., 2025). The recently introduced IgGM model leverages generative deep learning for antibody design. This model can simultaneously design antibody sequences and structures. It offers a novel paradigm for generating immunoglobulins with desired functional properties (Wang et al., 2025). Given a specific antigen, IgGM can co-generate the sequences and three-dimensional structures of nanobodies targeting that antigen. This approach demonstrates the potential to create novel antibody and nanobody variants. However, large-scale models such as IgGM entail substantial computational demands. These models consume significant GPU resources and processing time. When dealing with antigens of considerable length, truncation of the input is often necessary to alleviate memory pressure. This limitation presents a notable computational bottleneck for high-throughput applications (Lin et al., 2025).

This study proposes PhytoNB, an automated and parallel-accelerated pipeline specifically engineered to meet the high-throughput demands of plant biotechnology. While nanobodies have demonstrated immense potential for recognizing plant pathogens, detecting toxins, and enhancing crop resistance (Wang et al., 2021), their discovery has traditionally been a low-throughput process. The recent expansion of the AlphaFold Database (Fleming et al., 2025) has provided unprecedented structural coverage of plant proteomes; however, it also creates a significant computational bottleneck. Existing AI-driven design tools are often not optimized for large-scale execution, making the screening of thousands of plant targets prohibitively slow. PhytoNB addresses this gap by integrating advanced epitope identification and AI-driven generation (IgGM) within a GPU-parallelized framework. By implementing a dynamic multi-threading optimization strategy, our platform transforms nanobody design from a month-long endeavor into a high-throughput computational workflow. Beyond raw speed, PhytoNB emphasizes usability through a dedicated web interface, enabling plant scientists to bridge the gap between structural 'big data' and functional molecular tools. This integrated framework provides a robust foundation for the rapid development of nanobodies across diverse plant protein targets, as demonstrated by our experimental validation on key maize proteins.

RESULTS AND ANALYSIS

Overview of the PhytoNB design pipeline

The PhytoNB pipeline follows a sequential workflow for nanobody design. First, predicted structures of target plant proteins are retrieved from the AlphaFold Database. These structures then undergo domain-level segmentation using Chainsaw. Following segmentation, binding-site prediction is performed on each domain. This step identifies potential antigenic epitopes within the target protein. The identified epitope residues are subsequently used as input for IgGM. IgGM generates candidate nanobody sequences along with their corresponding three-dimensional structures. Finally, the resulting nanobody-antigen complexes are subjected to energy-based screening via Rosetta. Designs with the lowest energy values are retained, as lower energy indicates higher binding stability. The entire pipeline is accelerated through GPU parallelization and multi-threading optimization. This acceleration enables seamless, automated end-to-end execution from target protein structure to candidate nanobody design.

Data preparation and domain segmentation

The initial phase of the PhytoNB pipeline focuses on the standardized processing of plant target protein information. Input data may consist of either the amino acid sequence of a plant protein or its known

three-dimensional structure, such as Protein Data Bank (PDB) format. If only sequence information is provided by the user, the system first employs deep learning-based tools to predict high-confidence three-dimensional structures. For example, ESMFold can be used for structure prediction (Fig. 1-A).

Plants often possess long sequences or multi-domain proteins. To address this characteristic, the Chainsaw tool is introduced to perform domain-level segmentation of the overall structure. This step isolates individual domains from the full-length protein. The significance of this segmentation is twofold. First, decomposing large proteins into independent units enables parallel processing of individual domains. Second, this approach allows the design process to focus more specifically on larger surface interfaces. These regions are generally more amenable to nanobody binding. Consequently, the segmentation step substantially enhances design specificity and improves the computational success rate (Fig. 1-A).

Epitope identification and spatial constraint construction

Following the acquisition of preprocessed isolated domains, the system employs a multi-path computation strategy to achieve precise localization of potential binding epitopes (Fig. 1-B). For each

