Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Knowledge-driven approaches for engineering complex metabolic pathways in plants

Gemma Farré[1], Richard M Twyman[2], Paul Christou[3,4], Teresa Capell[3] and Changfu Zhu[3]

Plant metabolic pathways are complex and often feature multiple levels of regulation. Until recently, metabolic engineering in plants relied on the laborious testing of *ad hoc* modifications to achieve desirable changes in the metabolic profile. However, technological advances in data mining, modeling, multigene engineering and genome editing are now taking away much of the guesswork by allowing the impact of modifications to be predicted more accurately. In this review we discuss recent developments in knowledge-based metabolic engineering strategies, that is the gathering and mining of genomic, transcriptomic, proteomic and metabolomic data to generate models of metabolic pathways that help to define and refine optimal intervention strategies.

Addresses
[1] Metabolic Biology Department, John Innes Centre, Norwich, UK
[2] TRM Ltd, York, UK
[3] Department of Plant and Forestry Science, University of Lleida – Agrotecnio Center, Lleida, Spain
[4] Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

Corresponding author: Zhu, Changfu (zhu@pvcf.udl.cat)

Current Opinion in Biotechnology 2015, **32**:54–60

This review comes from a themed issue on **Plant biotechnology**

Edited by **Inge Broer** and **George N Skaracis**

http://dx.doi.org/10.1016/j.copbio.2014.11.004

Introduction

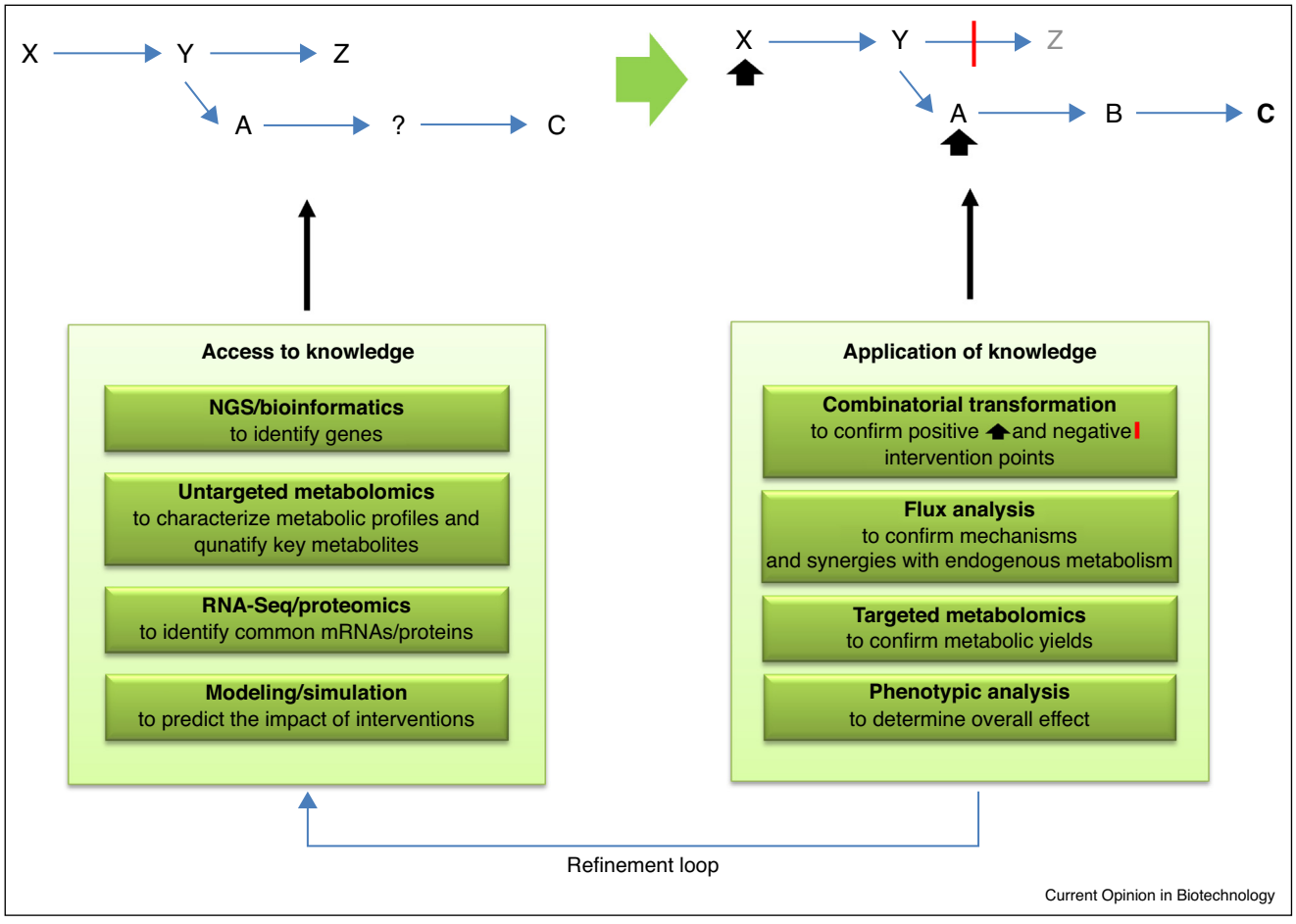
Metabolic engineering describes the targeted modification of endogenous metabolism to control the accumulation of one or more specific products [1]. One of the aims is to boost the levels of a metabolite normally accumulating in small amounts or to produce a compound that is not usually made at all, although in some cases metabolic engineering is used to reduce the level of an undesirable product. Metabolic pathways can be conceptually reduced to linear sequences of reactions but the reality is more complex, particularly in plants and other higher eukaryotes, because there are multiple layers of regulation as well as interactions with other pathways [2]. This makes the consequences of intervention difficult to predict. In this context, the traditional trial-and-error approach to metabolic engineering is being supplanted by a knowledge-driven strategy using combinations of genomics, transcriptomics, proteomics, metabolomics, metabolic libraries and modeling to identify priorities for targeted intervention before the design and introduction of optimized gene cassettes (Figure 1). The new wave of metabolic engineering therefore combines systems biology, synthetic biology, large-scale information gathering and predictive modeling to take as much of the guesswork away as possible, and takes advantage of multigene engineering strategies to simplify the process of generating metabolically reconfigured plants [3]. Precise genetic modulation can now be achieved using a range of genome editing tools, including most recently the CRISPR/Cas9 system which is described by Eudes *et al.* [4] in this issue.

