

Review

AI-Assisted Spatial Metabolic Engineering in Plants: Integrating Flux Design, Spatial Omics, and Synthetic Biology

Huize Chen ^{1,2}, Jia Yang ^{1,2} and Meiting Du ^{1,2,*}
¹ Shanxi Key Laboratory of Plant Macromolecules Stress Response, Taiyuan 030000, China; chenhz@sxnu.edu.cn (H.C.)

² School of Life Sciences, Shanxi Normal University, Taiyuan 030000, China

* Correspondence: 703255@sxnu.edu.cn

Abstract

Background: Plant synthetic biology reprograms metabolic networks for the sustainable production of high-value compounds. Recent computational advances incorporate machine learning to accelerate the design-build-test-learn (DBTL) cycle, enabling more predictable and scalable engineering in photoautotrophic chassis. However, the translation of AI-generated designs into stable plant phenotypes remains constrained by incomplete plant-specific training datasets, tissue heterogeneity, and limited in vivo validation. **Scope:** This review examines the convergence of machine learning methods with plant metabolic engineering across four spatial engineering levels: subcellular compartmentalization, cell/tissue/organ-specific control, developmental or inducible regulation, and genome-level organization. Spatial omics is considered a cross-cutting validation layer, and the evidence supporting each technology is classified as plant-demonstrated, non-plant proof-of-concept, or prospective. **Conclusions:** Integrating predictive machine learning with spatial engineering offers promising strategies to design complex biosynthetic pathways. Hybrid approaches, combining constraint-based metabolic models with generative algorithms, reduce trial-and-error in crop engineering. Future plant synthetic biology is likely to rely increasingly on automated and data-rich workflows to support more predictable plant bioproduction.

Keywords: plant metabolic engineering; spatial metabolomics; metabolic flux; plant synthetic biology; machine learning; plant biomanufacturing

1. Introduction: Plant Synthetic Biology in the Algorithmic Era

Plant synthetic biology applies engineering principles to systematically reprogram plant metabolic pathways [1]. Plant chassis circumvent the post-translational glycosylation bottlenecks and low solubility limits commonly encountered in prokaryotic hosts by coupling photosynthetic carbon fixation with highly compartmentalized subcellular environments, enabling the synthesis of structurally complex natural products that are otherwise challenging to produce heterologously [2]. In this review, spatial metabolic engineering refers to the deliberate redistribution and coordinated control of enzymes, transporters, cofactors, intermediates, and storage processes across multiple spatial scales in plants. These spatial engineering levels comprise (i) subcellular compartmentalization and inter-organelle routing; (ii) cell-, tissue-, and organ-specific pathway expression; (iii) developmental-stage-specific or externally inducible regulation; and (iv) genome-level or-

Academic Editor: Michał Tomczyk

Received: 24 June 2026

Revised: 12 July 2026

Accepted: 21 July 2026

Published: 23 July 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

ganization of multigene pathways. Spatial omics is treated as a cross-cutting measurement and validation layer rather than as an independent engineering scale. AI is considered a prioritization and data-integration layer operating across these levels rather than a substitute for plant experimentation.

To separate established evidence from future potential, the revised review uses three evidence categories: plant-demonstrated approaches supported by in planta metabolite, protein, or phenotype measurements; non-plant proof-of-concept approaches demonstrated in vitro, in microorganisms, or in animal cells; and prospective approaches whose application to plant metabolic engineering remains hypothetical or computationally inferred. Historically, research in this field focused primarily on optimizing the physical assembly of genetic circuits. While standardized DNA assembly frameworks, such as Golden Gate-based Modular Cloning (MoClo), have streamlined the physical assembly phase [3,4], they do not address the fundamental bottleneck located within the design phase. The multivariable nature of multi-gene expression dynamics within plant regulatory networks often eludes human intuition, where unpredictable epistatic interactions and metabolic burdens frequently arise. Machine learning algorithms address this regulatory multi-dimensionality, shifting the design-build-test-learn (DBTL) cycle from empirical screening toward a highly predictive, quantitative framework [5]. Deep learning models trained on specific multi-omics datasets computationally pre-screen circuit architectures, partially relieving the trial-and-error burden associated with multi-gene pathway construction [6].

Genomic language models have emerged as generative design engines. For instance, the Evo 2 model, a biological foundation model trained on 9 trillion DNA base pairs (the OpenGenome2 dataset) spanning all domains of life, processes a context window of 1 million tokens at single-nucleotide resolution [7]. This architecture can predict the functional impacts of genetic variation—from noncoding mutations to clinically relevant variants—without requiring task-specific fine-tuning.

However, while generalist biological foundation models like Evo 2 offer massive context windows and cross-domain zero-shot capabilities, they are trained predominantly on microbial and mammalian genomic datasets. Consequently, they often fail to fully decode the unique regulatory grammar of plants, which is characterized by highly repetitive heterochromatic regions, polyploid genomic complexities, and specific plant-specific cis-regulatory systems [8]. Nevertheless, the predictive capacity of such generalist genomic foundation models should not be directly equated with plant engineering performance. Plant genomes contain lineage-specific cis-regulatory elements, frequent whole-genome duplications, extensive repetitive sequences, and chromatin-dependent regulatory features that are underrepresented in many general biological training datasets. To address this limitation, the field has begun to move toward plant-specific genomic foundation models. A representative model is PlantGFM, a plant-oriented genomic foundation model built upon the Hyena operator architecture and pre-trained on 10.84 billion nucleotides from 12 representative plant species. Supporting a 64 kb context window, PlantGFM is optimized for the structural features of plant genomes and may enable the de novo design of novel plant-compatible genes. In initial validation trials, synthetic candidate sequences generated by PlantGFM successfully demonstrated transcription and translation in *Nicotiana benthamiana* leaves, suggesting the functional expression of generative plant DNA sequences. Importantly, detectable transcription or translation of synthetic sequences does not necessarily demonstrate predictable pathway-level function, stable inheritance, or agronomic utility. Therefore, plant-specific foundation models still require systematic validation across species, tissues, developmental stages, and environmental conditions.

Integrating this genomic foundation model with automated microfluidic screening provides a basis for a closed-loop platform. These automated infrastructures attempt to

close the loop between algorithmic design and in vivo phenotypic validation, accelerating the transition of plant biomanufacturing toward a predictable, data-driven engineering science.

The overall workflow proposed in this review is illustrated in Figure 1. AI-assisted design begins with computational prediction of genes, promoters, enzymes, and pathway architectures, followed by spatial targeting, plant-based construction, multi-omics validation, and iterative model refinement.

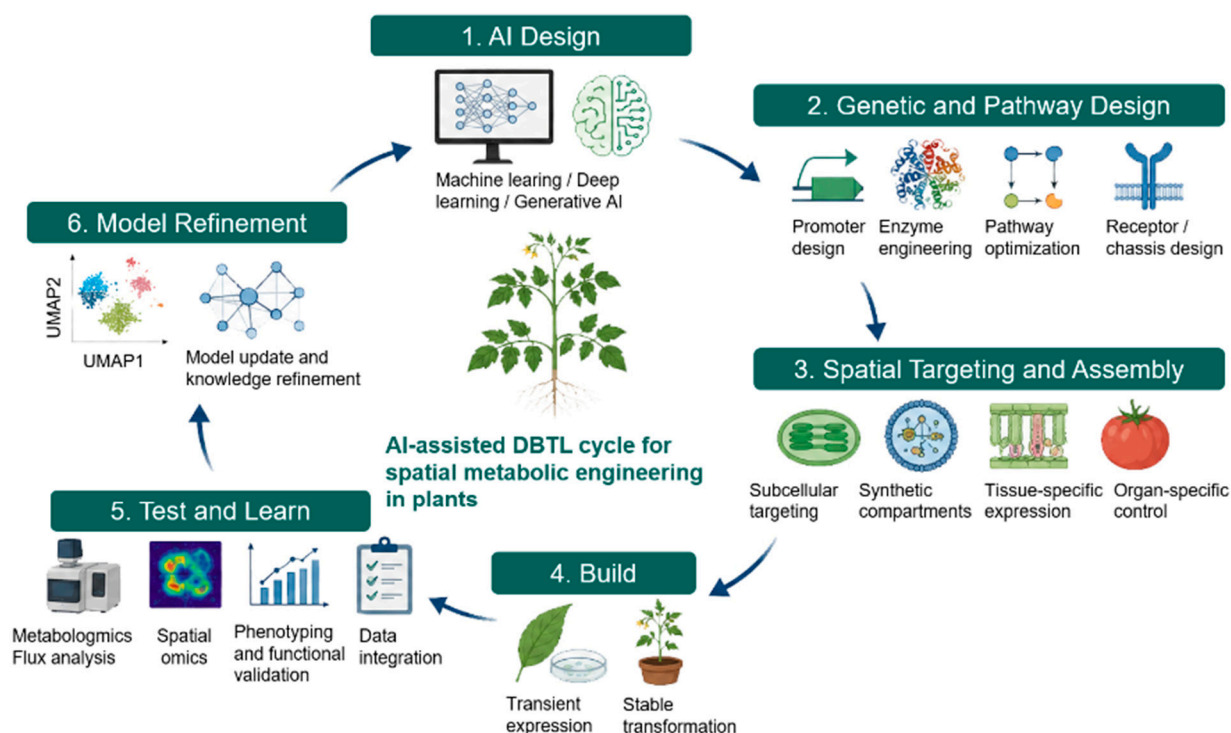


Figure 1. Conceptual framework of the AI-assisted design-build-test-learn cycle for plant spatial metabolic engineering. Machine learning and generative AI support metabolic flux prediction, pathway design, spatial targeting, plant-based construction, spatial omics-guided validation, and iterative model refinement.

