Scilight

Natural Products Analysis

https://www.sciltp.com/journals/npa

Review

Constraint Geometry in Natural Product Biosynthesis

Simon Gibbons

Centre for Natural Products Discovery (CNPD), School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK; S.Gibbons@ljmu.ac.uk or anabasis@mail.com

How To Cite: Gibbons, S. Constraint Geometry in Natural Product Biosynthesis. *Natural Products Analysis* **2026**, *2*(1), 100011. https://doi.org/10.53941/npa.2026.100011

Received: 11 April 2026
Revised: 4 June 2026
Accepted: 23 June 2026
Published: 30 June 2026

Abstract: Natural product biosynthesis is often framed through the lenses of enzymatic novelty, pathway diversity, and evolutionary contingency. Yet across systems, a deeper, more profound regularity emerges, biosynthetic space is shaped not only by genetic potential but by the geometric constraints imposed by ecological context, metabolic flux, and structural feasibility. This perspective outlines a minimal conceptual framework for understanding how constraint geometry governs the emergence, stability, and diversification of natural product families. Rather than proposing new theory, it synthesizes existing observations into a coherent lens that highlights why certain scaffolds recur, why others remain rare, and how ecological pressures delimit the accessible chemical landscape. The aim of this paper is to provide a compact, integrative reference point for researchers seeking to understand the deeper architecture underlying biosynthetic outcomes.

Keywords: natural product biosynthesis; chemical space constraints; scaffold recurrence; metabolic economy; chemical ecology; biosynthetic convergence

1. Introduction: Why Constraints Matter

Natural product chemistry has long celebrated novelty. The discovery of new molecular scaffolds, previously unseen biosynthetic pathways, and unusual enzymatic transformations remains a central driver of the field, from chemical ecology to drug discovery. Advances in genomics, metabolomics, and computational mining continue to expand access to previously uncharted regions of natural product chemical space, reinforcing the sense that nature's capacity for molecular invention is vast and only partially explored [1,2].

Yet novelty alone does not explain some of the most persistent and striking regularities in natural product chemistry. Across taxa and environments, certain scaffolds dominate chemical space, while others recur with remarkable consistency despite arising from distinct biosynthetic lineages. Comparative analyses of natural product databases show that newly reported compounds are, on average, structurally similar to previously known molecules, and that the range of scaffolds readily produced by biological systems is significantly more constrained than the space of theoretically possible small molecules [3].

These patterns are not limited to structural statistics. Independent biosynthetic families frequently converge on closely related molecular architectures, even when the underlying gene clusters, enzymes, and evolutionary histories differ. Chemical ecology further shows that ecological transitions, such as shifts in habitat, symbiotic association, or competitive pressure, often produce predictable changes in metabolite classes rather than arbitrary diversification. An example of this is the biosynthesis of the cyanogenic glycosides such as linamarin (**1**) and lotaustralin (**2**) in developing flax [4] (Figure 1).

Similarly, vast regions of combinatorially accessible chemical space remain sparsely populated by natural products, despite being synthetically feasible and, in some cases, pharmacologically attractive [3,5]. The number of potential organic molecules is enormous [6], natural products occupy only a small and highly biased subset of chemical space. Entire regions of synthetically accessible space remain unoccupied, including highly fluorinated

compounds, randomly substituted hydrocarbons, and many planar aromatic scaffolds. These gaps arise because biosynthetic enzymes can access only a limited set of chemical transformations, while evolutionary selection favours molecules with specific physicochemical properties suited to biological targets. Consequently, biologically relevant chemical space is highly structured, leaving vast regions of chemically plausible but biologically inaccessible molecules. Natural products can therefore be viewed as occupying privileged regions of chemical space shaped by evolutionary pressure, rather than sampling the space uniformly.

