CellPress
OPEN ACCESS

Trends in
Pharmacological Sciences

Forum

Advances in triterpene drug discovery

Zoë R. Goddard[1], Mark Searcey[2], and Anne Osbourn[1,*]

Triterpenes are structurally complex natural products with promising therapeutic properties. Recalcitrance to chemical synthesis has hindered their use in drug development. Recent advances now make it possible to access and harness triterpene structural diversity using engineering biology approaches, enabling the discovery and optimisation of a new generation of drug leads.

The triterpenes: untapped potential for drug discovery

Although the number of chemically synthesizable compounds has increased over the past 30 years, this has not translated into an increase in drug approvals, suggesting that the quality of the chemical structures rather than the quantity of new molecules is the limiting factor [1]. Natural products (chemical compounds synthesized by living organisms) have a far greater degree of chemical diversity in both molecular structures and physicochemical properties than the chemical space that can be accessed using synthetic chemistry. Furthermore, they have been honed by evolution to interact with biological targets and processes, making them a rich source of potential drug leads [2].

The triterpenes are a large structurally diverse group of natural products with considerable pharmaceutical potential. They are particularly abundant in plants (>20 000 distinct structures reported to date) but are also found in fungi and some marine animals. Four triterpenes have been approved for clinical use in humans to date. However, broader investigations of the triterpenes as potential new drug leads have been hampered by the recalcitrance of these structurally complex compounds to chemical synthesis and the dual challenges of purifying these compounds from natural sources and lead optimisation. Advances in plant genomics and improved understanding of triterpene biosynthesis now make it possible to access previously inaccessible triterpene chemical space using engineering biology approaches. This opens up the possibility of systematically evaluating the bioactivities, modes of action, and structure–activity relationships (SARs) of a diverse array of new triterpene chemistries for the development of next-generation drugs. In this forum article, we provide background on triterpene structure, biosynthesis, and current pharmacological status. We highlight recent advances that are now enabling the previously inaccessible triterpene chemical space to be accessed and harnessed using engineering approaches. We further review new developments and future potential for understanding the impact of triterpenes on cellular functions, identifying their protein targets and rational design of new triterpene-based therapeutics.

Triterpene structure, biosynthesis, and current pharmacological status

Triterpenes are complex hydrocarbons derived from the isoprenoid pathway by cyclization of 2,3-oxidosqualene. They consist of 30 carbon atoms cyclised to form one to five ring skeletons, the most common of which are pentacyclic (e.g., β-amyrin; Figure 1A). As more cycles are added, the rigidity of the triterpene increases. Cyclisation also leads to chiral centres, which, along with the conformational rigidity of the scaffold, result in complex, preorganised molecules. Subsequent scaffold decoration (e.g., the addition of hydroxyls, carbonyls, sugars, and acyl chains) confers even greater structural diversity and is responsible for the wide range of biological activities displayed by these compounds.

Despite the considerable pharmaceutical interest in this large diverse family of natural products, examples of progression of triterpenes to clinical use have so far been limited. By 2020 two triterpenes had been approved for therapeutic use in humans, the vaccine adjuvant QS-21 and carbenoxolone (for peptic and mouth ulcers). More recently omaveloxolone was approved for treatment of Friedreich's ataxia [3], and ibrexafungerp was approved for vulvovaginal candidiasis [4] (Figure 1B). QS-21 is a highly complex triterpene glycoside sourced by extraction from the bark of the Chilean soapbark tree *Quillaja saponaria*, while carbenoxolone and omaveloxolone are semisynthetic triterpenes generated from oleanolic acid. All three of these compounds are derived from β-amyrin, the most common pentacyclic triterpene scaffold (Figure 1A). Ibrexafungerp is generated by chemical derivatisation of the naturally occurring fungal triterpene enfumafungin [5]. Fifty clinical trials for triterpene-based therapies are currently registered (Figure 1C), further highlighting the potential of this family of natural products as drug leads.

