

Making drugs out of sunlight and 'thin air': an emerging synergy of synthetic biology and natural product chemistry

Michael J. Stephenson
and **Anne Osbourn**
(John Innes Centre, UK)

Nature has long served as a rich source of structurally diverse small organic molecules with medicinally relevant biological activities. Despite the historical success of these so-called natural products, the enthusiasm of big pharma to explore these compounds as leads in drug design has waxed and waned. A major contributor to this is their often inherent structural complexity. Such compounds are difficult (often impossible) to access synthetically, a hurdle that can stifle lead development and hinder sustainable large-scale production of promising leads for clinical evaluation. However, in recent years, an emerging synergy between synthetic biology and natural product chemistry offers the potential for a renaissance in our ability to access natural products for drug discovery and development. Advances in genome sequencing, bioinformatics and the maturing of heterologous expression platforms are increasing, enabling the study, and ultimately, the manipulation of plant biosynthetic pathways. The triterpenes are one of the most structurally diverse families of natural products and arguably one of the most underrepresented in the clinic. The plant kingdom is the richest source of triterpene diversity, with >20,000 triterpenes reported so far. Transient expression of genes for candidate enzymes and pathways in amenable plant species is emerging as a powerful and rapid means of investigating and harnessing the plant enzymes involved in generating this diversity. Such platforms also have the potential to serve as production systems in their own right, with the possibility of upscaling these discoveries into commercially useful products using the same overall basic procedure. Ultimately, the carbon source for generation of high-value compounds in plants is photosynthesis. Therefore, we could, with the help of plants, be producing new medicines out of sunlight and 'thin air' in green factories in the not too distant future.

Nature has long served as a rich source of structurally diverse small organic molecules. Many of these so-called natural products possess medicinally relevant biological activities. Indeed, exploration of this chemical space has yielded numerous clinically utilized drugs over the last century. This is particularly true in the treatment of infectious or malignant disease, where more than half of common pharmaceutical treatments are natural products or directly derived analogues thereof. The success of natural products as medicines is perhaps unsurprising, given the selection pressures imposed on living organisms in the natural world. This evolutionary 'arms race' has refined a wealth of structures to provide selective advantages, for example in providing protection against pests and pathogens and in inter-organismal signalling. These pharmacophores whose endogenous role can be directly exploited to treat human infections or conditions

of aberrant cell proliferation such as cancer. Indeed, the strictest definition of an antibiotic is a substance produced by one bacterium to kill another.

Despite the historical success of natural products, in recent decades, the enthusiasm of big pharma to explore these compounds as leads in drug design has waxed and waned. There are many potential reasons for this, but a major contributor is their often inherent structural complexity. This makes many natural products difficult/impossible to access synthetically. This is not only a problem for sustainable and practical large-scale production of promising candidates, but also stifles their refinement though analogue generation in the earlier lead development stages. Furthermore, as our understanding of the molecular biology underpinning many disease states has grown, the scope of potential drug targets has widened. The drive to more specifically

target these aberrant processes has often provided more fundamental focus on specific protein targets in the drug discovery process, rather than identifying a compound that has a positive effect on the system as a whole. When screening for activity against a known target protein, it makes sense to test libraries of compounds comprising structures that are easily accessible through practical established chemistries or seek fragments with complementary affinity that can be combined through high-yielding reactions. In doing so, one discovers not only a lead possessing the desired activity, but also a compound whose synthetic access is known.

It seemed for a while that this trend in declining interest might be set to continue. Increasing total synthesis of natural products seemed to serve more as an intellectual challenge for chemists in academia, or as a proving ground for new reactions, than a truly practical tool for drug development. While natural product chemists continued to discover more complex bioactive compounds from nature, medicinal chemistry was increasingly turning to high-throughput screening of synthetic libraries for inspiration and more reliable reactions for engineering diversity. However, the tide could be turning once more for natural products. In recent years, there has been a new resurgence of interest. This renaissance has arguably been driven, at least in part, by advances in genome sequencing, bioinformatics and the maturing of heterologous expression platforms. Natural product chemistry has provided, and continues to provide, a wealth of interesting structures. However, increasing numbers of genomes and transcriptomes of the species that produce diverse chemistries are now also becoming available. Combined with the development of powerful and rapid methods for heterologous expression, this has allowed exploration of the biosynthesis of these

complex compounds, and indeed, via the application of synthetic biology, the manipulation of these pathways to engineer novel analogues. Therefore, a new interface between synthetic biology and natural product chemistry is emerging as a potentially powerful tool for the medicinal chemists of the future.