By emphasizing metabolic flux prediction, pathway compartmentalization, and spatial metabolomics-guided validation, this review highlights how AI-assisted design can be connected to measurable metabolic outputs in plant systems.

To make these spatial levels explicit, Figure 2 classifies the strategies reviewed below and places spatial transcriptomics, single-cell omics, laser microdissection, and imaging mass spectrometry as a cross-cutting validation layer.

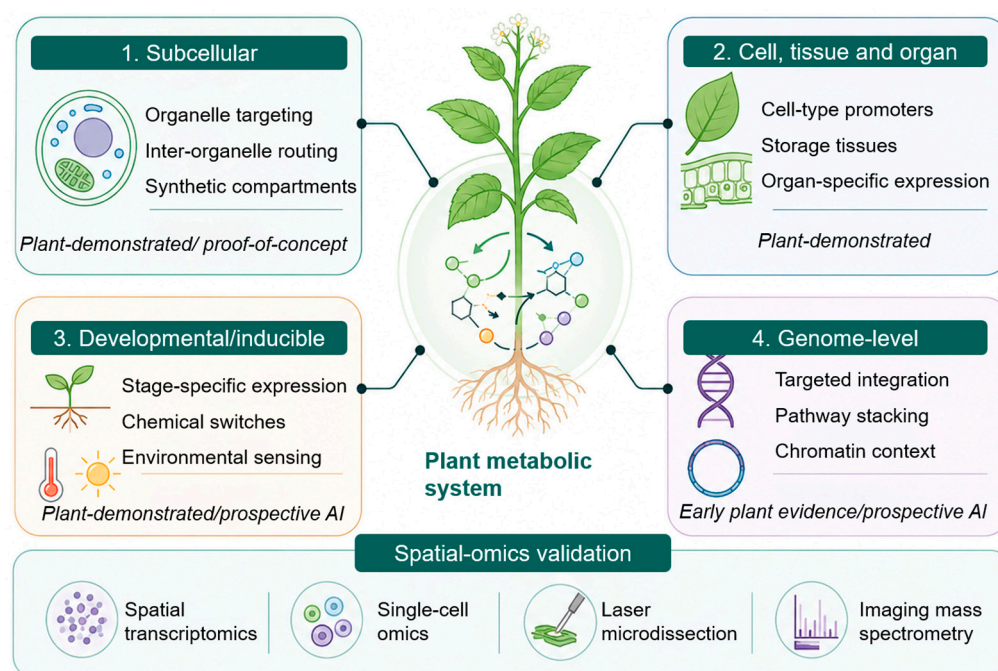


Figure 2. Spatial hierarchy and evidence levels of plant metabolic engineering. Spatial control operates at subcellular, cell/tissue/organ, developmental or inducible, and genome levels. Spatial transcriptomics, single-cell omics, laser microdissection, and imaging mass spectrometry form a cross-cutting validation layer that links computational prioritization to measured flux, product localization, transport, storage, and toxicity.

2. Predictive Flux Optimization: From Metabolic Network Modelling to Pathway Balancing

The Push-Pull-Block (PPB) framework provides a practical logic for redirecting metabolic flux toward desired products. In this framework, precursor supply is enhanced through “push” strategies, product-forming reactions are strengthened through “pull” strategies, and competing endogenous sinks are reduced through “block” strategies [9]. In plant chassis, however, the effective implementation of PPB strategies requires more than the selection of a few pathway genes. It depends on the accurate identification of rate-limiting reactions, competing branches, transport steps, cofactor constraints, and tissue- or organelle-specific metabolic bottlenecks.

Constraint-based modelling has long provided a useful computational basis for metabolic flux analysis. Flux Balance Analysis (FBA), for example, defines a feasible flux space by requiring mass balance under reaction-capacity constraints, commonly represented as $S \cdot v = 0$, where S is the stoichiometric matrix and v is the vector of reaction fluxes. Parsimonious FBA further narrows this solution space by assuming that cells tend to avoid unnecessary flux or enzyme investment. Although these methods are valuable for identifying potential intervention points, they remain limited by steady-state assumptions and incomplete metabolic annotations. These limitations are particularly important in plants, where metabolism is shaped by organellar compartmentation, diel regulation, cell-type specialization, source–sink transitions, developmental stage, and environmental responses. As a result, flux prediction pipelines developed in microbial or mammalian systems cannot be directly transferred to plant metabolic engineering without careful adaptation.

To overcome some of these limitations, recent systems biology approaches have increasingly integrated constraint-based modelling with supervised machine learning. Models such as Random Forest, XGBoost, and multilayer perceptrons can use tran-

scriptomic, proteomic, metabolomic, and fluxomic datasets to predict metabolic fluxes and prioritize candidate engineering targets [10]. Compared with purely stoichiometric models, these data-driven approaches can better capture non-linear relationships between gene expression, enzyme abundance, and metabolic output. Nevertheless, unconstrained machine-learning models may generate predictions that are statistically accurate but biologically implausible, especially when they violate mass balance, thermodynamic feasibility, or known pathway architecture. Hybrid frameworks, including multi-omics-integrated neural networks and dynamic artificial metabolic networks, attempt to reduce this problem by embedding stoichiometric or mechanistic constraints into model training [11]. These approaches are conceptually important because they combine the flexibility of machine learning with the biological realism of genome-scale metabolic models.

Despite these advances, plant flux prediction remains far from fully reliable. Genome-scale metabolic models for plants are often incomplete, especially for specialized metabolism, transport reactions, and compartment-specific metabolite pools. Many secondary metabolic enzymes remain poorly annotated, and the same pathway may behave differently across species, organs, developmental stages, and environmental conditions. Therefore, computational flux predictions should be viewed as tools for candidate prioritization rather than definitive evidence of pathway performance. Experimental validation remains essential, particularly through isotope-labelling, targeted metabolomics, enzyme assays, and spatially resolved metabolite analysis.

Several recent examples illustrate how flux balancing can support plant-based production of high-value metabolites. In *Nicotiana benthamiana*, the citrus-derived flavone glycoside diosmin has been reconstructed through transient co-expression of a multi-gene flavonoid pathway, yielding 37.7 µg/g fresh weight, corresponding to approximately 61.95 nmol/g FW, as calculated from the reported mass-based yield. [12,13]. This provides a useful quantitative benchmark for transient pathway reconstruction in leaves [14]. Similarly, multivariate modular optimization has been used to balance gene expression ratios within heterologous glucoraphanin biosynthetic modules, reducing intermediate accumulation and improving pathway performance [15,16]. These studies demonstrate that multi-gene pathway engineering in plants requires coordinated control of pathway dosage rather than simple overexpression of individual enzymes.

However, transient expression in *N. benthamiana* does not necessarily predict stable production in crops or long-term field performance. High-level expression of complex synthetic pathways can impose a metabolic burden, interfere with endogenous growth and defence programmes, or cause accumulation of toxic intermediates. Therefore, future pathway optimization will require dynamic and context-sensitive regulation rather than constitutive overexpression. AI-assisted promoter engineering provides one possible route toward this goal. For example, generative models trained on plant promoter datasets may help design synthetic promoters with tunable expression strength, enabling more precise control of pathway activation [17]. Such tools could be particularly useful for coordinating multi-enzyme pathways, reducing metabolic burden, and restricting expression to specific tissues or developmental stages.

At the same time, promoter activity measured in transient reporter assays should be interpreted cautiously [18]. Expression strength may change substantially in stable chromosomal contexts because of local chromatin state, DNA methylation, copy number variation, developmental regulation, and environmental responsiveness. Thus, AI-designed promoters require validation not only in transient assays but also in stable transgenic lines and, ultimately, under agronomically relevant conditions.

Accurate analytical feedback is also essential for closing the DBTL cycle. Targeted metabolite extraction, chemical derivatization, mass spectrometry, isotope tracing, and spatial metabolomics provide the experimental evidence needed to evaluate whether pre-

dicted flux redirection has actually occurred [19]. Without such feedback, computationally optimized designs may remain disconnected from plant physiological reality.

An evidence-gated validation sequence is therefore recommended: computational prediction and uncertainty ranking → transient plant expression → enzyme assays and isotope tracing → targeted metabolomics and spatial metabolite imaging → stable transformation and inheritance testing → field-relevant validation. Each step addresses a different failure mode. Transient assays test construct function and dosage; isotope tracing tests pathway flux; spatial measurements identify producing cells, transport routes, storage sites, and toxicity zones; stable lines reveal chromosomal and developmental effects; and field experiments quantify genotype-by-environment interactions. Computational output should therefore be treated as a ranked hypothesis until the relevant experimental gates have been passed [20–22].

Overall, machine learning can improve the identification of metabolic intervention points, but flux optimization alone is insufficient for plant synthetic biology. Once candidate pathways, enzymes, or promoters have been prioritized, their performance depends strongly on where they are placed within the plant cell or tissue. Enzyme localization, cofactor availability, organellar transport, intermediate diffusion, and tissue-specific expression all determine whether predicted flux improvements can be translated into functional biosynthesis. This spatial dimension provides the rationale for the subcellular and tissue-level engineering strategies discussed in the next section.