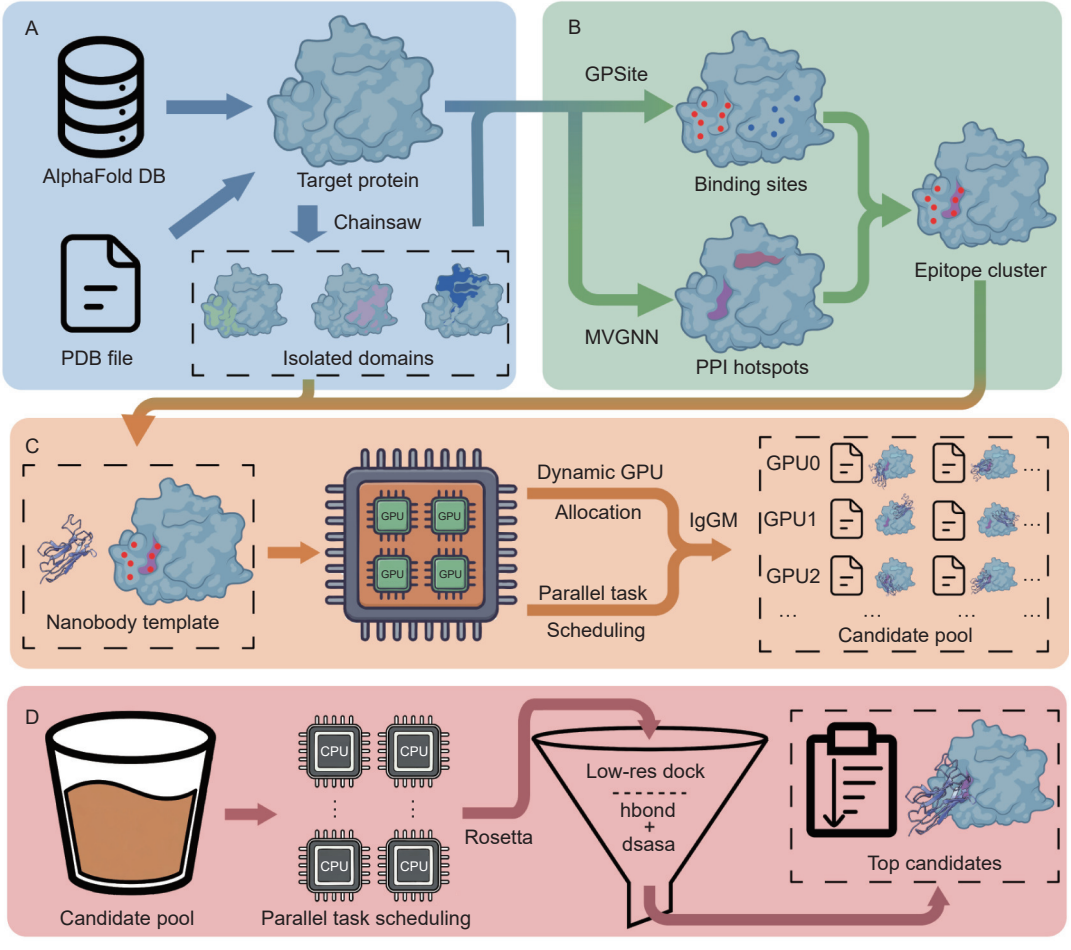

Fig. 1. Schematic overview of the PhytoNB automated parallel-accelerated pipeline for nanobody design.
A. Target processing module. Input protein structures are retrieved from the AlphaFold Database or uploaded as user-provided PDB files. For long or multi-domain proteins, Chainsaw performs domain-level segmentation to generate isolated domains for parallel processing. **B**. Epitope mapping module. For each isolated domain, GPSite predicts general ligand-binding sites, and MVGNN identifies protein-protein interaction hotspots. Predicted sites are filtered and clustered into epitope clusters, serving as spatial constraints for subsequent nanobody generation. **C**. AI-driven parallel generation engine. The IgGM model, supported by dynamic GPU allocation and parallel task scheduling, distributes design tasks across multiple GPU instances. Each instance outputs candidate nanobody sequences and their corresponding complex structures, which are aggregated into a candidate pool. **D**. Refinement, filtering, and output module. Candidate complexes are evaluated in parallel using Rosetta-based high-throughput screening, including low-resolution docking, hydrogen bond scoring, and solvent-accessible surface area analysis. Top-ranked candidates are retained, and key molecular interactions, such as hydrogen bonds and hydrophobic contacts.

domain, the pipeline concurrently utilizes two complementary prediction tools. GPSite is employed to predict sites on the domain surface that may interact with small molecules or protein ligands. These sites are referred to as binding sites. Simultaneously, MVGNN, a graph convolutional neural network-based method, is used to annotate potential protein-protein interaction hotspots, termed PPI hotspots.

The raw predicted sites from both tools undergo further processing. These sites are first filtered using predefined confidence thresholds. The remaining sites are then clustered into several key residue clusters, which are designated as epitope clusters. These clusters serve as spatial constraints for subsequent nanobody generation by the IgGM model. This approach ensures that the generated candidate sequences are directed toward the key functional regions of the antigen (Fig. 1-B).

IgGM generation module and dynamic GPU scheduling strategy

The IgGM module serves as the core engine for *de novo* nanobody generation within this pipeline. This module has been designed with a dual-mode operation strategy that balances flexibility and computational efficiency (Fig. 1-C). The standard mode processes individual epitopes sequentially, making it suitable for small-scale computing tasks. In contrast, the accelerated parallel mode is specifically designed to fully harness the computational power of multi-GPU hardware resources. This mode is supported by a dedicated parallel acceleration engine.

A sophisticated dynamic GPU scheduling strategy has been implemented within this engine. First, pending epitope tasks are prioritized based on their importance or computational complexity. Following prioritization, tasks are distributed according to a first-come first-served principle or grouped based on their memory requirements. Prior to subtask submission, the scheduler estimates memory footprint by considering model parameters and input dimensions. The scheduler also performs real-time checks of available GPU memory. These checks ensure that each task is assigned to the physical card with the lightest load and sufficient available memory. This approach prevents resource wastage and computation interruptions caused by out-of-memory errors. This adaptive allocation strategy enables IgGM to generate multiple antibody sequences and their corresponding complex structures in parallel across multiple GPUs. Compared to traditional serial execution modes, this parallel approach achieves an orders-of-magnitude improvement in generation throughput (Fig. 1-C).

Application cases of the PhytoNB platform

The PhytoNB platform employs an intuitive, interactive result visualization scheme. This scheme showcases *de novo* design case studies on four representative plant-associated proteins (Fig. 2-A to -D). These case studies encompass both endogenous plant physiological regulatory genes and exogenous agricultural trait genes. This multi-dimensional visualization approach aims to validate the platform's generalizability across target proteins with diverse functional categories.

For endogenous gene targets, two examples are presented. The maize gibberellin receptor ZmGID1 is shown in Fig. 2-A (Islam et al., 2025). The key carotenoid biosynthesis enzyme ZmPDS1 is shown in Fig. 2-B. For each target, the system clearly displays the domain-level segmentation performed by Chainsaw, with domains distinguished by red, green, and blue coloration. The system also displays epitope hotspots annotated by GPSite and MVGNN. ZmGID1 functions as a signaling hub that regulates plant height and seed germination (Islam et al., 2025). ZmPDS1 serves as a rate-limiting enzyme critical for photosynthetic efficiency (Peng et al., 2024). Both targets yielded conformationally stable nanobody complexes under the guidance of the platform. These complexes are depicted in cyan in the figures.

To further validate the broad applicability of the design platform, exemplary design capabilities are also demonstrated for exogenous genes which are widely employed in agricultural biotechnology. The acetyltransferase BAR confers herbicide resistance in plants and is shown in Fig. 2-C. CP4EPSPS, a core target in glyphosate-resistant crops, is shown in Fig. 2-D. These exogenous proteins are critical to modern agriculture (Mou et al., 2026). They were subjected to the identical automated parallel-accelerated pipeline. The system successfully identified their key interfacial regions and produced high-quality candidate nanobody models.

This concurrent successful design targeting both endogenous physiological targets and exogenous resistance markers carries two important implications. Theoretically, it underscores the robust compatibility