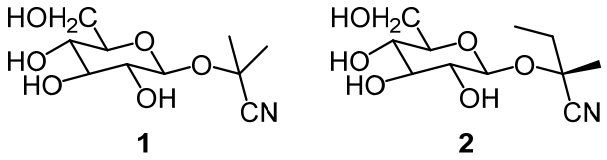


Figure 1. The chemical structures of linamarin (**1**) and lotaustralin (**2**).

These observations suggest that natural product chemistry is not primarily organised by unbounded exploration of novelty, but by structured constraint. Importantly, constraint here should not be understood as mere limitation or lack. Instead, constraints act as generative forces that shape what is chemically accessible, biosynthetically stable, ecologically functional, and evolutionarily persistent. Biological systems do not sample chemical space uniformly; they repeatedly occupy particular regions defined by enzymatic feasibility, metabolic economy, physicochemical compatibility, and ecological utility [2,7].

Seen in this way, constraints impose a form of geometry on natural product chemical space. They define not only what is possible, but what is probable, reusable, and robust under selective pressure. This geometric structure helps explain why unrelated organisms converge on similar metabolite frameworks, why certain molecular motifs are repeatedly elaborated rather than abandoned, and why biosynthetic innovation tends to proceed through modification of existing architectures rather than exploration of entirely new ones [3].

This perspective is not a new theory. Elements of it are already embedded across the literature in discussions of scaffold reuse, biosynthetic gene cluster conservation, chemical ecology, and natural product-informed exploration of chemical space. However, these insights are rarely synthesised into an explicit organising principle. As a result, constraint is often treated as background noise, an inconvenience to be overcome, rather than as a primary explanatory dimension of natural product structure and diversity.

In this article, I draw these threads together and argue that constraints, understood here as the structured shaping of chemical and biological possibility, provide a deeper organising principle for natural product science than novelty alone. Making these constraints explicit allows long-standing empirical patterns to be understood as outcomes of biological geometry, rather than as coincidental or purely historical artifacts.

2. Why Certain Scaffolds Dominate Chemical Space

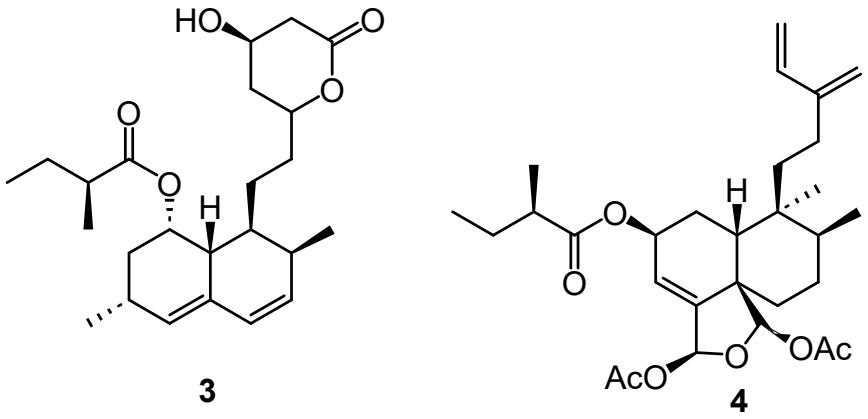
Large-scale analyses of natural product databases consistently show that a relatively small number of core scaffolds account for a disproportionate fraction of known natural products. Quantitative scaffold classifications and chemoinformatic studies demonstrate that natural products cluster into tightly bounded regions of biologically relevant chemical space rather than being uniformly distributed across all synthetically conceivable structures. This scaffold dominance persists even as new compounds are discovered, indicating that biological synthesis repeatedly returns to particular structural frameworks that are enzymatically accessible, metabolically economical, and functionally robust under selection pressure [3,8].

These results suggest that scaffold prevalence reflects not historical sampling bias alone, but the outcome of constrained evolutionary exploration, in which certain molecular architectures are preferentially retained because they occupy chemically and biologically stable positions in nature's chemical landscape.