3. Architecting Subcellular Space: From Transit Peptides to Synthetic Organelles

3.1. Organelle-Targeting Peptides and Short Spatial Tags

At the subcellular level, spatial metabolic engineering involves the deliberate organization of enzymes and reactions across organelles, membranes, and synthetic compartments. Targeting entire heterologous metabolic cascades to specific subcellular compartments isolates reactions from cytosolic feedback inhibition, exploits localized cofactor pools, and mitigates intermediate toxicity. Subcellular localization predictors are therefore essential for the rational design of non-native transit peptides. DeepLoc 2.0, which leverages pre-trained protein language model (pLM) embeddings, achieves a state-of-the-art Matthews Correlation Coefficient (MCC) of 0.90 for predicting plastid-targeted proteins [23].

Advanced machine learning architectures, such as the attention-based neural networks developed [24], have demonstrated that deep learning can capture complex, non-linear sequence features to generate functional *de novo* signal peptides, surpassing traditional profile-based models in both targeting accuracy and sequence diversity. Mechanistic interpretability analyses of the attention layers in TargetP 2.0 and DeepLoc 2.0 revealed that the second residue following the initial methionine (position 2) exerts a dominant influence on chloroplast import classification [25]. Specifically, these neural networks identified that approximately two-thirds of chloroplast and thylakoid transit peptides possess an alanine at position 2, compared with only 20% in other endogenous plant proteins.

However, synthetic biologists must evaluate this structural feature with caution to avoid computational misinterpretations. In plant genetic engineering, an alanine codon (GCT or GCC) is frequently inserted at position 2 as an artificial cloning scar—often a direct consequence of engineering an *Nco*I restriction endonuclease site (CCATGG) to facilitate plasmid construction, or optimizing Kozak-like translation initiation contexts in plant expression cassettes [26]. Training biological language models on uncured transgenic sequences containing these synthetic constructs introduces a systematic training

bias, wherein the model misattributes high import-probability scores to an artificial translation enhancer rather than to a genuine evolutionary sorting signal [23]. Consequently, de novo transit peptide design must partition training data to filter out these non-natural sequence artifacts, ensuring that predictive algorithms capture only the genuine biophysical constraints of natural plastid import pathways. To circumvent these dataset-specific biases, recent computational efforts have shifted toward model architectures tailored specifically for photoautotrophic systems. A representative framework is Chlamy_ChloroPred, a deep learning-based binary classifier that couples ProtBERT-BFD protein language model embeddings with stacked bidirectional long short-term memory (BiLSTM) networks and an attentive pooling layer [24]. By extracting locality-aware features within the highly variable N-terminal ~50-amino-acid region of transit peptides, Chlamy_ChloroPred achieves a binary classification accuracy of 0.8462 for the *Chlamydomonas reinhardtii* proteome. Remarkably, despite being trained exclusively on microalgal datasets, Chlamy_ChloroPred exhibits effective cross-species transferability, predicting the chloroplastic proteome of the land plant *Arabidopsis thaliana* with an accuracy of 0.7316, representing a 12.6% improvement over the generalized TargetP 2.0. The integration of such interpretability-focused, lineage-specific networks offers a vital computational filter to distinguish genuine, evolutionarily conserved targeting signals from artificial cloning scars. Despite their utility, localization predictors mainly infer targeting probability rather than import efficiency, processing accuracy, enzyme folding, or catalytic competence after import. Therefore, predicted localization should be experimentally verified using fluorescent fusion proteins, immunoblotting of processed peptides, and subcellular metabolite profiling. Short peptides serve several distinct spatial functions and should not be treated as a single design class. Cleavable transit peptides deliver full-length enzymes to chloroplasts or mitochondria; signal peptides and retention motifs route proteins through the endoplasmic reticulum and Golgi; 15–20-residue encapsulation peptides recruit cargo to bacterial microcompartment shells; and designed interaction fragments may modulate protein assemblies or condensates. AI can rank candidate sequences by predicted targeting or interaction propensity, but cleavage efficiency, import kinetics, cargo stoichiometry, proteolytic stability, off-target localization, and catalytic competence must be measured experimentally.

3.2. Inter-Organelle Routing of Plant Metabolic Pathways

When engineered with precise biophysical constraints, compartmentalization within the plastidial stroma is a highly effective strategy for capturing the rich geranylgeranyl diphosphate (GGPP) precursor pool generated by the endogenous methylerythritol phosphate (MEP) pathway [15]. A prominent example is the metabolic reconstruction of early paclitaxel (Taxol) intermediates in *Nicotiana benthamiana*. By targeting the soluble Taxus-derived taxadiene synthase (TS) to the chloroplast stroma using a cleavable transit peptide, researchers achieved a taxadiene accumulation of 56.6 µg per gram of fresh leaf weight (µg/g FW). This example illustrates that spatial optimization is not simply a matter of placing all enzymes in the same compartment. Instead, successful engineering requires matching each enzyme with the membrane environment, redox partner, cofactor availability, and intermediate transport route required for catalytic function.

However, scaling this pathway to synthesize downstream oxidized taxanes introduces severe biophysical and spatial barriers. The subsequent oxidation step is catalyzed by taxadiene-5 α -hydroxylase (T5 α H, a member of the CYP725A4 family), a hydrophobic cytochrome P450 monooxygenase that natively anchors to the endoplasmic reticulum (ER) membrane and requires an uninterrupted electron supply from its membrane-bound redox partner, NADPH-cytochrome P450 reductase (CPR). Native TS localizes to the chloroplast, whereas T5 α H and CPR are associated with the endoplasmic reticulum, creating

an initial spatial separation between the precursor-producing enzyme and the downstream P450 reaction. Co-expression of the enzymes in their native configurations did not produce detectable taxadiene-5 α -ol, indicating that inter-organelle separation was a major pathway constraint. To overcome this limitation, Li et al. redirected a truncated T5 α H-CPR fusion to the chloroplast using the transit peptide of TS, thereby co-localizing the committed biosynthetic reactions within the plastid. Further enhancement of the chloroplastic precursor pool through DXS and GGPPS expression increased taxadiene production to $56.6 \pm 3.2 \mu\text{g g}^{-1}$ FW, while the optimized chloroplastic P450 configuration produced $1.3 \pm 0.5 \mu\text{g g}^{-1}$ FW taxadiene-5 α -ol [15].

Rather than solely relying on downstream pathway tuning to resolve this bottleneck, Edgar et al. demonstrated that rational engineering of TS itself can bypass downstream P450 limitations. By performing site-directed mutagenesis on the cyclase active site, they achieved a 2.4-fold improvement in yield and selectivity for an alternative cyclization product, taxa-4(20)-11(12)-diene, and for the Taxol precursor taxadiene-5 α -ol when co-expressed with CYP725A4, highlighting how upstream enzymatic optimization can alleviate downstream flux bottlenecks [15,27].

3.3. Synthetic Compartments and Encapsulation Peptides

Beyond exploiting endogenous lipid-bound compartments, an emerging direction in spatial metabolic engineering is the deployment of synthetic bacterial microcompartments (BMCs) as functional, proteinaceous organelles within plant cells [28]. BMC shell proteins and carboxysome components have been explored as modular scaffolds for plant synthetic biology, particularly in attempts to introduce carbon-concentrating compartments into tobacco chloroplasts. However, fully functional and broadly applicable synthetic BMC nanoreactors have not yet been established in plants [29]. In addition to plastidial targeting, the engineering of other organellar genomes, such as mitochondrial DNA base editing via mitoTALEN and mitoTALECD platforms, has emerged as a stable methodological route to reprogram organellar physiology, further expanding the cellular space available for compartmentalized metabolic engineering [30]. However, the deployment of BMCs in plant cells remains at an early proof-of-concept stage. Key unresolved issues include cargo loading efficiency, shell stability during plant development, possible interference with chloroplast physiology, metabolite permeability, and the energetic cost of producing large proteinaceous compartments.

These synthetic chloroplastic scaffolds can be functionalized by appending short (15 to 20 amino acids) encapsulation peptides (EPs) that fold into amphipathic α -helices to target cargo enzymes to the BMC lumen. Encapsulation of IspG and IspH may facilitate enzyme colocalization and locally organize terminal MEP-pathway reactions, although shell permeability to charged substrates, products, and cofactors remains a major unresolved constraint. Quantitative in vitro and computational molecular permeability measurements on *Haliangium ochraceum* BMC shells by researchers suggested that while the uncapped shell does not present an impenetrable barrier to small metabolic intermediates, it functions as a semi-permeable scaffold that can be biochemically tuned to partition targeted enzymes and substrates, providing critical biophysical validation for synthetic organelle designs [31]. More precisely, stopped-flow spectrophotometric assays evaluating the transport kinetics of small molecules across a 40 nm icosahedral *Haliangium ochraceum* BMC shell revealed that the intact, capped (fully assembled) shell imposes a distinct diffusion barrier, slowing downstream substrate-product exchange compared to free enzymes. Biophysical modeling of these transport kinetics demonstrated that capped shells are approximately 50-fold less permeable to substrates than their uncapped (vacant-vertex) counterparts [32]. This 50-fold permeability differential highlights a critical design trade-off in spatial metabolic engineering: while a fully sealed protein shell effectively se-

questers volatile, toxic, or gaseous intermediates (such as isoprenoid precursors), it simultaneously restricts the passive influx of bulky hydrophilic cofactors like ATP and NADP. Consequently, the rational design of chloroplastic nanofactories must balance shell tiling stoichiometry and vertex capping efficiency to optimize pathway flux without starving encapsulated enzymes of essential energetic cofactors.