Opening up access to new chemical space for triterpene drug development

Increased understanding of the genes and enzymes of triterpene biosynthesis from species across the plant kingdom now makes it possible to generate triterpene structural diversity in heterologous systems. Microbial systems such as yeast are traditionally used for heterologous expression and are favoured by industry. However microbes lack many of the features needed for efficient and effective expression of plant natural product biosynthetic enzymes and pathways and require considerable engineering for pathways to be expressed successfully. Transient plant expression offers a solution to this problem (Figure 2A). In this approach, leaves of a wild relative of

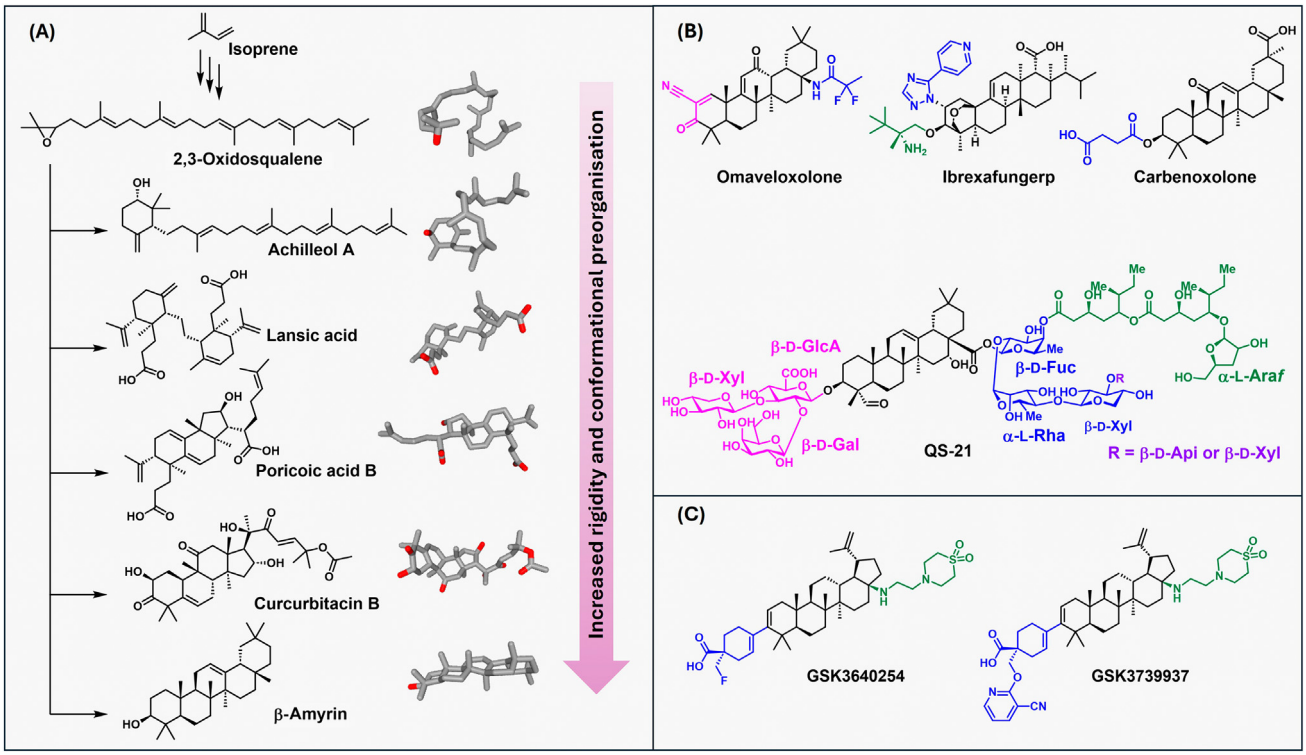

Figure 1. Structural diversity and pharmaceutical potential of triterpenes. (A) Triterpenes are synthesised from isoprene units via the mevalonate pathway. The precursor 2,3-oxidosqualene is cyclised into triterpenes with one to five cycles. Increasing the number of cycles restricts conformational freedom, resulting in highly preorganised triterpene skeletons. Representative triterpenes are achilleol A (monocyclic), lansic acid (bicyclic), poricoic acid B (tricyclic), cucurbitacin B (tetracyclic), and β-amyrin (pentacyclic). 3D structures are from PubChem. (B) Clinically approved triterpenes include the vaccine adjuvant QS-21 (the only natural triterpene clinically approved so far) and the semisynthetic triterpenes omaveloxolone (Friedreich's ataxia), ibrexafungerp (vulvovaginal candidiasis), and carbenoxolone (peptic and mouth ulcers). (C) GSK3640254 and GSK3739937 are semisynthetic triterpene HIV maturation inhibitors. GSK3739937 is currently in Phase 1 clinical trials[i].

tobacco (*Nicotiana benthamiana*) are infiltrated with agrobacteria containing genetic information for the expression of biosynthetic pathway enzymes, resulting in accumulation of the desired molecule. The process is very rapid, taking just 3–5 days from agroinfiltration to analysis of the metabolite content of the leaf extract.