2.1. Why Biosynthetic Families Converge on Similar Architectures

Independent biosynthetic pathways frequently converge on closely related molecular architectures despite diverging genetically and evolutionarily. Reviews of convergent and branching biosynthetic strategies in bacteria, fungi, and plants show that distinct enzyme families repeatedly generate overlapping skeletons from common metabolic precursors. Polyketides, terpenes, non-ribosomal peptides, and alkaloids all exemplify this phenomenon, where different enzymatic routes resolve similar chemical problems using convergent solutions [9,10]. A readily apparent example of this is lovastatin (Mevinolin) (**3**) and the clerodane diterpene (**4**), with both having structural and moietal similarities (Figure 2). However, compound **3** is from the fungus *Aspergillus terreus* [11] and **4** is

from the Costa Rican Rainforest tree *Casearia tremula* [12]. Such superficial similarities may hide a far more profound fact that polyketide synthases and terpene cyclases share a common ancestor and possibly even a similar tempo of action.



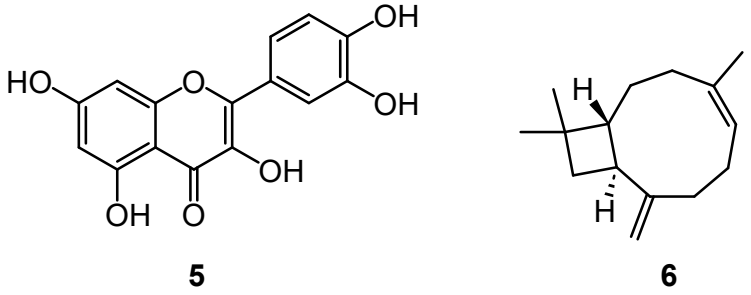
3 **4**

Figure 2. The chemical structures of lovastatin (Mevinolin) (**3**) and the clerodane diterpene (**4**).

This convergence reflects shared constraints imposed by substrate availability, reaction energetics, folding pathways, and functional interaction with targets. Rather than representing redundant invention, convergent architectures reveal the limited number of biosynthetically and ecologically viable ways to construct stable, bioactive molecules within living systems.

2.2. Why Ecological Transitions Produce Predictable Chemical Shifts

Chemical ecology demonstrates that changes in habitat, trophic interactions, stress exposure, or symbiotic relationships often produce systematic and repeatable changes in natural product profiles. Across microbial, plant, and marine systems, ecological pressures such as competition, predation, symbiosis, and abiotic stress reliably bias metabolite production toward particular compound classes. Examples of these include defensive phenolics such as quercetin (**5**) [13], signalling terpenes like β-caryophyllene (**6**) [14], and antimicrobial peptides rather than random diversification [15,16] (Figure 3).



5 **6**

Figure 3. The chemical structures of quercetin (**5**) and β-caryophyllene (**6**)

These observations indicate that natural product distributions are shaped by ecological constraints that favour metabolites occupying functionally effective regions of chemical space. Ecological transitions do not expand chemical possibility indiscriminately; they redirect biosynthesis toward regions already compatible with existing metabolic and signalling infrastructures.

2.3. Why Some Theoretically Possible Structures Never Appear

Despite the combinatorial immensity of chemical space, natural products occupy only a narrow, biologically validated subset. Comparative analyses between natural products, approved drugs, and synthetic libraries show that vast regions of synthetically accessible space remain empty of natural compounds, including many structures that are stable and human-synthesizable. Reviews of chemical space navigation emphasize that biological systems do not explore chemical space freely, but are constrained by enzymatic specificity, toxicity thresholds, solubility limits, transport constraints, and integration into cellular networks [5,17].

This absence is informative rather than incidental as it marks regions of chemical space that are chemically feasible yet biologically inaccessible or maladaptive. Natural products therefore map not the full space of what can exist, but the constrained geometry of what can persist within living systems.