3.4. Spatial-Omics-Guided Mapping and Validation

To validate the efficacy of these subcellular interventions and confirm the tight confinement of metabolic pathways, spatial metabolomics techniques are important. Mass spectral imaging (MSI) provides label-free, high-resolution mapping of localized metabolites across single-cell boundaries and tissue layers [33]. By overlaying spatial metabolite profiles onto tissue anatomy, MSI allows researchers to assess whether engineered pathways alter local metabolite distribution and reduce interference with native cytosolic metabolism, confirming the physical containment of target products within engineered subcellular domains [33].

A major missing layer in current AI-assisted spatial metabolic engineering is spatially resolved biological data. Most computational models are trained on bulk transcriptomic, proteomic, or metabolomic datasets, which average signals across heterogeneous tissues and obscure cell-type-specific metabolic states. Spatial transcriptomics, single-cell multi-omics, imaging mass spectrometry, and spatial metabolomics can provide the anatomical and cellular context required to train plant-specific predictive models. These methods are particularly important for engineering pathways in tissues with strong metabolic zonation, such as leaves, roots, seeds, glandular trichomes, vascular tissues, and developing fruits.

Nevertheless, spatial omics remains technically challenging in plants because of cell walls, high vacuolar content, strong autofluorescence, specialized metabolites, and difficulties in preserving labile compounds during tissue sectioning. Therefore, future AI-guided spatial engineering should integrate bulk omics, single-cell data, spatial metabolite imaging, and targeted biochemical validation rather than relying on a single data modality.

The analytical modalities provide complementary information. Spatial transcriptomics can identify tissue domains in which pathway genes and transporters are co-expressed, whereas single-cell or single-nucleus RNA sequencing can resolve rare biosynthetic cell types and developmental trajectories [21,22]. Laser-capture microdissection followed by targeted metabolomics links defined anatomical regions to metabolite abundance and is particularly useful for vascular bundles, glandular tissues, seed coats, or storage parenchyma [20]. Imaging mass spectrometry can then map products and intermediates without relying on reporter genes. Integrating these datasets allows AI models to infer candidate bio-synthetic sites, source–sink transport routes, storage tissues, and zones in which toxic intermediates accumulate; these inferences still require targeted biochemical validation. Representative plant-based case studies and their spatial engineering outcomes are summarized in Table 1.

Table 1. Plant-based case studies illustrating spatial strategies, metabolic outcomes, and unresolved bottlenecks.

Plant Chassis	Target Pathway/Product	Spatial Strategy	Reported Outcome	Remaining Bottleneck
<i>N. benthamiana</i>	Diosmin/flavonoid pathway	Transient reconstitution of ten diosmin-biosynthetic genes organized into three multigene modules in leaves	37.7 µg/g FW diosmin [12]	Transient output does not establish stable inheritance, tissue portability, or field performance
<i>N. benthamiana</i>	Taxadiene and taxadien-5α-ol	Chloroplast targeting of TS and a transit-peptide-directed truncated T5αH-CPR fusion, combined with DXS/GGPPS-mediated precursor enhancement	56.6 ± 3.2 µg/g FW taxadiene and 1.3 ± 0.5 µg/g FW taxadien-5α-ol [15]	Inter-organelle transfer, P450 redox coupling, and downstream oxidation remain limiting
<i>N. tabacum</i>	Artemisinic-acid pathway	Combinatorial nuclear supertransformation of a transplastomic recipient line (COSTREL)	120.4 ± 42 mg/kg FW artemisinic acid in the highest-performing T1 line, representing an increase of up to 77-fold relative to the recipient line [34]	Large genotype-to-genotype variation and the need to balance plastid and nuclear expression
<i>S. lycopersicum</i>	Anthocyanin biosynthesis	Fruit-specific expression of the snapdragon transcription factors Del and Ros1	Anthocyanin concentrations comparable to those of blackberries and blueberries, accompanied by an approximately threefold increase in hydrophilic antioxidant capacity [35]	Developmental promoter dependence, metabolic burden, and stability across cultivars and environments

Note: Reported outcomes reproduce quantitative or qualitative findings from the cited primary studies. The remaining bottlenecks represent the authors' critical synthesis based on the experimental systems and limitations described in those studies.

These cases show that spatial engineering can act through organelle targeting, inter-organelle partitioning, stable plastid–nuclear pathway distribution, and organ-specific pathway activation. The comparison also highlights why yield alone is insufficient: spatial localization, stability, transport, toxicity, inheritance, and environmental robustness must be evaluated together.

4. Secretory Pathway Trafficking and Glycoengineering

Spatial control in plant synthetic biology is not limited to small-molecule biosynthesis; it is also central to molecular farming, where recombinant protein folding, glycosylation, and accumulation depend on trafficking through the plant secretory pathway [36]. Transient expression using optimized viral vectors represents the current commercial standard [37]. The pEAQ-HT system, leveraging elements from the Cowpea Mosaic Virus and the P19 silencing suppressor, uncouples target mRNA translation from viral replication. This architecture drives expression limits up to nearly 50% of total

soluble protein in leaves. Similarly, deconstructed viral vector systems like MagnICON exhibit high-yield metrics; for instance, the expression of the hepatitis B virus core antigen (HBcAg) achieved yields of 2.4 mg/g of fresh leaf biomass, assembling into highly immunogenic virus-like particles (VLPs) [38]. However, high transient expression does not necessarily ensure correct folding, homogeneous glycosylation, or scalable downstream purification. For pharmaceutical proteins, product quality, batch consistency, host-cell protein contamination, and regulatory comparability remain as important as expression yield.

Accordingly, this section addresses compartment-specific synthesis and processing of recombinant proteins rather than direct control of small-molecule flux. The relevant spatial variables are ER entry, Golgi transit, retention or secretion, cell-type expression, and the localization of host glycosyltransferase activity.

The viability of molecular farming to address global health emergencies was distinctly validated by the regulatory progression of Medicago's Covifenz vaccine [39]. The regulatory trajectory of plant-made vaccines demonstrates the feasibility of the platform, but it also highlights that commercial success depends on manufacturing scale, cost competitiveness, public acceptance, and regulatory continuity. However, the immunogenicity of plant-specific N-glycans (specifically α -1,3-fucose and β -1,2-xylose linkages) remains a primary barrier to human therapeutics [40]. To humanize these glycan profiles, multiplex CRISPR/Cas9 editing has been employed to knock out glycosyltransferase genes in *Nicotiana benthamiana*. Researchers simultaneously targeted all seven glycosyltransferase homologs—including five α -1,3-fucosyltransferase genes and two β -1,2-xylosyltransferase genes (specifically targeting the previously omitted FucT5)—using multiplex CRISPR/Cas9 editing to generate stable, Cas9-free, homozygous glycoengineered *Nicotiana benthamiana* lines [41]. A remaining challenge is that humanized N-glycosylation is protein-dependent. Removal of plant-specific glycans reduces immunogenic risk, but it does not automatically generate fully human-like glycoforms for every recombinant protein. Given the high polyploid complexity and functional redundancy within the *Nicotiana benthamiana* genome, targeting all seven glycosyltransferase loci is necessary to minimize residual activity, as the previously neglected *FucT5* homolog can exert residual or conditional α -1,3-fucosyltransferase activity. Recent multiplex CRISPR/Cas9 editing pipelines utilizing non-repetitive guide RNA arrays successfully achieved complete, simultaneous knockouts of all seven glycosyltransferase loci in the T0 generation (specifically lines HL40 and HL64), carrying frame-shift indels up to 26 bp [41]. Subsequent segregation screening in the T1 and T2 generations isolated stable, homozygous lines that were fully Cas9-free. Removing the Cas9 transgene not only stabilizes the edited loci against further nuclease activity and off-target mobilization, but also aligns the resulting molecular farming platform with international biosafety standards, significantly lowering regulatory barriers and easing public acceptance for plant-produced human therapeutics. This strategy can be viewed as engineering of secretory-pathway processing, because glycan maturation is progressively shaped as proteins move from the ER to the Golgi apparatus [42]. The resulting engineered lines produce recombinant proteins with N-glycans lacking the major plant-specific β 1,2-xylose and core α 1,3-fucose residues.

The editing challenge in *N. benthamiana* is amplified by its polyploid-derived and highly duplicated genome. Extensive homeologous gene retention means that several functionally redundant copies can contribute to the same glycosylation reaction; disrupting only one or a subset of homologues may therefore leave residual activity [43]. Multiplex editing must also distinguish genuine paralogues from assembly artefacts and should be followed by segregation, off-target assessment, glycan profiling, and evaluation of growth and fertility.

5. Cellular and Tissue-Level Spatial Dynamics: AI-Guided Receptor Design

Reprogramming metabolic and water flux across plant tissues also demands spatial control at cellular and stomatal levels. Engineering crops for climate resilience requires interventions that do not inherently suppress vegetative growth or biomass yield. Synthetic immunity provides a framework to decouple stress tolerance from growth-yield penalties by deploying programmable genetic circuits and orthogonal receptor systems [1].