Knowledge-based metabolic engineering — large-scale data gathering and mining

Early examples of metabolic engineering in plants typically featured single-gene interventions aiming to remove rate-liming steps, and more often than not they served only to identify additional limiting steps in the pathway [5]. This created a need to generate large numbers of transgenic plants on a trial-and-error basis to test different intervention strategies. The aims of metabolic engineering in plants are now more ambitious, focusing on complex, branching pathways with enzymes in different subcellular compartments, so there is an even greater need to understand the pathways and their regulation prior to intervention. An understanding of at least the key enzymes and intermediates in a pathway is recommended, and this process generally begins by isolating the corresponding genes.

Core metabolic pathways are well-characterized because the components are present in all plants and many versions of the corresponding genes have been isolated and tested. In contrast, complex secondary pathways may be present in a limited number of species and the genes and corresponding enzymes for each pathway step may not be known. This lack of data is being addressed through the development of large-scale sequencing, transcriptomics and proteomics projects targeting metabolically relevant plants and plant tissues. For example, genomic and transcriptomic analysis in the opium poppy revealed a 10-gene cluster encoding five types of enzymes required

Figure 1



Knowledge-based metabolic engineering. To determine the best intervention strategy in the theoretical metabolic pathway leading to compound C, traditional approaches would involve the trial-and-error introduction of individual or multiple genes encoding the corresponding enzymes, possibly also with targeted blocks in competitive branches, even in the absence of a complete set of enzymes. The knowledge-driven approach involves the use of genomics, transcriptomics, proteomics and/or metabolomics to generate large datasets that can be mined for relevant information (e.g. homologous sequences, coexpression profiles, co-induction in response to particular substrates, co-induction when particular transcription factors are present, protein co-localization) and the correlation of interventions with detailed metabolic profiles. Rather than *ad hoc* intervention, strategies can be developed based on rational models and the results of untargeted interventions to generate metabolic libraries, again using metabolic profiling to gain detailed information. These strategies can be used to refine and fine–tune the interventions.

to synthetize the anti-cancer alkaloid noscapine, and six of these genes were selectively silenced to validate their role in alkaloid biosynthesis [6•]. Similarly, transcriptomic and proteomic analysis of trichomes has helped to identify enzymes required for the synthesis of several important terpenoids and flavonols [7,8]. Co-regulated genes often encode relevant enzymes and those enzymes may be colocalized or found in the same supramolecular complexes. These complexes can be isolated under native conditions so that the components can be identified by mass spectrometry or interaction-based methods such as tandem affinity purification, as recently shown for the lignin pathway [9•].

RNA-seq has been used to catalog large numbers of tissue-specific and inducible sequences that can be mined to identify candidate enzymes, and the corresponding proteins can then be tested *in vitro* or in bacteria to determine their precise catalytic roles. Even without knowledge of the complete pathway, significant progress can be made by increasing the supply of precursors and analyzing the impact on gene expression, based on the principle that pathway intermediates often regulate genes involved in the same pathway. One of the most exciting recent demonstrations of this approach was the elucidation of the seco-iridoid pathway in the Madagascar periwinkle *Catharanthus roseus*, which yields important pharmacological molecules such as the anti-cancer drugs vinblastine and vincristine [10••]. These compounds are produced at minute levels in nature and are therefore extremely expensive to extract and purify, making them ideal targets for metabolic engineering. The authors