Taken together, these patterns show that constraints act as organizing forces, shaping the structure, distribution, and recurrence of natural products. Far from hindering chemical diversity, constraints define the geometry of biologically meaningful diversity, that is specifically what is reachable, stable, and repeatedly rediscovered. Making these constraints explicit allows long-standing empirical observations in natural product chemistry to be understood as coherent outcomes of structured chemical possibility rather than as isolated anomalies.

3. Constraint Geometry as an Integrative Lens

Natural product diversity emerges at the intersection of chemistry, metabolism, and ecology. While these domains are often analysed separately, many of the strongest regularities in natural product chemistry arise from their simultaneous action. Viewed together, they impose a structured geometry on chemical space that explains recurrence, rarity, and convergence without invoking teleology or chance.

3.1. Why Scaffolds Recur

Recurrent natural product scaffolds, such as terpenoids, polyketides, alkaloids, and non-ribosomal peptides, consistently occupy regions of chemical space that satisfy three intersecting constraints.

Firstly, they are structurally feasible. These scaffolds arise from reaction manifolds that enzymes can reliably execute and control, such as terpene cyclization, polyketide chain extension and folding, and amine-centred heterocycle formation. Decades of biosynthetic analysis show that relatively few core enzymatic logics account for a large fraction of known natural product backbones, leading to repeated reuse of particular molecular frameworks [18,19].

Secondly, they are metabolically affordable. Comparative studies of biosynthetic gene clusters indicate that dominant scaffolds are constructed from abundant primary metabolites and leverage modular enzymatic machinery, minimizing energetic and regulatory cost. Scaffold reuse often reflects the ability to diversify products through late-stage tailoring rather than rebuilding costly backbones de novo [10].

Finally, they exhibit ecological utility. Recurrent scaffolds repeatedly mediate interactions such as defence, signalling, competition, or symbiosis across multiple lineages. Chemical ecology studies demonstrate that these frameworks occupy biologically effective regions of chemical space, where modest modifications can tune activity without sacrificing stability or bioavailability [1].

Together, these features make recurrent scaffolds geometrically stable solutions: they lie at intersections of feasibility, affordability, and function that biological systems can reach repeatedly and maintain over evolutionary time.

3.2. Why Some Scaffolds Remain Rare

In contrast, rare natural product scaffolds often violate one or more of the same constraints. Some are metabolically expensive, requiring long, tightly regulated pathways or unusual precursor pools that limit their distribution to narrow taxonomic or environmental contexts. Others are structurally unstable, positioned near physicochemical boundaries where small perturbations compromise folding, solubility, or persistence in cellular or extracellular environments. Still others are ecologically narrow, conferring benefits only under highly specific conditions that restrict their evolutionary retention [7].

Chemoinformatic analyses comparing natural products with synthetic libraries highlight that many theoretically possible and synthetically accessible scaffolds are absent from nature, not because they cannot exist, but because they cannot be robustly integrated into living systems. These absences mark chemically feasible yet biologically unstable regions of chemical space.

Such scaffolds represent geometrically unstable solutions. They may arise sporadically, but they do not persist or spread because they fail to satisfy the combined constraints that favour recurrence.

3.3. Why Biosynthetic Families Converge

Biosynthetic convergence is often described as surprising, but it becomes expected when viewed through a constraint-based lens. Independent biosynthetic families repeatedly converge on similar molecular architectures because they navigate the same underlying constraint landscape.

Distinct enzyme classes working from shared precursor pools face similar trade-offs involving reaction energetics, intermediate stability, and functional output. Convergent and branching biosynthetic pathways

documented across bacteria, fungi, and plants show how unrelated gene clusters resolve these trade-offs in comparable ways, yielding overlapping scaffold types despite genetic divergence [18,19].

In this view, convergence reflects neither redundancy nor limited creativity, but the fact that chemical space is unevenly accessible under biological constraints. Systems exploring similar ecological problems with similar metabolic vocabularies repeatedly settle on the same structurally and functionally viable regions.