Triterpenes are inherently complex molecules with multiple sites of oxidation, often with a number of sugar residues. SAR studies of triterpenes by traditional synthetic chemistry is therefore challenging and involves multistep syntheses [6]. Increased understanding of biosynthetic pathways and advances in large-scale transient expression have paved the way for *in planta* production of bespoke triterpenes through transient expression of a 'toolkit' of enzymes known to undertake characterised alterations at specific positions on the triterpene skeleton [7]. This combinatorial approach enables the generation of designer, highly substituted triterpenes on a milligram to gram scale and expands the chemical space available for SAR studies of novel triterpenes, thereby revolutionising the drug discovery pipeline. The 20-gene *Q. saponaria* QS-21 pathway has recently been reconstituted in *N. benthamiana* leaves by transient plant expression [8]. QS-21 has also recently been produced in yeast, but this required extensive retuning of the host's natural pathway fluxes and coexpression of 38 heterologous enzymes [9]. Production of QS-21 at scale will require considerable further optimisation. Nevertheless, the fact that this exceptionally complex triterpene can be made in heterologous systems holds great promise for the development of new vaccine adjuvants.

Understanding modes of action fuels triterpene drug development

Understanding modes of action will be critical for drug development. For many triterpenes, even some of those with clinical approval, the mechanism of action remains elusive. Methods to investigate triterpene bioactivity have generally relied on traditional techniques such as western blotting, ELISA, real-time PCR, and immunofluorescence. While these techniques are important for validating target engagement and downstream effects, they are