The triterpenes are one of the most structurally diverse families of natural products. Despite this tremendous diversity, all triterpene alcohols are derived from the same linear precursor – a compound known as 2,3-oxidosqualene. Enzymes called oxidosqualene cyclases (OSCs) control the differential cyclization of this substrate into a wealth of different basal scaffolds. These structures are then further modified by tailoring enzymes, such as, but not limited to, cytochromes P450 (CYP450s) (Figure 1). The result is a rich plethora of structural variety, but one that shares a common biogenic origin. Given the breadth of this chemical space, it is not surprising that a wide range of biological activities have been reported. However, this potential has not yet translated into a wealth of clinically utilized medicines. Indeed, it is arguable that triterpene-derived drugs are underrepresented in the clinic, with only a few examples in clinical use or late stage development, such as the vaccine adjuvant QS-21 and the semi-synthetic triterpene bardoxolone methyl. Both of these molecules are based on the same OSC product (β -amyrin). QS-21 is used for its immunostimulatory properties in a vaccine formulation to protect against shingles, a disease caused by the virus herpes zoster. Known by the proprietary name ‘Shingrix’, this vaccine is manufactured by GSK. QS-21 is currently isolated from the bark of the Chilean tree, *Quillaja saponaria*. The American company Reata Pharmaceuticals is currently seeking regulatory approval for the use of bardoxolone methyl for the

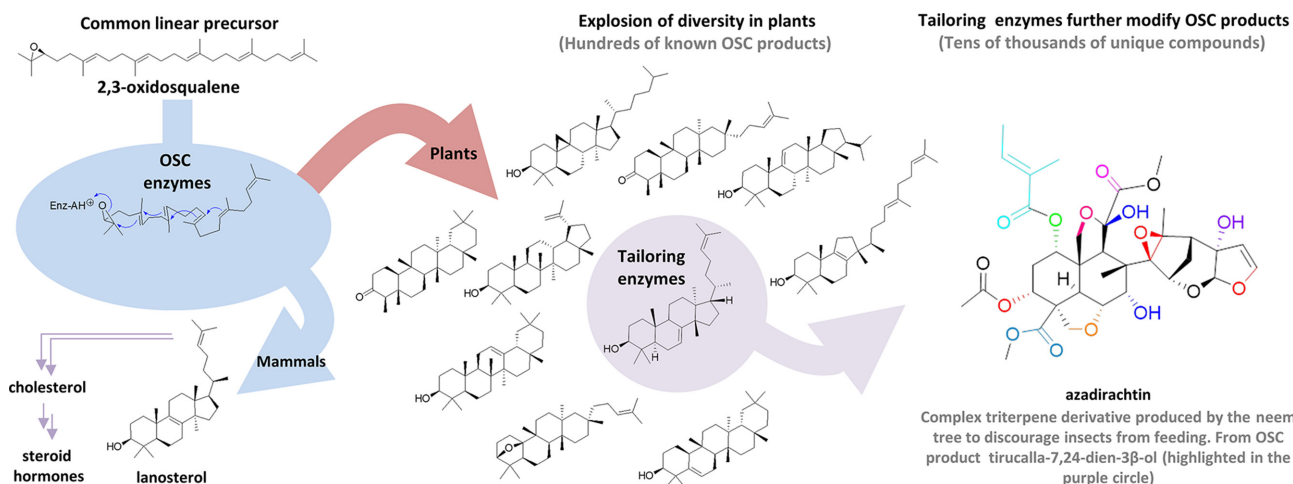


Figure 1. The structural diversity of all triterpenes shares a common origin. Plants are a rich source of this diversity. OSCs and tailoring enzymes convert 2,3-oxidosqualene into a wealth of unique compounds. Many tens of thousands of triterpenes from nature are known.