Together, scaffold recurrence, scaffold rarity, and biosynthetic convergence all reflect the same underlying principle: natural product chemistry is shaped by the geometry of constraint. The chemical universe available to biology is vast in theory but highly structured in practice. Understanding this structure helps explain why nature's chemistry is at once diverse, patterned, and predictably incomplete.

4. Case Examples: Scaffold Recurrence, Metabolic Economy, and Ecological Coherence

This section illustrates how recurring patterns in natural product chemistry arise from intersecting constraints using concrete, well-known compounds. The goal is not exhaustiveness, but clarity.

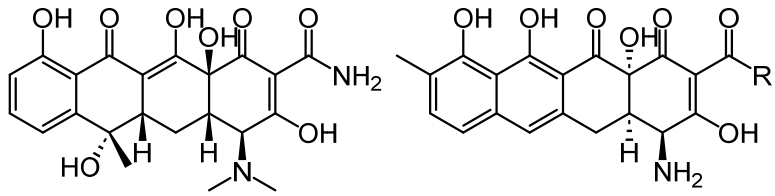
4.1. Scaffold Recurrence: Well-Known Exemplars

Several of the most intensively studied natural products illustrate scaffold recurrence across unrelated taxa and contexts.

Artemisinin and taxol, though pharmacologically distinct, are both terpenoid-derived structures built from isoprene units through robust cyclisation chemistry. Terpene biosynthesis relies on a small set of universally conserved precursors and enzyme classes, enabling the repeated emergence of complex polycyclic frameworks with high stereochemical density from simple linear substrates. Comparative biosynthetic studies show that such cyclisation reactions are unusually tolerant to variation, supporting repeated scaffold reuse across evolutionary time [18,19].

Menthol and β-caryophyllene illustrate the same principle at different scales: both are terpenoids derived from relatively short isoprene chains, yet they fulfil radically different ecological roles (volatile signalling versus membrane-active defence). Chemical ecology literature demonstrates that small modifications of shared terpenoid scaffolds can tune volatility, hydrophobicity, and molecular recognition while preserving biosynthetic accessibility [1].

Tetracycline (**7**) provides a parallel example in polyketide chemistry [20]. Despite extensive structural elaboration, its core scaffold arises from repeated β-keto extension and controlled cyclisation, mechanisms that recur across many polyketide families. Large-scale retrospectives of natural product discovery show that tetracycline-like frameworks are repeatedly rediscovered with peripheral variation rather than replaced by entirely new backbones [10]. A beautiful example of this is chelocardin (**8**) and its amide analogue amidochelocardin (**9**) [21]. Both are structural analogues of tetracycline, yet have subtle differences in polyketide substitution (but the same polyketide chain length) and functional group decoration (Figure 4).



7 tetracycline

8 R = CH_3, chelocardin
9 R = NH_2, amidochelocardin

Figure 4. The chemical structures of tetracycline (**7**), chelocardin (**8**) and amidochelocardin (**9**).

Together, these compounds illustrate scaffold recurrence as an outcome of reliable chemistry applied in many contexts, not isolated historical accidents.

4.2. Metabolic-Economy Contrasts: Penicillin G Versus Vancomycin

The contrast between penicillin G (**10**) and vancomycin (**11**) illustrates how metabolic cost shapes persistence without determining value (Figure 5).