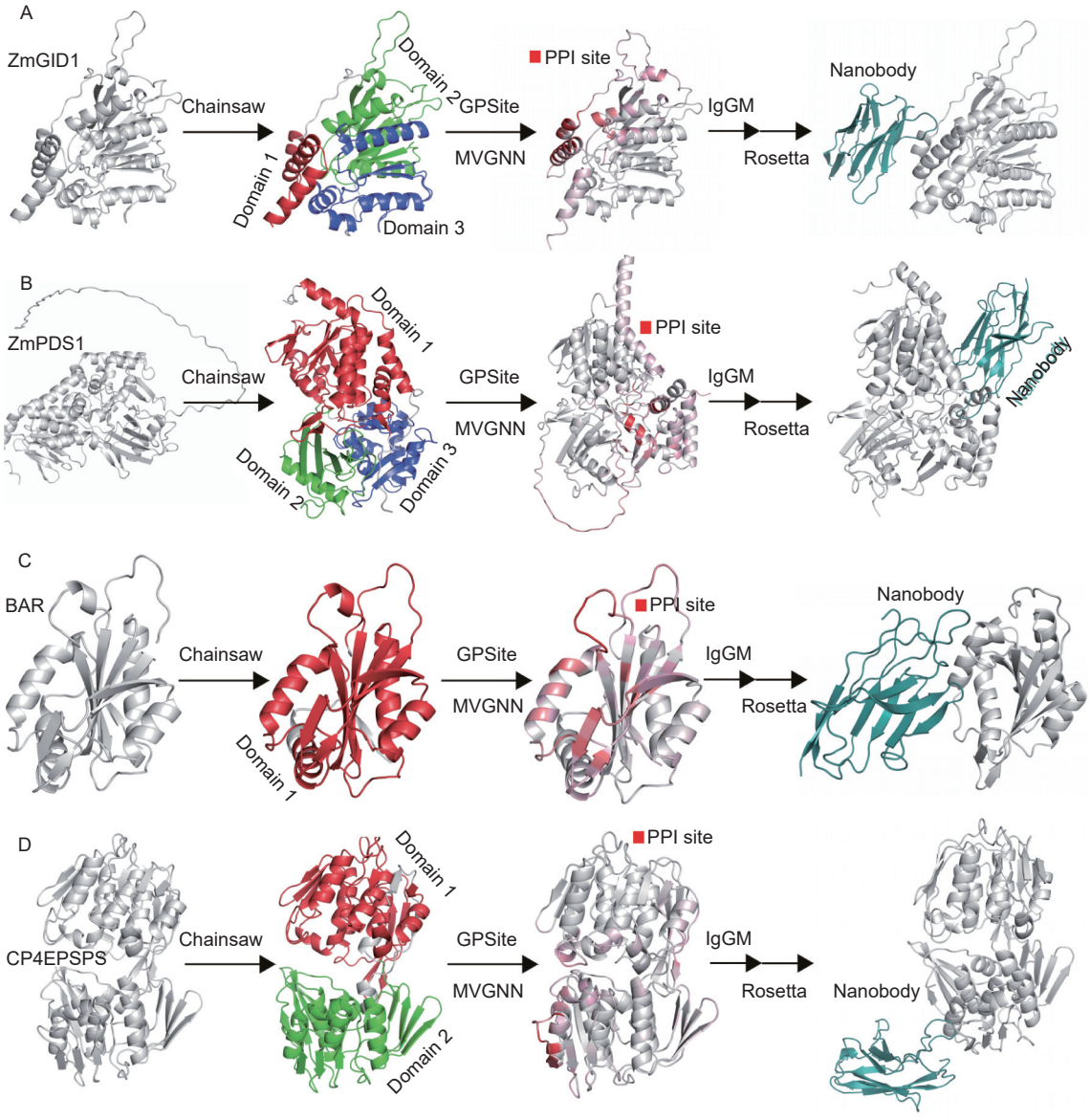

Fig. 2. Application examples of the PhytoNB platform on endogenous and exogenous protein targets in maize.
A–D. Complete design pipelines are shown for four representative targets: the endogenous regulatory proteins ZmGID1 (A) and ZmPDS1 (B), and the exogenous functional proteins BAR (C) and CP4EPSPS (D). For each target, three key stages are illustrated. Chainsaw: Domain-level segmentation results, with distinct colors representing isolated structural domains. GPSite/MVGNN: High-confidence epitope hotspots, indicated in red, are identified as core binding regions for antibody design. Nanobody (IgGM/Rosetta): Final complex structure models are shown, with the generated nanobody depicted in cyan and the target protein in gray. These four case studies demonstrate the fully automated workflow from raw sequence input to high-confidence candidate models. The results highlight the broad applicability of PhytoNB across protein targets of varying sizes, ranging from 175 to 571 amino acids, and of distinct biological origins.

of PhytoNB with targets of diverse origins and functions. Practically, it provides solid theoretical data to support subsequent experimental validation, such as *in vitro* affinity assays.

Web-based platform for PhytoNB

To improve the accessibility of the PhytoNB pipeline, we developed a web-based platform that enables users to perform nanobody design through an intuitive interface (Fig. 3). The platform supports two design modes, AFDB Design and Custom Design, allowing users either to retrieve protein structures from public databases or to upload user-defined PDB files. The web interface provides an integrated workflow for task submission, progress tracking, and result retrieval. Users can configure analysis parameters, submit jobs, and monitor execution status in real time, forming a complete and continuous analysis process. Designed with usability in mind, the platform simplifies complex computational procedures into a