An early example of this cellular-level spatial control is the development of PYR1-Mandi, a modified abscisic acid receptor with nanomolar sensitivity to the fungicide mandipropamid [44,45]. During water deficits, plants natively produce elevated levels of ABA to control guard cell aperture and promote drought tolerance. Utilizing structure-guided rational design and yeast-two-hybrid screening, researchers mutated six residues within the PYR1 ligand-binding pocket to shift its sensitivity from ABA to mandipropamid [46]. In transgenic plants, mandipropamid activated PYR1-Mandi-dependent ABA signalling, promoted stomatal closure, and reduced water loss [47]. This strategy is conceptually powerful, but field-level deployment would require careful evaluation of ligand persistence, off-target agrochemical effects, environmental safety, cost, and regulatory acceptance. Beyond its utility as a bespoke chemical switch, the PYR1-HAB1 chemically induced dimerization (CID) system is increasingly recognized as a highly programmable, generic scaffold for broad-spectrum small-molecule sensing due to its unique allosteric amplification mechanism [45]. While ligand binding occurs solely within the malleable pocket of the PYR1 monomer, subsequent recruitment of the PP2C coreceptor (such as HAB1 or ABI1) stabilizes the complex, lowering the ligand off-rate and boosting the apparent sensitivity of micromolar binders up to 100-fold *in vivo*. Capitalizing on this ‘antibody-like’ simplicity, recent high-throughput screens utilizing engineered PYR1 double-site mutant (DSM-Hao) and triple-site mutant (TSM) libraries against 2726 diverse ligands successfully isolated functional sensors for 181 structurally distinct molecules—including plant natural products, the explosive 2,4,6-trinitrotoluene (TNT), and per- and polyfluoroalkyl substances (PFAS)—achieving an overall success rate of 6.6% [47]. This provides a possible pipeline where fixed-conformer docking and rigid-body alignment algorithms can rapidly repurpose the START domain of PYR1 into a target-driven biosensing platform, bypassing the historical constraints of empirical trial-and-error mutagenesis. However, broad ligand responsiveness may also increase the risk of unintended activation by structurally related environmental compounds. Therefore, receptor engineering should balance sensitivity, specificity, reversibility, and ecological safety.

While PYR1-Mandi demonstrated the feasibility of chemical ligand-receptor hijacking via classic structural biology, AI-based screening may shorten the early stages of receptor engineering. Modern machine learning docking models, generative structural predictors, and deep learning-based activity engines can evaluate millions of virtual mutations *in silico*, calculating binding free energies for novel ligand-receptor pairs before physical synthesis starts [48]. This approach transitions receptor engineering from empirical mutagenesis to target-driven, computational design. When coupled with tissue-specific promoters or synthetic metabolic modules, such engineered receptors could provide inducible control over metabolic fluxes in defined cell types or organs.

For plant breeders, AI is most useful when it augments rather than replaces expert intuition. Breeders can use high-dimensional genomic, phenomic, environmental, and management data to rank crosses, identify non-obvious genotype-by-environment interactions, and quantify uncertainty, while retaining biological knowledge to reject

implausible recommendations [49]. Further progress requires interpretable feature attribution, uncertainty calibration, transfer learning across germplasm and environments, standardized phenotyping, and prospective multi-location trials. For tissue-level control specifically, model predictions must also be validated for cell-type specificity, inducibility, reversibility, developmental stability, and unintended effects on growth and reproduction.

6. Genomic Circuitry: Artificial Chromosomes and Chromatin Topology

Beyond cellular signalling modules, scalable plant engineering also requires genome-level strategies that can maintain large synthetic pathways across generations. Scaling metabolic engineering requires the simultaneous transfer of entire multi-gene pathways, often exceeding 50 kb in size. Standard *Agrobacterium*-mediated integration results in random genomic insertions, leading to unpredictable positional effects and severe epigenetic silencing. Engineering synthetic plant chromosomes or utilizing targeted integration sites may provide a route for modular pathway stacking [50]. However, synthetic chromosome engineering in plants remains less mature than in yeast or mammalian systems. Stable maintenance, meiotic transmission, centromere function, epigenetic silencing, and transformation efficiency remain major barriers.

Plant artificial-chromosome technology is therefore relevant as a long-term pathway-stacking strategy, but it is not yet a routine engineering platform. Engineered maize minichromosomes have been generated by telomere-mediated truncation and by *in vitro* assembly, demonstrating that autonomous or truncated chromosome platforms are biologically possible [51–53]. Their broader use remains limited by low recovery frequency, rearrangement, variable meiotic transmission, centromere epigenetics, loading of large DNA cargoes, and species-dependent transformation. Engineered minichromosomes should not be confused with polytene chromosomes. Polytene chromosomes arise through repeated DNA replication without cell division in specialized cells and are not currently established as artificial vectors for transferring multigene pathways into crops.

Sequence-based models developed primarily in mammalian systems, including Akita [54] and Orca [50], illustrate how DNA sequence can be used to predict aspects of three-dimensional genome organization. Their application to plant chromatin and transgene landing-pad selection remains prospective and requires benchmarking with plant Hi-C and transgene-expression datasets. By predicting topologically associating domains (TADs) and boundary elements, these tools may help researchers prioritize candidate genomic landing pads—stable chromosomal loci insulated from the suppressive spread of heterochromatin. Because most 3D genome prediction models have been developed using animal or human datasets, their direct applicability to plant genomes requires careful benchmarking using plant Hi-C, epigenomic, and transgene-expression datasets. In principle, directing CRISPR-Cas9 to candidate landing pads could improve targeted integration and reduce positional effects, although stable inheritance and expression of large synthetic operons in plants remain to be systematically validated.

7. Generative AI: From Structural Prediction to De Novo Protein Design

The integration of specialized AI tools has transitioned from analyzing biological structures to explicitly generating them. The latest iteration of predictive models, AlphaFold 3, expands modeling parameters beyond single-chain proteins. By replacing previous architectures with the Pairformer module, it improves the modelling of protein–ligand, protein–nucleic acid, and multi-chain complexes [55,56]. This capacity to directly model molecular interactions reduces barriers in understanding enzyme–substrate

transitions in complex plant metabolic nodes. Although AlphaFold 3 improves molecular interaction modelling, predicted binding poses or complex structures should not be interpreted as direct evidence of catalytic activity. Enzyme engineering still requires kinetic assays, substrate specificity tests, and validation under plant-relevant pH, redox, and cofactor conditions.

These models predict intermediate representations rather than plant performance. AlphaFold 3 can propose structures and interaction geometries, but it does not establish catalytic turnover, substrate specificity, product profile, compartment-specific folding, or metabolic flux. RFDiffusion [57] generates candidate backbones, and ProteinMPNN [58] proposes sequences compatible with a backbone, but neither guarantees expression, stability, cofactor use, or activity in a chloroplast, ER, vacuole, or cytosol. Promoter and genomic foundation models similarly estimate sequence–activity relationships within their training domain; they do not directly predict chromatin effects, stable inheritance, environmental responsiveness, or agronomic phenotype.

Protein design has expanded beyond structural prediction to de novo backbone generation [59]. While earlier iterations required predefined scaffold libraries, architectures like RFDiffusion can generate candidate protein backbones around predefined structural motifs or active-site geometries. Once the 3D backbone is formulated, graph neural networks like ProteinMPNN map the coordinates back into amino acid sequences optimized for heterologous expression. To complement de novo scaffold generation, the downstream optimization of complex multi-enzyme cascades requires fast, cooperative evolution. The recently developed MULTI-evolve framework addresses this by combining protein language models with epistatic modeling to predict synergistic multi-site mutations [60]. When paired with high-efficiency MULTI-assembly mutagenesis, this pipeline systematically engineers functional variants across multikilobase sequences, achieving up to 10-fold catalytic improvements in a single round of machine learning-guided evolution [57,58]. In plant metabolic engineering, de novo-designed enzymes face additional constraints, including chloroplast or ER folding environments, post-translational modification, codon usage, protein turnover, and possible toxicity of novel intermediates.

AI may provide new opportunities for physical control over the cellular space by regulating biomolecular condensates. The AI-driven FragFold method computationally designs protein fragments to control liquid–liquid phase separation (LLPS). FragFold-like strategies may provide a conceptual basis for future plant applications [61–63]. Technologically, FragFold employs a computational pipeline to screen isolated peptide fragments against their full-length parental proteins, identifying specific interfaces that can be targeted to modulate phase separation. This approach operates via two distinct biophysical mechanisms: monovalent inhibition, where designed fragments competitively block the weak, multivalent domain–domain interactions (such as the FERM-kinase domain interface) required for phase separation; and bridging enhancement, where fragments act as non-covalent cross-linkers that physically strengthen the macromolecular network. In experimental validations, FragFold-designed fragments achieved a 50% success rate (9/18 designs) in controlling condensate dynamics across structurally diverse proteins (including G3BP1 and TDP-43) both in vitro and in living cells [63]. Deploying these short peptide modulators provides an alternative to overexpressing bulky, endogenous intrinsically disordered regions, as it may reduce the burden associated with overexpressing large intrinsically disordered regions. The spatially targeted delivery of these engineered condensates into chloroplasts may provide a future strategy for membrane-free spatial organization of metabolic enzymes. In plant metabolic engineering, designed peptide fragments could, in principle, be used to organize artificial membrane-free condensates within the chloroplast stroma or cytosol,

physically sequestering volatile or toxic secondary metabolites without constructing physical lipid membrane barriers. Because current evidence for FragFold-like condensate control is largely derived from non-plant systems and includes preprint-level findings, its application to plant metabolic engineering should be presented as a future possibility rather than an established platform.

Plant cells already use biomolecular condensates in signalling, development, RNA regulation, and stress responses [64], but this biological precedent does not yet demonstrate that designed peptide fragments can increase plant metabolic flux. FragFold is cited as a non-plant proof-of-concept and preprint-level study. Plant application would require demonstration of controllable assembly and dissolution, enzyme partitioning, metabolite permeability, material-state measurements, organelle compatibility, and a measurable improvement in pathway output without toxicity or growth penalties.