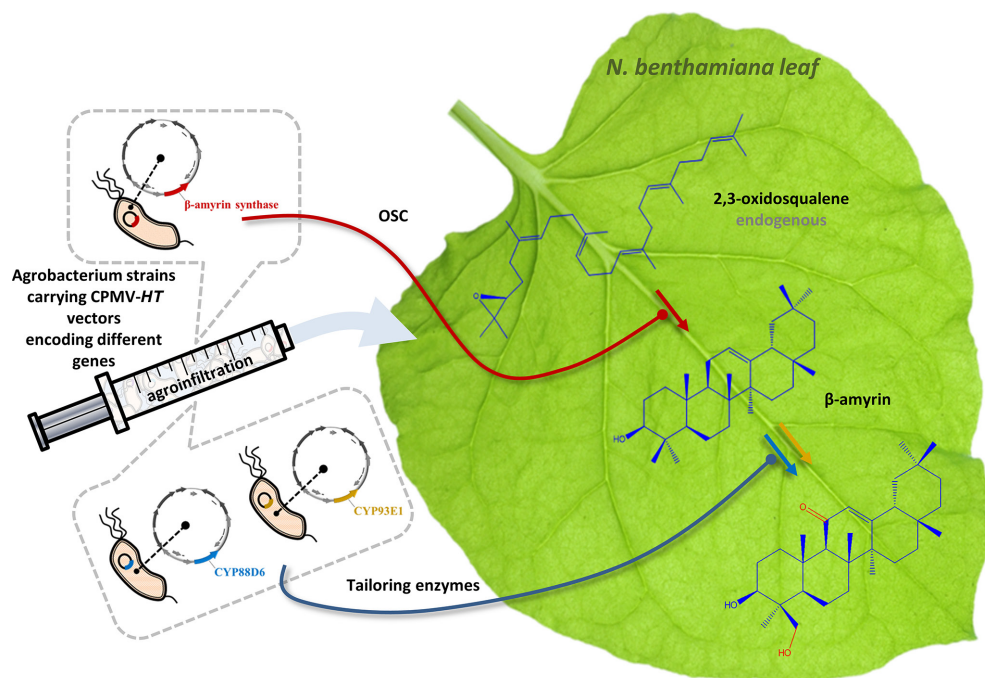


Figure 2. Schematic representation of the agroinfiltration process to afford transient expression of biosynthetic genes encoded in the CPMV-*HT*⁺ vector in *Nicotiana benthamiana*. The translated proteins can divert endogenous supplies of 2,3-oxidosqualene towards production of triterpene products not normally made by the plant.

treatment of chronic kidney disease in the USA. The anti-inflammatory properties of this compound have been associated with significant improvements in kidney function in patients suffering from Alport syndrome during phase III clinical trials.

The rigid, but shapely nature of many polycyclic triterpene scaffolds would seem an excellent basal structure for exploring modifications to increase binding affinity to target proteins. Functional groups added at different positions would be held in consistent relative space, restricting change from high-affinity conformations to lower ones. Indeed, rigidification is a common strategy employed in the optimization of lead compounds during drug development. However, unlike nature which has evolved enzymes capable of selectively adorning these bare scaffolds with different chemical features, synthetic chemistry is limited in the modifications that can be made directly. Furthermore, analogue generation via a series of complex total synthesis is an unappealing prospect when the goal is rapid probing of structure–activity relationships.

Triterpenes are widespread in nature, but the richest source of triterpenoid diversity is the plant kingdom. While animals typically possess just one OSC which produces an important intermediate in the biosynthesis of essential metabolites such as cholesterol and the steroid hormones, plant genomes often encode more than 10. This reflects the extent to which plants have exploited triterpene chemistry to fulfil more diverse and

specialized roles, including, e.g., producing compounds to protect against insect attack (Figure 1). A major focus in our laboratory is the application of transient plant expression to explore and, increasingly, to manipulate, biosynthesis of this chemical space.