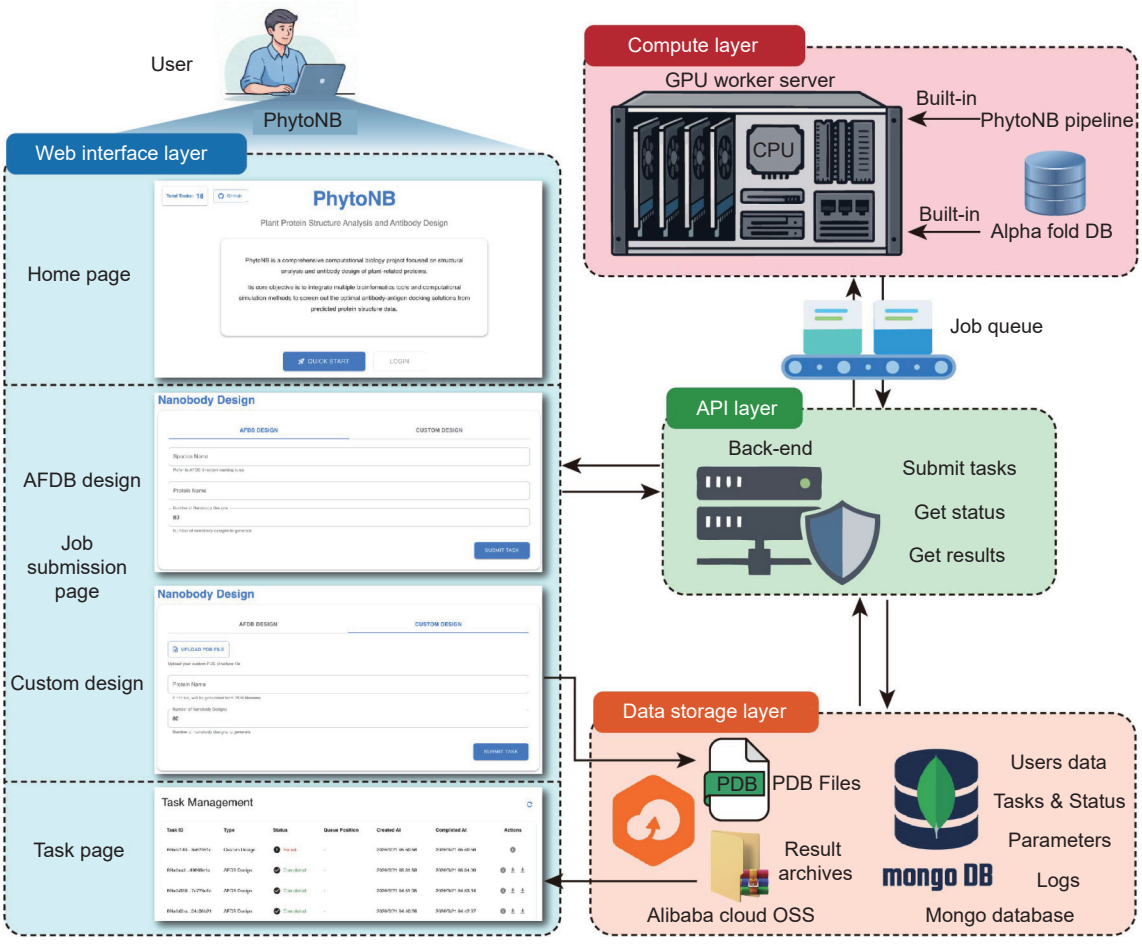

Fig. 3. Architecture of the PhytoNB web platform.
The system is organized into four layers: web interface, API, data storage, and computational layer. Users submit and monitor tasks through the web interface, while the back-end processes requests via RESTful APIs and schedules jobs to GPU worker servers. Input files and analysis outputs are uploaded to object storage (OSS), which facilitates data transfer and connectivity between the front-end and back-end. Task metadata and status information are maintained in a database system.

streamlined user experience. By integrating automated task scheduling with remote computational resources, PhytoNB allows users to perform large-scale nanobody design without requiring specialized computational environments. This design lowers the technical barrier for advanced AI-driven protein engineering and facilitates broader application in plant research.

Comparative analysis of computational efficiency

Following the detailed demonstration of the design pipeline applied to targets with diverse biological functions (Fig. 2-A to -D), Table 1 further summarizes the statistical outcomes and performance metrics of the PhytoNB platform. This table encompasses comprehensive quantitative data, ranging from protein sequence lengths to the final filtering of candidate molecules. The sequence lengths of the tested proteins span from 175 to 571 amino acids. These data intuitively illustrate the broad generalizability and system robustness of the platform in handling both endogenous plant physiological regulatory genes and exogenous agricultural trait genes.

Table 1 quantifies the substantial improvement in computational efficiency achieved through the parallel scheduling strategy, with the platform demonstrating remarkable optimization across varied hardware resources. To clearly evaluate these gains, we defined a "Speed-up Factor" representing the ratio of single-core execution time to PhytoNB's parallel execution time. On the GPU side, dynamic task allocation enables the nanobody generation phase to achieve an average 6.7-fold acceleration. For example, even for the complex, long-sequence protein ZmPDS1, the generation time was reduced from over

Table 1. Computational efficiency and design statistics of the PhytoNB platform on various plant-related protein targets.

Protein Name	Protein Length (aa)	Total Designs	Single-Core Time		Parallel Scheduling			Top-Ranked Designs
			Nanobody Generation (hh:mm:ss)	Filtering and Sorting (hh:mm:ss)	Task Count	Nanobody Generation (hh:mm:ss)	Filtering and Sorting (hh:mm:ss)	
ZmGID1	351	300	3:20:49	24:08:30	16	0:29:56	1:56:30	5
ZmPDS1	571	300	9:32:28	43:02:44	8	1:18:36	3:18:47	8
BAR	175	300	1:03:09	12:18:29	32	0:10:56	0:53:28	5
CP4EPSPS	443	300	5:30:56	31:41:34	16	0:49:08	2:34:13	7

9 hours to approximately 1 hour and 18 minutes. This acceleration effect is particularly vital for large-scale plant proteome screening, where PhytoNB transforms what was previously a day-long task into a high-throughput computational workflow.

On the CPU side, the platform addresses the most intensive computational bottleneck: the Rosetta-based filtering and sorting stage. By implementing a multi-threaded parallelization strategy, PhytoNB achieved a remarkable 12.4 to 13.8-fold increase in speed during this stage. For the ZmPDS1 target, while single-core serial processing required a prohibitive 43 hours and 2 minutes, our multi-threading optimization successfully reduced this to approximately 3 hours and 18 minutes. This order-of-magnitude improvement successfully overcomes the primary bottleneck in *de novo* design, allowing for the rapid evaluation of thousands of candidates within a manageable timeframe.

The intelligence of the platform is further reinforced by its heuristic scheduling strategy, where the "Task Count" is dynamically optimized based on protein length and structural complexity. To balance the computational load and maximize hardware efficiency, the scheduler assigns parallel granularity inversely to protein size to prevent hardware strain. For instance, the shorter BAR protein was assigned the maximum parallel granularity of 32 tasks to fully utilize available threads. In contrast, the longest protein, ZmPDS1, was intelligently allocated 8 tasks to prevent out-of-memory (OOM) errors and ensure stable execution during memory-intensive Rosetta scoring. Ultimately, this adaptive allocation yielded 5 to 8 high-quality, top-ranked designs for each target, providing a consistent and robust candidate pool for subsequent experimental validation.

Experimental validation of designed nanobodies

Based on comprehensive energy scoring and structural rationality criteria, top-ranked candidate molecules were selected from the 300 initial designs for each target. These high-quality theoretical models subsequently underwent biochemical validation via pull-down assays. This experimental approach was employed to confirm the physically interacting effects predicted through computational methods (Fig. 4).

For plant endogenous proteins, PhytoNB successfully identified molecules with binding potential. In the validation of ZmGID1 (Fig. 4-A), nanobody design #04 exhibited a clear co-precipitation signal in the experimental group. For the longer ZmPDS1 sequence (Fig. 4-B), molecule #08 was precisely identified among the eight top-ranked candidates. This nanobody demonstrated physical interaction capability. The interaction signals for endogenous proteins were relatively weak, which may be attributed to conformational dynamics. Nevertheless, these results confirm the platform's capacity to capture designs targeting complex plant endogenous targets.

In the validation of exogenous agricultural trait proteins, the platform demonstrated higher design enrichment and stronger interaction signals. For the BAR protein (Fig. 4-C), two designs identified by the system exhibited clear interaction effects. Design #04 showed positive binding, while design #05 displayed stronger binding capability. In the CP4EPSPS validation (Fig. 4-D), the platform exhibited a relatively high validation success rate. Three positive clones were successfully yielded, specifically designs #03, #07, and #02. The interaction strengths displayed a clear hierarchical distribution, with #03 showing the strongest binding, followed by #07 and then #02.