A more detailed tool-level comparison of representative AI-assisted technologies, their applications, advantages, and limitations is provided in Supplementary Table S1. To distinguish established plant applications from emerging or speculative technologies, Table 2 summarizes the major technology layers, their current evidence status, and the principal barriers to implementation in plant spatial metabolic engineering.

Table 2. Major technology layers, metabolic applications, and current bottlenecks in AI-assisted spatial metabolic engineering in plants.

Technology Layer	Representative Tools or Strategies	Main Applications in Plant Synthetic Biology	Current Bottlenecks and Unresolved Issues	Evidence Status
Sequence-level AI design	Genomic foundation models, promoter design models, regulatory sequence predictors	Gene discovery, synthetic promoter design, regulatory element optimization, variant-effect prediction	Limited plant-specific training datasets, weak interpretability, insufficient validation across species, tissues, developmental stages, and environmental conditions	Early plant proof-of-concept for plant-specific genomic models; most generative applications remain prospective
Flux and pathway optimization	FBA, pFBA, machine learning-based flux prediction, multi-omics-constrained metabolic models	Identification of rate-limiting reactions, pathway balancing, precursor supply optimization, reduction in competing metabolic sinks	Incomplete plant metabolic models, poorly annotated specialized metabolism, uncertain transport reactions, and limited compartment-specific flux data	Plant-demonstrated for metabolic modelling and pathway optimization, although many AI models are trained or benchmarked mainly in microbial and mammalian systems
Subcellular targeting and compartmentalization	Transit peptide design, protein localization predictors, chloroplast targeting, ER-plastid pathway partitioning	Organelle-specific pathway construction, improved precursor utilization, reduction in intermediate toxicity, cofactor-aware pathway design	Predicted localization does not guarantee import efficiency, correct processing, protein folding, catalytic activity, or compatibility with local redox and cofactor environments	Plant-demonstrated for chloroplast, mitochondrial, ER, Golgi, vacuolar, and tissue-specific targeting; AI-guided de novo targeting-sequence design remains early-stage
Synthetic organelles and spatial scaffolds	Bacterial microcompartment s, encapsulation peptides, synthetic chloroplastic scaffolds,	Enzyme clustering, metabolite sequestration, pathway insulation, prevention of volatile	Cargo loading efficiency, shell permeability, developmental stability, energetic burden, organelle compatibility, and	Early plant proof-of-concept for carboxysome-related and shell-protein components; broadly functional BMC

	proteinaceous nanoreactors	or toxic intermediate loss	scalability remain poorly resolved	nanoreactors remain prospective
Secretory pathway and glycoengineering	Viral expression vectors, ER/Golgi targeting, multiplex CRISPR/Cas9-mediated glycan remodeling	Recombinant protein production, vaccine manufacturing, therapeutic protein expression, humanized glycosylation	Product quality consistency, protein-specific glycoform variation, downstream purification, regulatory approval, and large-scale manufacturing remain major challenges	Plant-demonstrated and commercially relevant for transient expression and glycoengineering
Cellular, tissue, and organ-level control	Synthetic receptors, chemically inducible dimerization systems, tissue-specific promoters, AI-assisted receptor design	Environmental sensing, inducible pathway activation, stomatal regulation, stress-responsive metabolic or signalling control	Field-level ligand persistence, unintended receptor activation, ecological safety, specificity, reversibility, and regulatory acceptance require further evaluation	Plant-demonstrated for tissue-specific expression and PYR1-MANDI; broad AI-guided receptor redesign remains prospective
Genome-level pathway stacking	CRISPR-based targeted integration, landing-pad prediction, artificial chromosomes, chromatin-aware transgene design	Stable multigene pathway assembly, reduced positional effects, long-term inheritance of complex synthetic traits	Plant artificial chromosome systems remain immature; epigenetic silencing, meiotic stability, transformation efficiency, and reliable landing-pad prediction remain limiting	Early plant evidence for engineered minichromosomes and targeted integration; routine artificial-chromosome deployment remains prospective
Generative protein and spatial validation platforms	AlphaFold 3, RFdiffusion, ProteinMPNN, condensate engineering, spatial omics, automated DBTL platforms	Enzyme redesign, catalytic optimization, membrane-free compartmentalization, spatial validation, iterative model refinement	Most designs require extensive experimental validation; plant-specific folding environments, subcellular physiology, data standardization, and biofoundry infrastructure remain insufficient	Predominantly non-plant proof-of-concept; plant metabolic applications remain largely prospective
Biomolecular condensate engineering	FragFold-like peptide design, phase-separation modulators, synthetic interaction fragments	Membrane-free enzyme organization, dynamic pathway clustering, and potential sequestration of selected enzymes or intermediates	Limited peer-reviewed validation, uncertain condensate stability in plant cells, possible interference with endogenous phase-separated systems, uncontrolled material properties, off-target interactions, and unknown physiological burden	Non-plant proof-of-concept, including preprint-level evidence; plant applications are prospective
Spatial-omics-guided validation	Spatial transcriptomics, single-cell or single-nucleus RNA sequencing, laser-capture microdissection, imaging mass	Identification of biosynthetic cell types, mapping of pathway expression, metabolite localization, transport-route inference, storage-site	High cost, limited throughput, tissue-sectioning artefacts, cell-wall constraints, metabolite instability, autofluorescence, incomplete cross-platform standardization, and	Plant-demonstrated for spatial mapping and cell-type resolution; integration with AI-guided metabolic engineering remains emerging

	spectrometry, spatial metabolomics	identification, and toxicity assessment	difficulties linking spatial abundance to actual metabolic flux	
Integrated AI- DBTL workflow	High-throughput screening, robotics, microfluidics, multi-omics feedback, spatial metabolomics	Closed-loop design- build-test-learn optimization for predictable plant biomanufacturing	Infrastructure-intensive; plant systems remain slower and less standardized than microbial chassis; stable in planta validation is still a major bottleneck	Early-stage in plants; more mature in microbial systems

8. Conclusions and Future Perspectives

Plant biomanufacturing increasingly depends on the integration of precise genetic perturbation, spatial pathway organization, and AI-assisted design. The maturation of the Push-Pull-Block paradigm, coupled with advancements in synthetic compartmentalization and viral vector-driven molecular farming, provides useful platforms for generating high-value biomolecules. Despite these advances, several bottlenecks remain unresolved. First, plant-specific AI models are limited by the scarcity of standardized, high-quality training datasets. Second, many computational predictions are validated only in transient expression systems rather than stable plant lines. Third, the spatial organization of metabolism is influenced by development, environment, organellar physiology, and tissue context, which are difficult to capture using bulk omics datasets. Finally, regulatory approval, field performance, biosafety, and public acceptance will determine whether AI-designed plant systems can move beyond laboratory proof-of-concept studies.

Achieving a more predictive design-build-test-learn cycle remains constrained by data availability. Current generative AI models and protein language models face a major limitation: many are trained predominantly on microbial and mammalian datasets. They may incompletely capture the unique biophysical constraints of plant microenvironments, such as the specific pH gradients or redox potentials within chloroplasts and vacuoles. Overcoming this requires the establishment of automated plant biofoundries. Therefore, the future success of AI-assisted plant engineering will depend not only on the generation of candidate genetic designs, but also on metabolite-level validation through flux analysis, targeted metabolomics, and spatially resolved metabolic profiling. High-throughput, standardized data collection from single-cell plant omics is essential to fine-tune existing language models. Once the loop between in silico generation and high-throughput in vivo plant validation is closed, plant synthetic biology may gradually move from empirical optimization toward more predictable and data-supported engineering.

AI should therefore be framed as a prioritization and hypothesis-generation tool, not as a replacement for experimental validation or breeder expertise. Plant metabolic phenotypes emerge from genetic background, cell identity, developmental stage, organellar physiology, transport, environment, and management. The most credible future workflows will report model uncertainty, distinguish interpolation from extrapolation, and require metabolite-level, spatial, stable-line, inheritance, and field validation before claiming improved pathway performance.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/metabo16080519/s1>, Table S1: Detailed tool-level comparison of AI-assisted technologies for plant spatial metabolic engineering.