Transient plant expression is a rapid strategy for determining the functions of heterologous genes. The gene(s) of interest are introduced into the plant in expression constructs. The host's cellular machinery is hijacked and the heterologous gene(s) are expressed, resulting in the production of the encoded protein by the host cells, much like a virus. Indeed, many of the vectors used in transient expression are based on viral genomes. One highly successful example is the CPMV-*HT*⁺ system. This vector exploits elements of a natural plant leaf pathogen called cowpea mosaic virus. The gene of interest is flanked by untranslated regions of code derived from this virus. Modifications in the 5' sequences result in high levels of protein production when the vector is introduced into leaf cells. Indeed, *HT* stands for *HyperTranslatable*.

Unlike the virus, the CPMV-*HT*⁺ vector does not replicate or encode proteins that allow new copies to gain entry to neighbouring cells. Instead, the vector is introduced into leaf cells via agrobacteria following a process known as agroinfiltration. In this process, a suspension of the bacteria carrying the vector is forced through small holes in the underside of the leaf known as stomata. This is typically achieved via the application

of pressure using a needle-less syringe, with the mouth forming a seal on the leaf's surface. The result is to displace air and fill the intercellular space with the suspension. Once inside the leaf, the agrobacteria fulfil their natural function and transfer genetic information into the interior of the leaf cells, including the T-DNA from the introduced CPMV-HT⁺ vector (Figure 2).

The plant most commonly used for transient expression is *Nicotiana benthamiana*, a wild relative of the tobacco plant. *N. benthamiana* is fast growing, particularly amenable to the technique and has a leaf morphology that suits infiltration. Typically, 5-week-old plants are used for transient expression experiments. The process is quick with high levels of the desired protein accumulating in the infiltrated leaf tissue in just a matter of days. Using a plant as the heterologous expression host has a number of advantages over microbial systems such as yeast and bacteria when studying plant proteins. The cell architecture intrinsically supports appropriate mRNA and protein processing, and proper compartmentalization. Furthermore, for studies of biosynthetic enzymes, *N. benthamiana* intrinsically has many of the necessary coenzymes, reductases (for CYP450s) and metabolic precursors needed.

The fact that all triterpenoid diversity stems from a single linear precursor makes the triterpenes particularly attractive to study. Through the transient expression of the relevant biosynthetic enzymes, this extensive chemical space can be explored in *N. benthamiana* by diverting the endogenous supply of 2,3-oxidosqualene towards the production of triterpene-derived compounds from other species (Figure 2). Furthermore, the carbon source is photosynthesis. Thus plants can simply be grown in good-quality compost, requiring only water, carbon dioxide (from the air) and sunlight as inputs.

A great advantage of transient expression is the ease with which the activity of combinations of enzymes can be investigated. There is no need to build large multigene vectors. Co-expression can be achieved simply by the co-infiltration of different strains of agrobacteria containing different expression constructs, mixed in a single suspension at the appropriate concentrations (Figure 2). This allows for the quick and convenient screening of candidate enzymes which are suspected to function together. Exploitation of this advantage has enabled the rapid piecing together of biosynthetic pathways. Here, the common biosynthetic origin of the triterpenes provides another advantage. Since all triterpene pathways are initiated by OSC-mediated generation of a particular scaffold, these enzymes provide convenient bioinformatic handles to probe downstream steps. OSC genes can be used as a bait to search for genes potentially encoding downstream tailoring enzymes by using co-expression analysis to identify genes with similar expression patterns. At a genome level, this information can also be combined with physical proximity to further prioritize likely genes involved in the pathway of interest. Increasing numbers of examples of so-called biosynthetic gene clusters are being discovered in plants. These are chromosomal regions containing high densities of co-localized and co-expressed genes that together comprise the biosynthetic pathways for different types of specialized metabolites (Figure 3).

As these strategies mature, the synthetic manipulation of these pathways is becoming a real possibility for engineering novel structures. The ease of co-expression allows unnatural combinations of enzymes to be screened for complementary activity, i.e., enzymes which function on the same scaffold but in different biosynthetic pathways. We have used such

