The declining prospects of innovator pharmaceutical companies have been attributed to their inability to discover leads that bind to novel protein domains. All of the 21,000-odd drug products that have ever been approved by the US FDA bind to just 130 of the over 10,000 functional protein domains that exist in the human body. The high degree of drug redundancy, in turn, stems from the industry’s persistence with combinatorial chemistry to synthesize drug candidates. Not only is the chemical space being screened for biological activity usually one that has been previously interrogated, but the products of combinatorial chemistry too are simply not drug-like. So dire are the odds that launching a single synthetic compound necessitates screening as many as 100,000 molecules from a combinatorial library.

In comparison, as many as 20 molecules of the 7,000-odd polyketide natural products that have been screened thus far have made it to the market. Nevertheless, drug companies remain underwhelmed as synthetic inaccessibility has heretofore stymied drug prospection in natural product space. However, recent applications of genetic approaches to decipher the biosynthetic code that guides the assembly of several natural product families have revealed that living systems methodically and repetitively combine carbohydrate-derived building blocks in metabolic unit reactions to generate unprecedented molecular diversity. By clustering combinations, permutations and mutations of these biosynthetic genes into operons to be expressed as heterologous metabolic pathways in tractable microbial hosts, one can readily generate pharmacophores that are more likely to bind to novel targets. Ordered combinations of known genes generate pharmacophores that are natural products in toto, whereas permutations and mutations produce molecules that possess a scaffold that is an unnatural isomer of a natural product. The atom economy of microbial synthesis of pharmaceutical intermediates and active ingredients also far exceeds those of conventional schemes.
We recently demonstrated the feasibility of this new paradigm for drug discovery by integrating two of the four metabolic unit reactions involved in taxane assembly into the metabolism of *E. coli* to produce an unsubstituted, 20-carbon drug scaffold at 1 g/L titers. The taxane family of molecules includes, among others, the blockbuster anti-cancer drug, taxol. We then proceeded to express the first of many cytochrome P450 monooxygenases that oxidize the taxane scaffold in the third metabolic unit reaction. However, plant cytochrome P450s such as those involved in taxane biosynthesis are notoriously promiscuous and non-selective. A likely candidate that meets the synthetic requirements must first be selected, following which it must be customized to enhance its substrate specificity, product selectivity and activity. The use of structure-driven approaches is preferred herein as the structure-activity insights gained from redesigning one active site could potentially be extended to re-engineering downstream homologues. However, conventional methodologies such as X-ray crystallography are fairly lengthy and laborious. Ergo, we combined phylogeny-guided mutagenesis with a model-driven assessment of the active site of taxadiene-5α-hydroxylase, the cytochrome P450 that catalyzes proto-oxidation of taxane scaffold, to investigate its catalytic mechanism. The reactive topography of the active site was then charted to relate the impact of structural features to activity and selectivity, opening up the future possibilities to specifically tune the reactivity of cytochrome P450s to catalyze the regio- and stereospecific oxidation of the taxane scaffold for prospection of targeted regions of natural product chemical space.