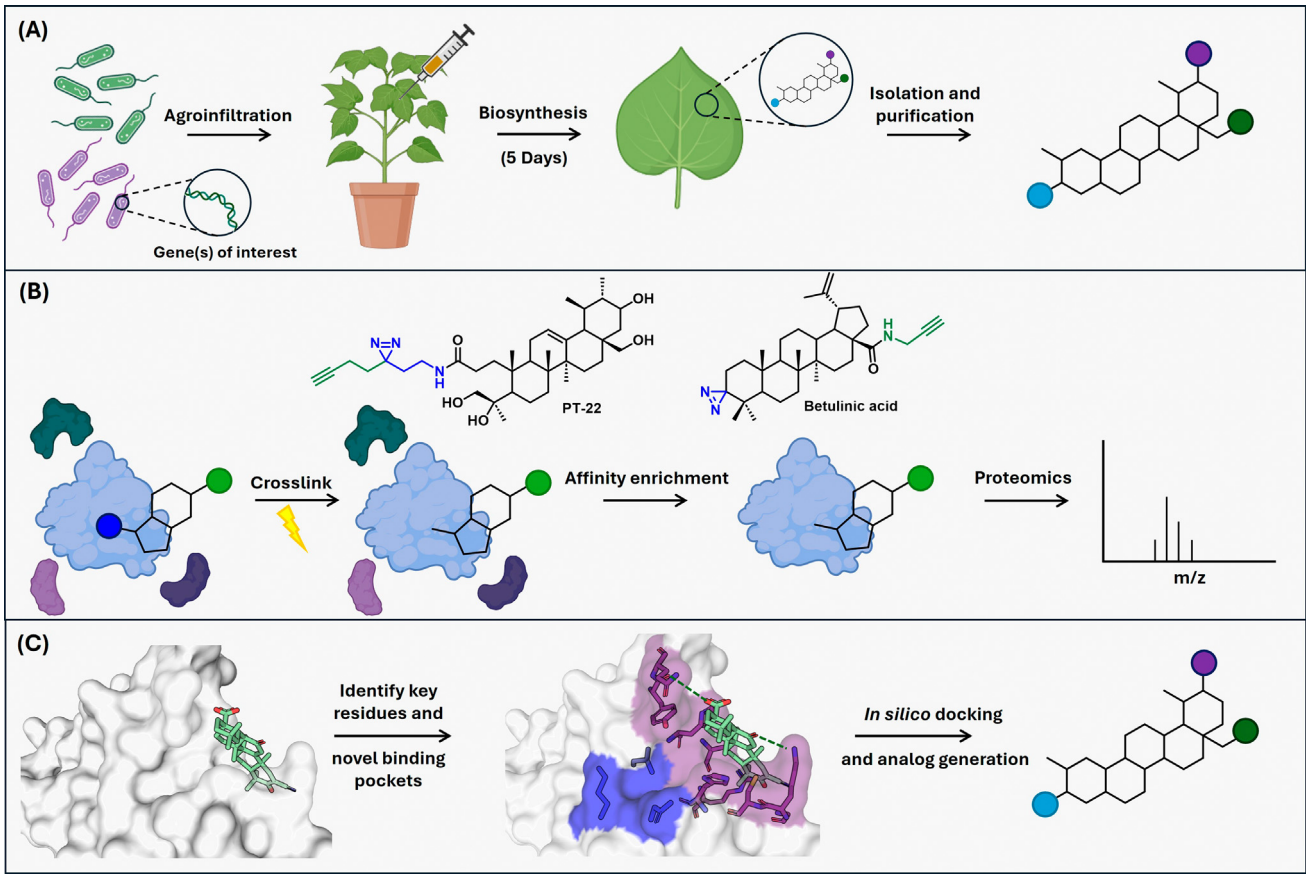

Figure 2. Strategies for generating triterpene diversity and investigating function. (A) Engineering triterpenes in *Nicotiana benthamiana*. *N. benthamiana* leaves are infiltrated with agrobacteria containing plasmids for the expression of the enzymes needed to synthesise the desired triterpene. (B) Chemoproteomics pipeline. Probes containing photoreactive and enrichment tags are incubated with cells. Irradiation leads to a covalent bond between the protein and the probe. Cells are lysed and subjected to affinity enrichment, and tagged proteins are collected by affinity chromatography. The tagged proteins are eluted, digested, and analysed by proteomics. The structures of the photoaffinity probes PT-22 [13] and betulinic acid [12] are shown with the photoreactive diazirine (blue) and the alkyne for affinity enrichment (green) highlighted. (C) Structure-based drug design workflow. The structure of a molecule in its binding site is analysed to improve its binding affinity. Key residues (purple) and novel binding pockets (blue) are identified. *In silico* docking is used to screen for higher-affinity binders, and the best compounds are synthesised and analysed for biological activity. Representative crystal structure Protein Data Bank (PDB): 4CXT. Images were created using BioRender.com.

focused on potential target proteins of interest, not on *de novo* discovery of unknown targets.

Notably, three of the four clinically approved triterpenes are semisynthetic (Figure 1B). Semisynthesis allows for the development of analogues with improved pharmacokinetics. This is approached through SAR studies, which depend on some understanding of the biological target. For instance, the development of GSK3640254 and GSK3739937 (Figure 1C), HIV maturation inhibitors, relied on the identification of a critical carboxylic acid [10]. The growing number of enzymes available for regio- and stereo-specific modification of triterpene scaffolds now offers opportunities to make new cohorts of analogues that cannot be attained using chemical synthesis, further empowering SAR studies.