In summary, with the exception of a few sequences that failed to produce output due to heterologous expression limitations, PhytoNB achieved positive candidate molecule generation across all tested targets. This high conversion rate,

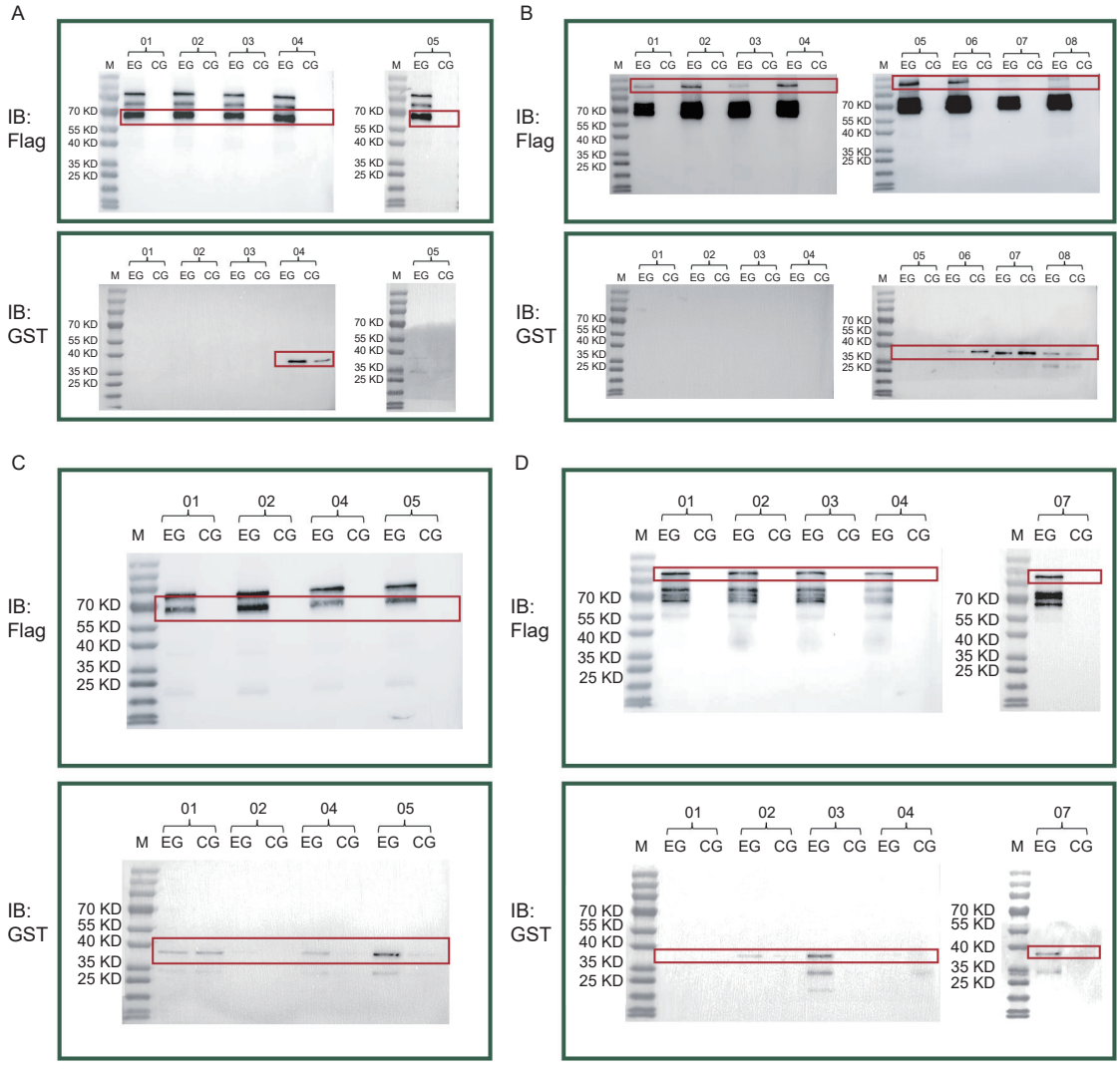

Fig. 4. Pull-down validation of interactions between PhytoNB-designed nanobodies and endogenous or exogenous maize target proteins. **A**–**D**. Western blot results are shown for four targets: the endogenous proteins ZmGID1 (A) and ZmPDS1 (B), and the exogenous proteins BAR (C) and CP4EPSPS (D). M, protein marker with molecular weights indicated in kDa. EG, experimental group containing both Flag-tagged target protein and GST-tagged candidate nanobody. CG, control group containing only GST-tagged candidate nanobody without target protein, used to assess background noise and exclude non-specific binding. Upper panels (anti-Flag) show captured target proteins, and lower panels (anti-GST) show co-precipitated nanobodies. Positive interactions are indicated by the presence of a clear GST signal in the experimental group and its absence in the control group. For endogenous targets, positive clones were identified: candidate #04 for ZmGID1 (A) and candidate #08 for ZmPDS1 (B). For exogenous targets, strong interaction signals were observed for BAR candidates #04 and #05 (C) and for CP4EPSPS candidates #02, #03, and #07 (D). Missing bands for certain constructs, such as BAR #03 in (C) and CP4EPSPS #05 and #06 in (D), were attributed to failed expression of the corresponding nanobody sequences in the heterologous expression system.

from high-throughput computation to biochemical validation, provides experimental support for the practical utility of the platform in identifying functional nanobody candidates in automated plant nanobody design and screening.

DISCUSSION AND CONCLUSION

PhytoNB establishes a fully automated and parallel-accelerated pipeline that integrates domain-level segmentation, multi-view binding site prediction, AI-driven nanobody generation, and energy-based structural filtering for plant nanobody discovery. By converting plant target selection into a scalable structure-to-binder workflow, this platform addresses a key challenge in plant biotechnology: the systematic interrogation of large, multi-domain, and underexplored proteins that are increasingly accessible through structure prediction resources (Chen et al., 2026; Zou et al., 2025). The successful application of PhytoNB to both endogenous regulatory proteins and exogenous agricultural targets demonstrates that computational nanobody design can

be effectively scaled for plant-specific applications.

The results presented here have several implications for plant immune and signaling research. PhytoNB enables targeted interrogation of receptor systems whose modularity and evolvability have been established through recent functional studies (Tsitsikli et al., 2025). The platform's ability to generate binders against specific epitopes positions it as a tool not only for discovery but also for probing receptor conformational states and stabilizing functional interfaces (Ngou et al., 2025). In particular, PhytoNB could be extended to design nanobodies as state-selective conformational probes for dynamic plant proteins, such as receptor kinases regulated by ligand binding or phosphorylation. By targeting distinct structural states (e.g., inactive versus activated conformations), these nanobodies could enable state-specific detection and stabilization of signaling intermediates *in vivo*. This is particularly relevant given that receptor activation often depends on discrete structural features rather than sequence alone, as demonstrated in NLR and resistosome assemblies (Gu et al., 2025; Guo et al., 2025). For subcellularly specialized immune components, including plastid-associated factors that shape disease resistance phenotypes (Peppino et al., 2025), PhytoNB offers a route to generate conformation-specific probes that traditional methods cannot easily provide.