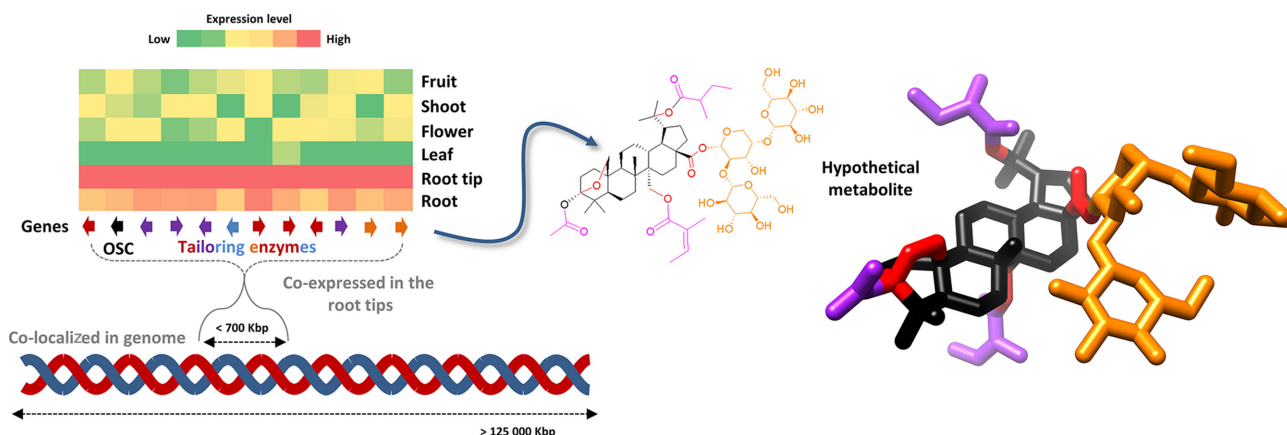


Figure 3. Schematic representation of a hypothetical gene cluster. OSCs can serve as a bait to discover tailoring enzymes when investigating triterpene biosynthetic pathways. They are present in low numbers compared to the different families of tailoring enzymes.



Vacuum infiltration allows the process to be scaled-up to produced large quantities of high-value product

Figure 4. Transient expression via agroinfiltration can be scaled up using vacuum infiltration, allowing the production of high-value products. The image shows over £20,000 worth of pure β -amyryn produced in this way. The process required little optimization and was achievable on a laboratory scale. The technique is already used at an industrial scale to produce protein products for vaccines. Intermediate scale-up facilities have also been established to aid translation (e.g. Leaf Expression Systems, Norwich, UK).

a combinatorial biosynthetic approach to generate new-to-nature variants of β -amyryn. This has allowed us to begin to probe the structure–activity relationships that give rise to the reported anti-inflammatory properties of compounds derived from this triterpene.

The inventor of the CPMV-*HT*⁺ vector and pioneer in the use of transient expression for the production of proteins such as virus-like particles for vaccines, Professor George Lomonosoff, is often quoted as saying, “you get your failures quickly with transient expression”. This is true and a fact that allows freedom and confidence to explore more speculative and ambitious ideas. However, the system is not just a platform for discovery. It is also a potential production system in its own right, with the possibility of translating these discoveries to commercially useful products using the same basic procedure. The process can be linearly and reliably scaled simply by increasing the number of plants used in the experiment. Moving from hand infiltration of individual leaves to vacuum infiltration of whole plants allows large-scale production (Figure 4). Indeed, this is already performed on an industrial scale for the production of proteins. The Canadian company Medicago is in the

process of building a new facility capable of producing 50m doses of influenza vaccine annually using transient plant expression technology. Even on a laboratory scale, we have used vacuum infiltration to produce gram-scale quantities of triterpene products. This may not sound significant to those unfamiliar with such technologies, but this is many orders of magnitude greater than the typical quantities produced in such experiments (or indeed from typical total synthesis of complex natural products) and requires no lengthy optimization. To offer some perspective, the isolated product had a commercial value of over £20,000 (Figure 4). Indeed, even low gram-scale production opens the door to the practical possibility of more extensive clinical studies such as early toxicology screening and pharmacokinetic investigations in animal models.

