

that early planting tends to increase oil, late planting tends to increase protein, and improved nutrition at sowing can raise both yield and seed protein concentration, although the response depends on region and field productivity (Di Mauro et al., 2023).

7.3 Future research directions in dynamic regulation of seed composition

Future progress depends on resolving the developmental mechanisms that produce the protein-oil trade-off rather than only selecting around its final outcome. Current reviews argue that breeders need seed development-based studies using mutants, multi-omics, and metabolic flux analysis to explain how soybean seeds rebalance protein, oil, and sucrose during filling (Kumar et al., 2025). This need is reinforced by evidence that only a few soybean-specific oil and protein regulators have been functionally characterized so far, despite the likely existence of many crop-specific transporters, proteases, and transcription factors that are not predictable from Arabidopsis homologs alone.

The most promising future breeding framework combines multi-omics-guided target discovery with precise editing and broader use of untapped diversity. Spatial transcriptomics and metabolomics in wild soybean already show cell-type separation between protein- and lipid-associated metabolism and identify regulators such as *GsMAPK23-4* that alter amino acid and protein accumulation. At the translational end, gene editing now extends beyond natural alleles: DNA-free and AI-assisted editing pipelines are being developed for genotype-independent improvement, and AlphaFold-guided editing of *GmSWEET10a/b* has already increased oil or protein in multi-year, multi-site field trials without reducing yield (Wang et al., 2025; Kim et al., 2026).

8 Conclusions

Protein and oil accumulation in developing soybean seeds follow coordinated but often antagonistic temporal patterns that are shaped by shared metabolic resources and stage-specific regulation. Developmental studies show that the inverse association between mature seed protein and oil reflects cumulative changes across seed filling rather than a single static relationship, and late seed development is especially important because lipid content can decline during maturation while carbohydrates increase as maternal nutrient supply diminishes. Complementing this metabolic view, transcriptomic network analysis across contrasting high-oil and high-protein cultivars indicates that the major transcriptional divergence appears in later developmental stages, when lipid-centered and nitrogen-centered programs become antagonistically activated. These dynamic patterns also help explain why soybean seed composition is difficult to improve through endpoint-based selection alone. Comparative metabolomics of extreme phenotypes identified key intermediates such as glucose, citric acid, and α -ketoglutarate, and showed increased Calvin cycle, TCA cycle, and glycolytic activity in lines with strong protein or oil accumulation, supporting a model in which carbon rerouting underlies reserve deposition. Proteomic analysis further shows that differences in seed oil and protein content arise largely from the peripheral proteome, which fluctuates strongly across seed developmental stages rather than from a fixed constitutive protein set.

For soybean quality improvement, the central implication is that protein and oil should be treated as complex quantitative traits with partially shared but partly separable genetic control. Genome-wide studies have identified many loci affecting seed composition, including 87 chromosomal regions in one diverse panel and additional candidate genes involved in nitrogen fixation, amino acid biosynthesis, and fatty acid metabolism, providing a substantial marker base for breeding. More recent resequencing-based GWAS likewise detected 23 loci for protein and 29 for oil, including multiple new regions and nine candidate genes related to biosynthesis, transport, signaling, and development, reinforcing the feasibility of marker-assisted improvement. At the same time, breeding strategies must account for persistent trade-offs among composition, yield, and adaptation. Historical analyses indicate that selection for yield has reduced seed protein concentration by 0.06% per year while increasing residual low-energy fractions, and century-scale surveys in China found declining protein concentrations with largely stable oil content across decades of cultivar release. Conventional breeding therefore remains constrained by the negative correlation between protein and oil, but newer work suggests that simultaneous improvement is possible when breeders target independent loci, pleiotropic regulators, or specific genomic regions that affect one trait without penalizing the other.