From a methodological perspective, this work demonstrates the value of coupling generative design with complementary filtering strategies. The current PhytoNB pipeline leverages Rosetta for robust energy-based scoring and hydrogen bond analysis, forming a foundational evaluation layer for candidate nanobody-antigen complexes. Future iterations could further enhance this framework by integrating geometric deep learning for surface interaction fingerprinting (Gainza et al., 2020) and residue-level interaction characterization via PLIP (Salentin et al., 2015). Such extensions would reinforce the closed-loop logic of target inference, binder generation, and interaction scoring—an approach particularly valuable in plant nanobody design, where candidate numbers are large but experimental validation capacity remains limited. The parallel acceleration strategy implemented here directly addresses the computational demands of large-scale generative models, enabling screening scenarios where serial execution would be prohibitive.

The translational relevance of this work is supported by expanding experimental applications of nanobodies in plants. FLAG-tag-based workflows have demonstrated that nanobody-mediated imaging and biochemical capture can be broadly deployed in plant systems. Recent advances in synthetic antibody design further suggest that generative and ranking pipelines can be combined to improve binder specificity and functional utility (Kong et al., 2025). These developments position PhytoNB for extension beyond single-target binding toward multivalent probes, state-specific reagents, and modular tools for protein tracking and perturbation in plants.

Several limitations of the current approach should be acknowledged. First, structural prediction and interface scoring methods necessarily simplify protein dynamics and cellular context (Kryshtafovych et al., 2021). Epitope accessibility in plant systems may be shaped by cellular context and post-translational processing that are not captured by static structural models, while misfolding and aggregation during protein production can further limit functional epitope exposure *in vivo* (Beygmoradi et al., 2023). Second, despite rapid advances in plant structural biology, most deep learning models remain trained predominantly on non-plant data, potentially limiting transferability to lineage-specific domains or unusual receptor classes (Shanker et al., 2024). Third, energy-based ranking via Rosetta, while useful for prioritization and robust in assessing binding stability through hydrogen bond and energy analysis, cannot replace biochemical validation of expression, folding, and *in vivo* binding behavior (Barlow et al., 2018). Addressing these limitations will require expansion of plant-specific training data, improved modeling of conformational ensembles, and tighter integration of computational predictions with experimental feedback.

In conclusion, PhytoNB provides an integrated framework that connects plant structural prediction, epitope inference, nanobody generation, and binding-based filtering within a unified design pipeline. By combining AI-driven design with parallel computation, this platform reduces practical barriers to plant nanobody discovery and enables more systematic engineering of plant protein function. Importantly, the incorporation of a user-oriented platform extends the applicability of this framework beyond specialized computational settings, facilitating its use in broader experimental and translational contexts. Leveraging Rosetta for reliable energy scoring and hydrogen bond analysis, PhytoNB

establishes a solid foundation for candidate evaluation, with room for future enhancement through integration of complementary analytical tools. As plant structural resources continue to expand and generative methods improve, PhytoNB is well positioned to support the development of functional probes, synthetic receptors, and engineered proteins for agricultural applications.

METHODS

Data sources

In this study, predicted structural data for plant proteins were obtained from the AlphaFold Protein Structure Database, hereafter referred to as AlphaFold DB. This database provides hundreds of thousands of high-confidence protein structures for major plant species, including *Arabidopsis thaliana*, soybean (*Glycine max*), rice (*Oryza sativa*), and maize (*Zea mays*). According to the official download page, 27 40 2 predicted structures are available for *Arabidopsis thaliana*, 55 796 for soybean, 43 645 for rice, and 39 139 for maize. All these structures were generated by AlphaFold with high confidence and were used as candidate target proteins in this study. Special emphasis was placed on proteins involved in plant physiological pathways or defense against pathogens. All target protein structures were retrieved from AlphaFold DB version 6, the latest release at the time of analysis, to ensure data consistency and accuracy.

Domain segmentation

Chainsaw is a deep learning-based tool for protein domain segmentation (Wells et al., 2024). This tool takes protein three-dimensional structures as input. It employs a two-dimensional convolutional neural network to predict the probability of each residue pair belonging to the same structural domain. Compared to traditional methods, Chainsaw exhibits superior performance on the CATH domain classification benchmark. This tool correctly identifies 78% of domains, compared to 72% for the next best method (Wells et al., 2024). To achieve this, within the PhytoNB pipeline, Chainsaw is utilized to segment target protein structures into individual domains. This process decomposes multi-domain proteins into independent structural units. Consequently, this approach enhances the specificity and accuracy of subsequent binding site prediction and molecular docking.

Binding site prediction

Two state-of-the-art deep learning models are integrated synergistically for binding site prediction. GPSite is a multi-task, geometry-aware protein binding site prediction network (Yuan et al., 2024). This network is capable of simultaneously predicting residues that bind various ligands, including DNA, RNA, peptides, proteins, ATP, heme, and metal ions. It takes protein sequences as input and leverages pre-trained language models, such as ProtTrans (Elnaggar et al., 2022), and folding models, such as ESMFold (Lin et al., 2023), to obtain sequence embeddings and predicted structural information. Deep learning is then performed via a residue-level geometric graph network. GPSite enables rapid prediction without requiring multiple sequence alignments. This tool demonstrates significantly superior performance compared to existing methods (Yuan et al., 2024). Within this pipeline, GPSite is employed to detect potential binding sites on the surface of each target protein domain. This approach thereby provides explicit design directions for IgGM.

Furthermore, MVGNN is specifically designed for protein-protein interaction site prediction. Its derivative model, MVGNN-PPIS, utilizes AlphaFold3-predicted structures combined with transfer learning to construct a multi-view graph network (Meng et al., 2025). This network simultaneously incorporates k-nearest neighbor graph and adjacency matrix views. It extracts both local and global features through graph convolutional and graph transformer layers. MVGNN accurately predicts protein-protein interface residues (Meng et al., 2025). In this study, MVGNN is applied to annotate potential protein-protein binding sites on target proteins. This approach complements the predictions from GPSite and enhances the comprehensiveness of binding site identification.

Nanobody generation

IgGM is a recently introduced generative model for immunoglobulins (Wang et al., 2025). This model simultaneously outputs antibody and nanobody sequences, along with their predicted structures, for a given antigen. The core of the model comprises three main components. These components include a pre-trained language model, a feature learning module, and a sequence-structure co-prediction module (Wang et al., 2025). Upon input of an antigen and the target epitope, IgGM generates nanobody sequences with

specific binding affinities. Within this pipeline, IgGM serves as the core design engine. It takes the selected binding sites as input and produces multiple candidate nanobody sequences along with their complex structures. Given the substantial GPU memory requirements of IgGM, a dedicated parallel acceleration strategy is introduced in subsequent steps to enhance its runtime performance.