Chemical probes can be used to elucidate the mechanism of action of molecules in the complex environment of a cell. While not a new concept, the application of such approaches for the study of biologically active triterpenes is underexplored and has considerable potential for increasing the understanding of modes of action. The design of chemical probes involves the addition of a tag to the molecule of interest, for example, fluorescent markers to track localisation and reactive functional groups that can interact with proteins, making sure that any alteration does not affect biological activity. The development of fluorescently tagged triterpenes has allowed for studies of their cellular localisation, narrowing down the potential molecular targets. For example, in a recent study, a cytotoxic

rhodamine conjugate of the triterpene asiatic acid was localised to the mitochondria, suggesting a mitochondrial target [11].

Further information about target engagement can be obtained using chemoproteomics, which involves developing a probe (often a photoaffinity probe) of the molecule of interest that can react with protein partners. After incubation of cells with the probe, UV irradiation leads to the reaction of the photosensitive linker with proteins in close proximity. The inclusion of an alkyne in the linker allows for isolation of these proteins through click chemistry to agarose beads, a process known as affinity enrichment. Proteins that interact with the molecule of interest are then identified by protein digestion and proteomic analysis. This approach is growing in popularity for unbiased identification of drug targets. The development of a photoaffinity probe for betulinic acid, a triterpene reported to interact with a plethora of targets to activate apoptosis, implicated the actin-binding protein tropomyosin as its putative target in breast cancer cells [12]. Photoaffinity probes have also identified the chromatin-binding protein high-mobility group box 1 as a possible target for the semisynthetic triterpene PT-22 [13] (Figure 2B). However chemoproteomics may produce misleading information because it directs the focus on abundant proteins that may be incidentally labelled as bystanders. Such results should therefore be interpreted with caution and require further validation by additional thorough pharmacological analysis.

The role of structural and computational biology in triterpene drug development

Understanding of the binding sites of triterpenes will accelerate with the continued rapid development of structural biology techniques such as cryoelectron microscopy (cryo-EM). As cryo-EM does not rely on the crystallisation of proteins with their binding partner (unlike X-ray crystallography), it opens avenues to determine previously inaccessible structural information for triterpenes bound to their protein partners. For example, the antifungal semisynthetic triterpene ibrexafungerp inhibits the fungal β-(1,3)-d-glucan synthase FKS1 by binding to the catalytic subunits [14]. A recent cryo-EM structure of FKS1 [15] highlights the potential for the use of structure-based design to develop further analogues active against ibrexafungerp-resistant strains. *In silico* docking of triterpenes with known molecular targets will help to identify key residues and novel binding pockets and inform structure-based drug design (Figure 2C).

Computational approaches to predict protein–small-molecule interactions now offer a powerful means of narrowing potential targets down to a small number of candidates that can then be evaluated experimentally. Target-based approaches exploit three-dimensional structural information for potential target proteins. Reverse docking, by contrast, screens the compound against a database of numerous protein structures to identify possible binding sites. AlphaFold2 allows binding of ligands to the predicted human proteome via reverse docking [16]. A new model (AlphaFold3) with substantially improved predictive capability for protein–ligand complexes has also recently been developed [17]. The application of such approaches should in the future provide insights into understanding the impact of triterpenes on cellular functions, identify potential protein targets, and enable rational design of new triterpene-based therapeutics.

Concluding remarks

Advances in bioengineering are now opening up unprecedented opportunities to access and harness a vast array of previously inaccessible triterpene chemical space. Unbiased approaches to target elucidation continue to develop and, coupled with novel predictive computational innovations (e.g., AlphaFold3), will accelerate understanding of the molecular targets and modes of action of these structurally diverse compounds. The triterpene structural diversity attainable with the rapidly growing toolkit of biosynthetic enzymes harvested from the Plant Kingdom can be triaged *in silico* using machine learning-based methods that predict the structures of bioactives for particular therapeutic indications [18], thereby guiding and informing directed biosynthesis of promising new triterpene therapeutics. Collectively these advances ultimately offer an entirely new pipeline for discovery and optimisation of next-generation triterpene therapeutics.