Author Contributions: H.C. and J.Y. performed the literature search and data synthesis. H.C. wrote the original draft. J.Y. and M.D. contributed to literature collection and prepared the tables and figures. H.C. and M.D. critically revised the manuscript for intellectual content. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (31900251), Fundamental Research Program of Shanxi Province (202403021211031, 202203021212402), Shanxi Normal University Graduate Curriculum Ideological and Political Education Demonstration Project (2024 YJSKCSZSFK-01).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wagh, S.G.; Patil, A.M.; Patil, G.B.; Bhor, S.A.; Pawar, K.R.; Shinde, H. Programmable Plant Immunity: Synthetic Biology for Climate-Resilient Agriculture. *SynBio* **2026**, *4*, 1. <https://doi.org/10.3390/synbio4010001>.
2. Nagegowda, D.A.; Gupta, P. Plant terpenoids for growth, defense and environmental adaptation. *Annu. Rev. Plant Biol.* **2020**, *71*, 123–148.
3. Engler, C.; Kandzia, R.; Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **2008**, *3*, e3647. <https://doi.org/10.1371/journal.pone.0003647>.
4. Taylor, G.M.; Mordaka, P.M.; Heap, J.T. Start-Stop Assembly: A functionally scarless DNA assembly system optimized for metabolic engineering. *Nucleic Acids Res.* **2019**, *47*, e17. <https://doi.org/10.1093/nar/gky1182>.
5. Najafabadi, M.Y.; Jackson, S.A. Hybrid AI in synthetic biology: Next era in agriculture. *Trends Plant Sci.* **2025**, *31*, 555–565. <https://doi.org/10.1016/j.tplants.2025.08.011>.
6. Ballard, J.L.; Wang, Z.; Li, W.; Shen, L.; Long, Q. Deep learning-based approaches for multi-omics data integration and analysis. *BioData Min.* **2024**, *17*, 38. <https://doi.org/10.1186/s13040-024-00391-z>.
7. Brixi, G.; Durrant, M.G.; Ku, J.; Naghipourfar, M.; Poli, M.; Sun, G.; Brockman, G.; Chang, D.; Fanton, A.; Gonzalez, G.A.; et al. Genome modelling and design across all domains of life with Evo 2. *Nature* **2026**, *652*, 1349–1361. <https://doi.org/10.1038/s41586-026-10176-5>.
8. Li, C.; Zhang, Q.; Chen, H.; Lin, K.; Luo, C.; Yang, M.; Xu, W.; Yao, F.; Yan, J.; Yang, Q.; et al. PlantGFM: A Genomic Foundation Model for Discovery and Creation of Plant Genes. *Adv. Sci.* **2026**, *13*, e75772. <https://doi.org/10.1002/adv.75772>.
9. Nielsen, J.; Keasling, J.D. Engineering cellular metabolism. *Cell* **2016**, *164*, 1185–1197.
10. Gonçalves, D.M.; Henriques, R.; Costa, R.S. Predicting metabolic fluxes from omics data via machine learning: Moving from knowledge-driven towards data-driven approaches. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 4960–4973. <https://doi.org/10.1016/j.csbj.2023.10.002>.
11. Zampieri, G.; Angione, C.; Coggins, M.; Valle, G. A poly-omics machine-learning method to predict metabolite production in CHO cells. In Proceedings of the 2nd International Electronic Conference of Metabolomics, Online, 20 November–1 December 2017.
12. Lee, H.; Park, S.; Lee, S.B.; Song, J.; Kim, T.-H.; Kim, B.-G. Tailored biosynthesis of diosmin through reconstitution of the flavonoid pathway in *Nicotiana benthamiana*. *Front. Plant Sci.* **2024**, *15*, 1464877. <https://doi.org/10.3389/fpls.2024.1464877>.
13. Wu, Q.-W.; Wei, M.; Feng, L.-F.; Ding, L.; Wei, W.-K.; Yang, J.-F.; Lin, X.-J.; Liang, H.-L.; Zhan, R.-T.; Ma, D.-M. Rhamnosyltransferases involved in the biosynthesis of flavone rutinosides in *Chrysanthemum* species. *Plant Physiol.* **2022**, *190*, 2122–2136. <https://doi.org/10.1093/plphys/kiac371>.
14. Kong, Q.; Zheng, L. Implementing OR gates in Arabidopsis protoplasts for dynamic regulation. *Nat. Commun.* **2025**, *16*, 1120–1132.
15. Li, J.; Mutanda, I.; Wang, K.; Yang, L.; Wang, J.; Wang, Y. Chloroplastic metabolic engineering coupled with isoprenoid pool enhancement for committed taxanes biosynthesis in *Nicotiana benthamiana*. *Nat. Commun.* **2019**, *10*, 4850. <https://doi.org/10.1038/s41467-019-12879-y>.
16. Park, S.; Mani, V.; Ha, K.; Kim, J.A.; Lee, S. Plant synthetic biology: From knowledge to biomolecules. *Front. Plant Sci.* **2025**, *16*, 1562216. <https://doi.org/10.3389/fpls.2025.1562216>.
17. Xiang, X.; Yao, Q.; Deng, K.; Ge, Y.; Xiong, Q.; Lu, Y.; Hu, X. TargetGAN: A generative AI framework for the design of plant core promoters with targeted activity. *Plant Commun.* **2026**, *7*, 101851. <https://doi.org/10.1016/j.xplc.2026.101851>.
18. Hernandez-Garcia, C.M.; Finer, J.J. Identification and validation of promoters and cis-acting regulatory elements. *Plant Sci.* **2014**, *217–218*, 109–119. <https://doi.org/10.1016/j.plantsci.2013.12.007>.
19. Ding, J.; Wu, J.-H.; Liu, J.-F.; Yuan, B.-F.; Feng, Y.-Q. Improved methodology for assaying brassinosteroids in plant tissues using magnetic hydrophilic material for both extraction and derivatization. *Plant Methods* **2014**, *10*, 39. <https://doi.org/10.1186/1746-4811-10-39>.
20. Schad, M.; Mungur, R.; Fiehn, O.; Kehr, J. Metabolic profiling of laser microdissected vascular bundles of *Arabidopsis thaliana*. *Plant Methods* **2005**, *1*, 2. <https://doi.org/10.1186/1746-4811-1-2>.

21. Giacomello, S.; Salmén, F.; Terebienieć, B.K.; Vickovic, S.; Navarro, J.F.; Alexeyenko, A.; Reimegård, J.; McKee, L.S.; Mannapperuma, C.; Bulone, V.; et al. Spatially resolved transcriptome profiling in model plant species. *Nat. Plants* **2017**, *3*, 17061. <https://doi.org/10.1038/nplants.2017.61>.
22. Denyer, T.; Ma, X.; Klesen, S.; Scacchi, E.; Nieselt, K.; Timmermans, M.C. Spatiotemporal developmental trajectories in the Arabidopsis root revealed using high-throughput single-cell RNA sequencing. *Dev. Cell* **2019**, *48*, 840–852.e5. <https://doi.org/10.1016/j.devcel.2019.02.022>.
23. Thumhuri, V.; Armenteros, J.J.A.; Johansen, A.R.; Nielsen, H.; Winther, O. DeepLoc 2.0: Multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res.* **2022**, *50*, W228–W234. <https://doi.org/10.1093/nar/gkac278>.
24. Choi, H.I.; Lee, S.H.; Lee, I.H.; Lee, Y.J.; Yun, J.-H.; Choi, D.-Y.; Cho, D.-H.; Shin, B.-S.; Chun, J.; Lee, D.W.; et al. Chlamy_ChloroPred: A deep learning-based, highly accurate binary classifier for chloroplast protein prediction in the model microalga, *Chlamydomonas reinhardtii*, with potential cross-proteome versatility. *Front. Microbiol.* **2026**, *17*, 1744805. <https://doi.org/10.3389/fmicb.2026.1744805>.
25. Almagro Armenteros, J.J.; Salvatore, M.; Emanuelsson, O.; Winther, O.; von Heijne, G.; Elofsson, A.; Nielsen, H. Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* **2019**, *2*, e201900429. <https://doi.org/10.26508/lisa.201900429>.
26. Muzaffar, A.; Kiani, S.; Khan, M.A.U.; Rao, A.Q.; Ali, A.; Awan, M.F.; Iqbal, A.; Nasir, I.A.; Shahid, A.A.; Husnain, T. Chloroplast localization of Cry1Ac and Cry2A protein—an alternative way of insect control in cotton. *Biol. Res.* **2015**, *48*, 14. <https://doi.org/10.1186/s40659-015-0005-z>.
27. Edgar, S.; Li, F.-S.; Qiao, K.; Weng, J.-K.; Stephanopoulos, G. Engineering of taxadiene synthase for improved selectivity and yield of a key taxol biosynthetic intermediate. *ACS Synth. Biol.* **2017**, *6*, 201–205. <https://doi.org/10.1021/acssynbio.6b00206>.
28. Kerfeld, C.A.; Aussignargues, C.; Zarzycki, J.; Cai, F.; Sutter, M. Bacterial microcompartments. *Nat. Rev. Microbiol.* **2018**, *16*, 277–290.
29. Long, B.M.; Hee, W.Y.; Sharwood, R.E.; Rae, B.D.; Kaines, S.; Lim, Y.-L.; Nguyen, N.D.; Massey, B.; Bala, S.; von Caemmerer, S.; et al. Carboxysome encapsulation of the CO₂-fixing enzyme Rubisco in tobacco chloroplasts. *Nat. Commun.* **2018**, *9*, 3570. <https://doi.org/10.1038/s41467-018-06044-0>.
30. Nicolai, A.; Scotti, N.; D’aGostino, N.; Festa, G.; Sannino, L.; Aufiero, G.; Arimura, S.-I.; Cardi, T. Mitochondrial DNA editing in potato through mitoTALEN and mitoTALECD: Molecular characterization and stability of editing events. *Plant Methods* **2024**, *20*, 4. <https://doi.org/10.1186/s13007-023-01124-9>.
31. Yadav, N.S.; Raza, S.; Wang, Y.; Landa, J.F.; Hegg, E.L.; Hausinger, R.P.; Vermaas, J.V. Quantifying selective metabolite transport for the bacterial microcompartment from *Haliangium ochraceum* with molecular dynamics simulations. *Nanoscale* **2026**, *18*, 6092–6104. <https://doi.org/10.1039/d5nr03156g>.
32. Young, E.J.; Kirst, H.; Dwyer, M.E.; Vermaas, J.V.; Kerfeld, C.A. Quantitative measurement of molecular permeability to a synthetic bacterial microcompartment shell system. *ACS Synth. Biol.* **2025**, *14*, 1405–1413. <https://doi.org/10.1021/acssynbio.4c00290>.
33. Myers, R.J., Jr.; Tretter, Z.M.; Daffron, A.G.; Fritsch, E.X.; Santos, W.T.; Foster, M.L.; Klotz, M.; Stafford, K.M.; Kasch, C.; Taylor, T.J.; et al. Spatially Resolved Plant Metabolomics. *Metabolites* **2025**, *15*, 539. <https://doi.org/10.3390/metabo15080539>.
34. Fuentes, P.; Zhou, F.; Erban, A.; Karcher, D.; Kopka, J.; Bock, R. A new synthetic biology approach allows transfer of an entire metabolic pathway from a medicinal plant to a biomass crop. *eLife* **2016**, *5*, e13664. <https://doi.org/10.7554/elife.13664>.
35. Butelli, E.; Titta, L.; Giorgio, M.; Mock, H.-P.; Matros, A.; Peterek, S.; Schijlen, E.G.W.M.; Hall, R.D.; Bovy, A.G.; Luo, J.; et al. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat. Biotechnol.* **2008**, *26*, 1301–1308. <https://doi.org/10.1038/nbt.1506>.
36. Margolin, E.A.; Strasser, R.; Chapman, R.; Williamson, A.-L.; Rybicki, E.P.; Meyers, A.E. Engineering the plant secretory pathway for the production of next-generation pharmaceuticals. *Trends Biotechnol.* **2020**, *38*, 1034–1044. <https://doi.org/10.1016/j.tibtech.2020.03.004>.
37. Sainsbury, F.; Thuenemann, E.C.; Lomonossoff, G.P. Lomonossoff, pEAQ: Versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* **2009**, *7*, 682–693. <https://doi.org/10.1111/j.1467-7652.2009.00434.x>.
38. Mardanov, E.S.; Vasyagin, E.A.; Ravin, N.V. Virus-like Particles Produced in Plants: A Promising Platform for Recombinant Vaccine Development. *Plants* **2024**, *13*, 3564. <https://doi.org/10.3390/plants13243564>.
39. Vo, D.-K.; Trinh, K.T.L. Molecular Farming for Immunization: Current Advances and Future Prospects in Plant-Produced Vaccines. *Vaccines* **2025**, *13*, 191. <https://doi.org/10.3390/vaccines13020191>.