Structure scoring and filtering

Rosetta is a comprehensive software suite widely employed for molecular modeling and energy calculations (Leman et al., 2020). For each candidate nanobody–target protein complex generated by IgGM, systematic evaluation is performed using the Rosetta docking and refinement protocol "RosettaDock", followed by scoring with the standard all-atom energy function "ref2015" (Alford et al., 2017; Lyskov and Gray, 2008). These protocols, originally developed for antibody–antigen modeling, have also been shown to be applicable to nanobody design, although they may exhibit relatively high false-negative rates, necessitating extensive screening to improve hit identification (Shrestha et al., 2025).

In this work, candidate complexes are ranked using a two-step criterion based on both interaction geometry and energetic stability. Specifically, models are first prioritized by the number of intermolecular hydrogen bonds (as computed by the Rosetta hydrogen bond scoring term "hbond_sc"), and among models with comparable hydrogen bond counts, complexes with more favorable (i.e., more negative) total energy scores, calculated by "ref2015", are preferentially selected. Thus, the "low energy" criterion refers to complexes that simultaneously exhibit enriched hydrogen bonding interactions and minimized Rosetta total energy. Designs satisfying these criteria are retained for subsequent analysis. To expedite the screening process, Rosetta calculations are executed in parallel via multi-threading, enabling simultaneous evaluation of multiple complex structures.

System and computational resources

The computational platform utilized in this study is based on a high-performance GPU cluster architecture. This system is equipped with eight NVIDIA GeForce RTX 3090 graphics cards, each with 24 GB of video random access memory. Hardware resources are harnessed via CUDA and multi-process parallelization techniques. The operating system installed on this platform is Ubuntu 24.04. The central processing unit comprises dual Intel Xeon Silver 4410T processors. These processors provide a total of 40 physical cores, which support 80 logical threads. The base frequency is 2.7 GHz, and the maximum turbo frequency is 4.0 GHz. The system is equipped with 512 GB of DDR4 RAM. Regarding task scheduling, a dynamic GPU memory monitoring mechanism has been implemented. Prior to the submission of each deep learning task, such as IgGM, GPSite, or MVGNN, the system automatically queries the available memory on each GPU. The task is then assigned to a GPU with sufficient available memory. This mechanism prevents computation interruptions caused by out-of-memory errors. For computationally intensive CPU-based steps, such as Rosetta scoring, multi-threaded parallelization is employed. This approach utilizes up to 80 threads to fully leverage all CPU cores. It enables the simultaneous evaluation of multiple complex structures. The entire workflow is orchestrated by a scheduler. This scheduler ensures balanced and efficient utilization of both GPU and CPU resources.

Web platform architecture and implementation

PhytoNB adopts a front–back-end separated microservice architecture combined with containerization to ensure efficient deployment and maintenance. Given the high computational demand of protein structure analysis, particularly the reliance on GPU resources, the system is designed as a multi-layer decoupled framework consisting of a web interface layer, a data storage layer, and a computational analysis layer, interconnected via APIs. This architecture improves scalability and optimizes resource utilization. The front-end is developed using React with Vite for modularization and performance optimization. It provides a tab-based interface supporting two core modes, AFDB Design and Custom Design, and communicates with the back-end through RESTful APIs for task submission, monitoring, and result retrieval. The back-end is implemented using Node.js and Express.js, with JWT-based authentication to ensure secure API access. MongoDB is used to manage user information, task states, and configuration data. To handle unstructured data such as PDB files and result packages, the system integrates object storage services, enabling efficient cross-server data access and distribution.

Computational tasks are executed by independent worker processes, which poll the database for pending jobs and utilize GPU resources for large-scale analysis. The system is deployed using Docker containers, with multi-service orchestration and centralized environment configuration, improving both security and maintainability.

Flag-tag pull-down assay

To validate the physical interactions between the nanobodies designed by the PhytoNB platform and their target proteins, a pull-down assay employing Flag-tag magnetic beads was implemented in this study. Recombinant nanobodies were expressed in Escherichia coli BL21(DE3) cells as soluble proteins with a C-terminal "GST-tag", while target proteins were expressed with a "Flag-tag" using either E. coli or heterologous expression systems, depending on protein properties. Proteins were purified using standard affinity chromatography prior to downstream assays and adjusted to working concentrations for interaction analysis. The experimental procedure was performed as follows.

First, 20 μL of Flag magnetic beads were washed and equilibrated with TBS buffer. Subsequently, approximately 50 μg of the target protein bearing a Flag tag was added to the beads. The mixture was incubated at 25 °C with constant shaking at 1 200 r/min for 1 hour to achieve bead capture. After multiple washes with TBS buffer to remove unbound proteins, 100 μL of the GST-tagged nanobody to be validated was introduced into the system. Incubation was continued under identical conditions for an additional 1.5 hours. Following incubation, repeated TBS washes were performed, and the tubes were replaced to minimize non-specific binding.

Thereafter, 30 μL of 3×Flag peptide at a concentration of 1 mg/mL was added for competitive elution. The eluted products were denatured by adding 5×SDS loading buffer and heating at 98 °C for 10 minutes. Finally, qualitative detection was carried out using Western blotting with anti-Flag and anti-GST specific antibodies. By observing the co-precipitation of target bands in the experimental group, which contained the target protein and designed nanobody, compared to the control group, which contained empty control and designed nanobody, the physical interaction between the nanobody and the target was comprehensively evaluated.

ACKNOWLEDGEMENTS

We thank the providers of the AlphaFold Protein Structure Database for making structural data publicly available. We also acknowledge the computational resources that supported the GPU-accelerated analyses in this study.

FUNDING

This work was supported by grants from the Major Science and Technology Special Project of Liaoning Province (2025JH1/11700012).

SUPPLEMENTAL DATA

Supplemental data is available at *Maize Sciences* online.

DECLARATION OF COMPETING INTEREST

The authors declare no conflict of interest.

REFERENCES

Ahern W, Yim J, Tischer D, et al. 2026. Atom-level enzyme active site scaffolding using RFdiffusion2. *Nature Methods*, **23**(1): 96−105.

Alexander E, Leong K W. 2024. Discovery of nanobodies: a comprehensive review of their applications and potential over the past five years. *J Nanobiotechnology*, **22**: 661.

Alford R F, Leaver-Fay A, Jeliazkov J R, et al. 2017. The Rosetta all-atom energy function for macromolecular modeling and design. *J Chem Theory Comput*, **13**(6): 3031−3048.

Arbabi-Ghahroudi M. 2022. Camelid single-domain antibodies: Promises and challenges as lifesaving treatments. *Int J Mol Sci*, **23**(9): 5009.