Acknowledgments

We acknowledge funding from Biotechnology and Biological Sciences Research Council (BBSRC) Follow-on-Fund grant APP4941, Wellcome Discovery Award 227375/Z/23/Z, the BBRSC Institute Strategic Programme Grant 'Harnessing biosynthesis for sustainable food and health' (BB/X01097X/1), and the John Innes Foundation.

Declaration of interests

A.O. is a cofounder and shareholder of HotHouse Therapeutics and serves as a consultant for this company. The remaining authors have no interests to declare.

Resources

[i]https://clinicaltrials.gov/study/NCT04493684?term=GSK3739937&rank=1

[1]Department of Biochemistry and Metabolism, John Innes Centre, Norwich Research Park, Colney Lane, Norwich NR4 7UH, UK
[2]School of Chemistry, Pharmacy, and Pharmacology, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK

*Correspondence:
anne.osbourn@jic.ac.uk (A. Osbourn).

https://doi.org/10.1016/j.tips.2024.10.003

© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

References

1. Brown, D.G. and Boström, J. (2016) Analysis of past and present synthetic methodologies on medicinal chemistry: where have all the new reactions gone? *J. Med. Chem.* 59, 4443–4458
2. Domingo-Fernández, D. *et al.* (2024) Natural products have increased rates of clinical trial success throughout the drug development process. *J. Nat. Prod.* 87, 1844–1851

3. Lee, A. (2023) Omaveloxolone: first approval. *Drugs* 83, 725–729
4. Lee, A. (2021) Ibrexafungerp: first approval. *Drugs* 81, 1445–1450
5. McInturff, E.L. *et al.* (2023) Synthetic approaches to the new drugs approved during 2021. *J. Med. Chem.* 66, 10150–10201
6. Alekseychuk, M. and Heretsch, P. (2023) Biogenetic space-guided synthesis of rearranged terpenoids. *Chem. Commun. (Camb.)* 59, 6811–6826
7. Reed, J. *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
8. Martin, L.B.B. *et al.* (2024) Complete biosynthesis of the potent vaccine adjuvant QS-21. *Nat. Chem. Biol.* 20, 493–502
9. Liu, Y. *et al.* (2024) Complete biosynthesis of QS-21 in engineered yeast. *Nature* 629, 937–944
10. Regueiro-Ren, A. *et al.* (2022) The discovery of GSK3640254, a next-generation inhibitor of HIV-1 maturation. *J. Med. Chem.* 65, 11927–11948
11. Heise, N. *et al.* (2023) Mitochondria-targeting 1,5-diazacyclooctane-spacered triterpene rhodamine conjugates exhibit cytotoxicity at sub-nanomolar concentration against breast cancer cells. *Int. J. Mol. Sci.* 24, 10695
12. Martín-Acosta, P. *et al.* (2022) A clickable photoaffinity probe of betulinic acid identifies tropomyosin as a target. *Acta Pharm. Sin. B* 12, 2406–2416
13. Shen, P. *et al.* (2024) Natural triterpenoid-aided identification of the druggable interface of HMGB1 occupied by TLR4. *RSC Chem. Biol.* 5, 751–762
14. Apgar, J.M. *et al.* (2021) Ibrexafungerp: an orally active β-1,3-glucan synthesis inhibitor. *Bioorg. Med. Chem. Lett.* 32, 127661
15. Hu, X. *et al.* (2023) Structural and mechanistic insights into fungal β-1,3-glucan synthase FKS1. *Nature* 616, 190–198
16. Varadi, M. *et al.* (2022) Alphafold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439–D444
17. Abramson, J. *et al.* (2024) Accurate structure prediction of biomolecular interactions with AlphaFold3. *Nature* 630, 493–500
18. Ross, J. *et al.* (2021) Large scale chemical language representations capture molecular structures and properties. *Nat. Mach. Intell.* 4, 1256–1264