40. Montero-Morales, L.; Steinkellner, H. Advanced plant-based glycan engineering. *Front. Bioeng. Biotechnol.* **2018**, *6*, 81. <https://doi.org/10.3389/fbioe.2018.00081>.
41. Kaur, C.; Song, H.; Lee, M.; Kim, S.-Y.; Seo, D.-H.; Kang, H.; Sohn, E.-J.; Ran, Y.; Koo, O.; Lee, G.-J. Multiplex CRISPR/Cas9-mediated editing of seven glycosyltransferase homologs in *Nicotiana benthamiana* to produce stable, Cas9-free, glycoengineered plants. *Front. Plant Sci.* **2025**, *16*, 1701668. <https://doi.org/10.3389/fpls.2025.1701668>.
42. Wang, K.Y.; Guo, Y. Editorial: Plant molecular farming for biopharmaceutical production and beyond. *Front. Plant Sci.* **2026**, *16*, 1765245. <https://doi.org/10.3389/fpls.2025.1765245>.
43. Bombarely, A.; Rosli, H.G.; Vrebalov, J.; Moffett, P.; Mueller, L.A.; Martin, G.B. A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. *Mol. Plant-Microbe Interact.* **2012**, *25*, 1523–1530. <https://doi.org/10.1094/mpmi-06-12-0148-ta>.
44. Zeng, H.; Yang, A. Modelling overflow metabolism in *Escherichia coli* with flux balance analysis incorporating differential proteomic efficiencies of energy pathways. *BMC Syst. Biol.* **2019**, *13*, 3. <https://doi.org/10.1186/s12918-018-0677-4>.
45. Park, S.-Y.; Peterson, F.C.; Mosquna, A.; Yao, J.; Volkman, B.F.; Cutler, S.R. Agrochemical control of plant water use using engineered abscisic acid receptors. *Nature* **2015**, *520*, 545–548. <https://doi.org/10.1038/nature14123>.
46. Ruiz-Partida, R.; Rosario, S.M.; Lozano-Juste, J. An update on crop ABA receptors. *Plants* **2021**, *10*, 1087. <https://doi.org/10.3390/plants10061087>.
47. Tian, H.; Beltrán, J.; George, W.; Lenert-Mondou, C.; Seder, N.; Davis, Z.I.; Swift, S.D.; Girke, T.; Whitehead, T.A.; Wheeldon, I.; et al. Unusually Broad-spectrum small-molecule sensing using a single protein scaffold. *Proc. Natl. Acad. Sci. USA* **2025**, *122*, e2519924122. <https://doi.org/10.1073/pnas.2519924122>.
48. Snoeck, S.; Lee, H.K.; Schmid, M.W.; Bender, K.W.; Neeracher, M.J.; Fernández-Fernández, A.D.; Santiago, J.; Zipfel, C. Leveraging coevolutionary insights and AI-based structural modeling to unravel receptor–peptide ligand-binding mechanisms. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2400862121. <https://doi.org/10.1073/pnas.2400862121>.
49. Crossa, J.; Pérez-Rodríguez, P.; Cuevas, J.; Montesinos-López, O.; Jarquín, D.; de los Campos, G.; Burgueño, J.; González-Camacho, J.M.; Pérez-Elizalde, S.; Beyene, Y.; et al. Genomic selection in plant breeding: Methods, models, and perspectives. *Trends Plant Sci.* **2017**, *22*, 961–975. <https://doi.org/10.1016/j.tplants.2017.08.011>.
50. Birchler, J.A.; Swyers, N.C. Engineered minichromosomes in plants. *Exp. Cell Res.* **2020**, *388*, 111852.
51. Yu, W.; Han, F.; Gao, Z.; Vega, J.M.; Birchler, J.A. Construction and behavior of engineered minichromosomes in maize. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 8924–8929. <https://doi.org/10.1073/pnas.0700932104>.
52. Carlson, S.R.; Rudgers, G.W.; Zieler, H.; Mach, J.M.; Luo, S.; Grunden, E.; Krol, C.; Copenhaver, G.P.; Preuss, D. Meiotic transmission of an in vitro-assembled autonomous maize minichromosome. *PLoS Genet.* **2007**, *3*, e179. <https://doi.org/10.1371/journal.pgen.0030179>.
53. Zhou, J. Sequence-based modeling of three-dimensional genome architecture from kilobase to chromosome scale. *Nat. Genet.* **2022**, *54*, 725–734. <https://doi.org/10.1038/s41588-022-01065-4>.
54. Fudenberg, G.; Kelley, D.R.; Pollard, K.S. Predicting 3D genome folding from DNA sequence with Akita. *Nat. Methods* **2020**, *17*, 1111–1117. <https://doi.org/10.1038/s41592-020-0958-x>.
55. Krokidis, M.G.; Koumadorakis, D.E.; Lazaros, K.; Ivantsik, O.; Exarchos, T.P.; Vrahatis, A.G.; Kotsiantis, S.; Vlamos, P. AlphaFold3: An Overview of Applications and Performance Insights. *Int. J. Mol. Sci.* **2025**, *26*, 3671. <https://doi.org/10.3390/ijms26083671>.
56. Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A.J.; Bambrick, J.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493–500. <https://doi.org/10.1038/s41586-024-07487-w>.
57. Watson, J.L.; Juergens, D.; Bennett, N.R.; Trippe, B.L.; Yim, J.; Eisenach, H.E.; Ahern, W.; Borst, A.J.; Ragotte, R.J.; Milles, L.F.; et al. De novo design of protein structure and function with RFdiffusion. *Nature* **2023**, *620*, 1089–1100. <https://doi.org/10.1038/s41586-023-06415-8>.
58. Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R.J.; Milles, L.F.; Wicky, B.I.M.; Courbet, A.; de Haas, R.J.; Bethel, N.; et al. Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **2022**, *378*, 49–56. <https://doi.org/10.1126/science.add2187>.
59. Kuang, J.; Liu, N.; Wang, J.; Sun, C.; Ji, T.; Wu, Y. PDFBench: A Benchmark for De novo Protein Design from Function. *arXiv* **2025**, arXiv:2505.20346.

60. Tran, V.Q.; Nemeth, M.; Bartie, L.J.; Chandrasekaran, S.S.; Fanton, A.; Moon, H.C.; Hie, B.L.; Konermann, S.; Hsu, P.D. Rapid directed evolution guided by protein language models and epistatic interactions. *Science* **2026**, *392*, eaea1820. <https://doi.org/10.1126/science.aea1820>.
61. Lin, Y.; Li, S.; Zhang, Y. Convergence rate analysis of accelerated forward-backward algorithm with generalized Nesterov momentum scheme. *arXiv* **2021**, arXiv:2112.05873.
62. Qi, X.; Wang, D.; Shi, Z.; Liao, X.; Ma, H. Generative Artificial Intelligence for Function-Driven De Novo Enzyme Design. *Synth. Biol. Eng.* **2025**, *3*, 10015.
63. Savinov, A.; Sadasivan, J.; White, K.J.; Rubien, J.D.; Li, G.-W.; Case, L.B. AI-discovered protein fragments as generalizable regulators of biomolecular condensates. *bioRxiv* **2026**, bioRxiv:2026.05.08.723928.
64. Emenecker, R.J.; Holehouse, A.S.; Strader, L.C. Biological phase separation and biomolecular condensates in plants. *Annu. Rev. Plant Biol.* **2021**, *72*, 17–46. <https://doi.org/10.1146/annurev-arplant-081720-015238>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.