Armstrong E M, Larson E R, Harper H, et al. 2023. One hundred important questions facing plant science: an international perspective. *New Phytol*, **238**(2): 470−481.

Barlow K A, Ó Conchúir S, Thompson S, et al. 2018. Flex ddG: Rosetta ensemble-based estimation of changes in protein-protein binding affinity upon mutation. *J Phys Chem B*, **122**(21): 5389−5399.

Bennett N R, Coventry B, Goreshnik I, et al. 2023. Improving de novo protein binder design with deep learning. *Nat Commun*, **14**: 2625.

Beygmoradi A, Homaei A, Hemmati R, et al. 2023. Recombinant protein expression: Challenges in production and folding related matters. *International Journal of Biological Macromolecules*, **233**: 123407.

Chen J, Feng Y, Zhang Y, et al. 2026. Structure-guided discovery of protein functions in plants. *Plant Cell*, **38**(2): 1−13.

Dauparas J, Anishchenko I, Bennett N, et al. 2022. Robust deep learning-based protein sequence design using ProteinMPNN. *Science*, **378**(6615): 49−56.

Elnaggar A, Heinzinger M, Dallago C, et al. 2022. ProtTrans:

Towards cracking the language of lifes code through self-supervised deep learning and high performance computing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**(10): 7112−7127.

Fleming J, Magana P, Nair S, et al. 2025. AlphaFold protein structure database and 3D-beacons: New data and capabilities. *J Mol Biol*, **437**(15): 168967.

Gainza P, Sverrisson F, Monti F, et al. 2020. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat Methods*, **17**(2): 184−192.

Gu Q, Liu S, He Z, et al. 2025. NLRs in plant immunity: Structural insights and molecular mechanisms. *Crop Design*, **4**(2): 100103.

Guo G, Zhao H, Bai K, et al. 2025. An activated wheat CCG10-NLR immune receptor forms an octameric resistosome. *bioRxiv*, **189**(10): 2955−2970.

Islam S, Park K, Xia J, et al. 2025. Structural insights into gibberellin-mediated DELLA protein degradation. *Mol Plant*, **18**(7): 1210−1221.

Jumper J, Evans R, Pritzel A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**: 583−589.

Kırboğa K K, Küçüksille E U. 2026. Integration of evolutionary analysis with RFdiffusion for *de novo* design of aggregation-resistant frataxin. *Proteins*, **94**(6): 1225−1244.

Kong Y, Shi J, Wu F, et al. 2025. A synergistic generative-ranking framework for tailored design of therapeutic single-domain antibodies. *Cell Discovery*, **11**: 85.

Kryshtafovych A, Schwede T, Topf M, et al. 2021. Critical assessment of methods of protein structure prediction (CASP)-Round XIV. *Proteins*, **89**(12): 1607−1617.

Leman J K, Weitzner B D, Lewis S M, et al. 2020. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat Methods*, **17**: 665−680.

Lin P-Y, Huang S-C, Chen K-L, et al. 2025. Analysing protein complexes in plant science: insights and limitation with AlphaFold 3. *Bot Stud*, **66**: 14.

Lin Z, Akin H, Rao R, et al. 2023. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, **379**(6637): 1123−1130.

Liu J, Wu L, Xie A, et al. 2025. Unveiling the new chapter in nanobody engineering: advances in traditional construction and AI-driven optimization. *Journal of Nanobiotechnology*, **23**: 87.

Lyskov S, Gray J J. 2008. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res*, **36**(suppl_2): W233−W238.

Meng L, Wei L, Wu R. 2025. MVGNN-PPIS: A novel multi-view graph neural network for protein-protein interaction sites prediction based on Alphafold3-predicted structures and transfer learning. *Int J Biol Macromol*, **300**: 140096.

Mou Q, Zhang J, Si Z, et al. 2026. Transgenic Lepidopteran-Pests-Resistant and Herbicide-Tolerant Cotton Through Transfer of *Cry1Ab-vip3Aa* and *Cp4-epsps+bar* Genes. *Plant Biotechnology Journal*, **0**: 1−3.

Ngou B P M, Wyler M, Schmid M W, et al. 2025. Systematic discovery and engineering of synthetic immune receptors in plants. *Science*, **389**(6764): eadx2508.

Peng Y, Liang Z, Cai M, et al. 2024. ZmPTOX1, a plastid terminal oxidase, contributes to redox homeostasis during seed development and germination. *The Plant Journal*, **119**: 460−477.

Peppino Margutti M Y, Cislaghi A P, Herrera-Vásquez A, et al. 2025. The *Arabidopsis* TNL immune receptor BNT1 localizes to the plastid envelope and is required for the flg22-induced resistance against *Pseudomonas*. *Plant J*, **122**(6): e70295.

Salentin S, Schreiber S, Haupt V J, et al. 2015. PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res.*, **43**(W1): W443−W447.

Shanker V R, Bruun T U J, Hie B L, et al. 2024. Unsupervised evolution of protein and antibody complexes with a structure-informed language model. *Science*, **385**(6704): 46−53.

Shrestha P, Talwar C S, Kandel J, et al. 2025. NanoBinder: a machine learning assisted nanobody binding prediction tool using Rosetta energy scores. *J Cheminform*, **17**: 96.

Tsitsikli M, Simonsen B, Luu T-B, et al. 2025. Two residues reprogram immunity receptors for nitrogen-fixing symbiosis. *Nature*, **648**: 443−450.

Wang R, Wu F, Gao X, et al. 2025. IgGM: A generative model for functional antibody and nanobody design. *The Thirteenth International Conference on Learning Representations*.

Wang W, Yuan J, Jiang C. 2021. Applications of nanobodies in plant science and biotechnology. *Plant Mol Biol*, **105**: 43−53.

Watson J L, Juergens D, Bennett N R, et al. 2023. *De novo* design of protein structure and function with RFdiffusion. *Nature*, **620**: 1089−1100.

Wells J, Hawkins-Hooker A, Bordin N, et al. 2024. Chainsaw: protein domain segmentation with fully convolutional neural networks. *Bioinformatics*, **40**(5): btae296.

Yuan Q, Tian C, Yang Y. 2024. Genome-scale annotation of protein binding sites via language model and geometric deep learning. *Elife*, **13**: (RP93695).

Zhu H, Ding Y. 2025. Nanobodies: From discovery to AI-driven design. *Biology*, **14**(5): 547.

Zou J, Yuan Q, Yang Y. 2025. An Online Server for Geometry-Aware Protein Function Annotations Through Predicted Structure. In L. Kurgan & D. Kihara (Eds.). *Protein Function Prediction: Methods in Molecular Biology*, **2947**: 191−208.

(Managing Editor: 封琳娜)