The future is bright once more for natural products in drug discovery. There is still a vast array of untapped chemical diversity waiting to be exploited from the natural world. As the synergy between synthetic biology and natural product chemistry grows, we could, with the help of plants, really be producing new medicines out of sunlight and ‘thin air’ in the not too distant future. ■

Further Reading

- Feng, L., Yongli, W., Dapeng, L. et al. (2019) Are we seeing a resurgence in the use of natural products for new drug discovery? *Expert Opin. Drug Discov.* **14**, 417–420 10.1080/17460441.2019.1582639
- Amirkia, V. and Heinrich, M. (2015) Natural products and drug discovery: a survey of stakeholders in industry and academia. *Front. Pharmacol.* **6**, 237 10.3389/fphar.2015.00237
- Stephenson M. J., Field R. A., and Osbourn A. (2019) The protosteryl and dammarenyl cation dichotomy in polycyclic triterpene biosynthesis revisited: has this 'rule' finally been broken? *Nat. Prod. Rep.* **36**, 1044–1052 10.1039/c8np00096d
- Nützmänn, H., Huang, A. and Osbourn, A. (2016) Plant metabolic clusters – from genetics to genomics. *New Phytol.* **211**, 771–789
- Orme A., Louveau T., Stephenson M. J. et al. (2019) A noncanonical vacuolar sugar transferase required for biosynthesis of antimicrobial defense compounds in oat. *Proc. Natl. Acad. Sci. USA* **116** 27105–27114 10.1073/pnas.1914652116
- Hodgson H., De La Pena R., Stephenson M. J. et al. (2019) Identification of key enzymes responsible for protolimonoid biosynthesis in plants: Opening the door to azadirachtin production. *Proc. Natl. Acad. Sci. USA* **116** 17096–17104 10.1073/pnas.1906083116
- Lau, W., and Sattely E.S. (2015) Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science* **349**, 1224–1228 10.1126/science.aac7202
- Dong, L., Jongedijk, E., Bouwmeester, H. et al. (2016), Monoterpene biosynthesis potential of plant subcellular compartments. *New Phytol.* **209**, 679–690 10.1111/nph.13629
- Stephenson M. J., Reed J., Patron N. J. et al. (2019) Engineering Tobacco for Plant Natural Product Production in *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering* 10.1016/B978-0-12-409547-2.14724-9
- Sainsbury, F., Thuenemann, E.C. and Lomonosoff G.P. (2009) pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* **7**, 682–693 10.1111/j.1467-7652.2009.00434.x
- Thuenemann, E.C., Meyers, A.E, Verwey, J. et al. (2013). A method for rapid production of heteromultimeric protein complexes in plants: assembly of protective bluetongue virus-like particles. *Plant Biotechnol. J.* **11**, 839–846 10.1111/pbi.12076
- Stephenson M. J., Reed J., Brouwer B. et al. (2018) Transient expression in *Nicotiana Benthamiana* leaves for triterpene production at a preparative scale. *J. Vis. Exp.* **138**, e58169 10.3791/58169
- Reed J., Stephenson M. J., Miettinen K. et al. (2017) A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metab. Eng* **42**, 185–193 10.1016/j.ymben.2017.06.012
- Video of large-scale vacuum infiltration in *Medicago*'s commercial production facility <https://youtu.be/4St743KvBeo>



Dr Michael J. Stephenson received his MPharm degree in 2010 and qualified as a registered pharmacist the following year. He then returned to academia to pursue a PhD in medical chemistry with Professor Mark Searcey's group at the University of East Anglia. This was awarded in 2015 for the development of novel methodology for the rapid synthesis of analogues of the potent antitumour-antibiotic duocarmycin. Michael then joined the Osbourn group at the John Innes Centre, where his research focusses on triterpene biosynthetic engineering and on utilization of transient plant expression for the preparative production of high-value triterpenes. Michael also holds an honorary lectureship in medicinal chemistry in the School of Chemistry at the University of East Anglia.



Anne Osbourn FRS OBE is a group leader at the John Innes Centre, an honorary professor at the University of East Anglia and Director of the Norwich Research Park Industrial Biotechnology Alliance. Her research focusses on plant natural products—biosynthesis, function and mechanisms of metabolic diversification. An important advance from the Osbourn laboratory has been the discovery that genes for specialized metabolic pathways are organized in 'operon-like' clusters in plant genomes, a finding that has opened up new opportunities for elucidation of new pathways and chemistries through genome mining. Anne also developed Science, Art and Writing (SAW), a cross-curricular science education programme (www.sawtrust.org). Email: anne.osbourn@jic.ac.uk