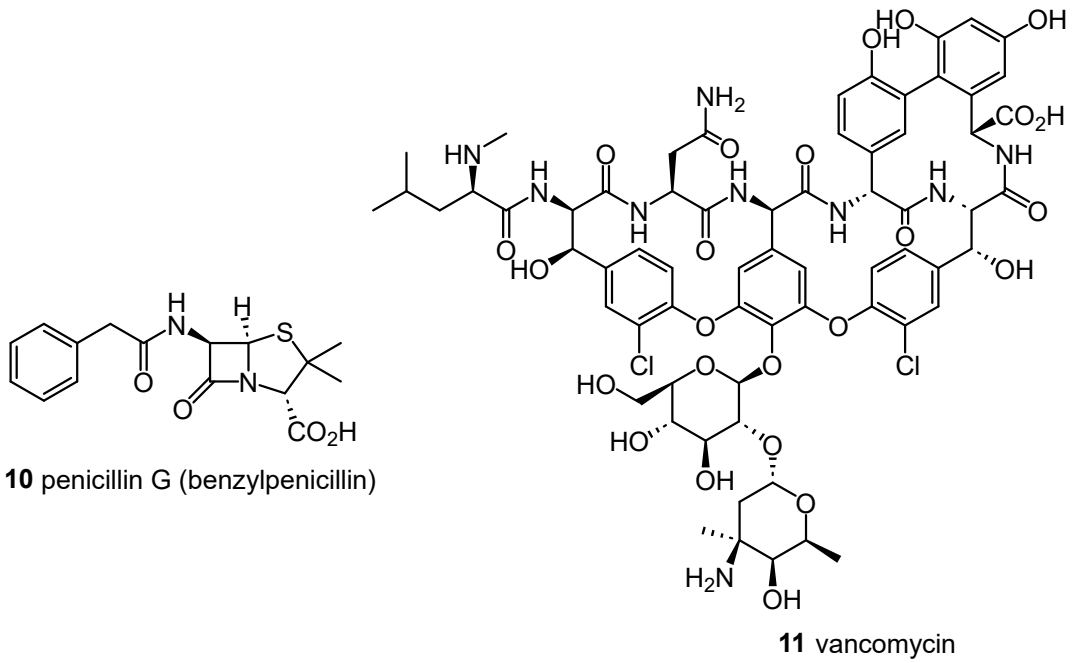


Figure 5. The chemical structures of penicillin G (**10**) and vancomycin (**11**).

Penicillin G is biosynthetically inexpensive and is assembled from readily available amino acid precursors using a compact gene cluster and a small number of enzymatic steps. Its β-lactam core is chemically simple yet biologically effective, enabling large-scale production and broad taxonomic distribution among fungi and industrial strains [19]. Vancomycin, by contrast, is among the most biosynthetically complex antibiotics known. Its assembly requires extensive oxidative cross-linking, glycosylation, and halogenation, and depends on tightly coordinated enzymatic choreography [22]. These features impose substantial metabolic and regulatory cost, which helps to explain why vancomycin-type scaffolds are comparatively rare and phylogenetically restricted [7].

Yet vancomycin persists because its ecological and pharmacological payoff is exceptionally high. Its mode of action targets cell-wall synthesis with extraordinary specificity, conferring strong selective advantage in competitive microbial environments. This contrast illustrates how expensive scaffolds are maintained only when functional returns justify their cost.

4.3. Ecological Coherence: Capsaicin, Quinine, Pyocyanin, Siderophores

Many natural products achieve persistence not through abundance, but through tight ecological alignment. Capsaicin exemplifies metabolites whose structure and activity are finely tuned to ecological niches. Its pungency deters mammalian herbivores while permitting avian dispersal, demonstrating how specific biological interactions can sustain chemically specialised compounds despite biosynthetic investment [23].

The alkaloid quinine, illustrates a similar principle in chemical defence. Nitrogen incorporation makes alkaloid biosynthesis energetically demanding, but quinine's potent bioactivity against parasites and herbivores confers sufficient advantage to maintain its production in *Cinchona* species over evolutionary time [7].

Pyocyanin shows how ecological function shapes persistence even in chemically reactive compounds. Though redox-active and potentially destabilising, pyocyanin plays a central role in microbial competition and signalling in *Pseudomonas* species, linking its production directly to ecological success rather than chemical elegance [24].

Siderophores present a final example of ecological coherence. These molecules are structurally diverse yet functionally unified, optimized for iron sequestration in resource-limited environments. Their widespread occurrence across microbial taxa reflects the universal constraint of iron availability and the strong selective pressure to solve it chemically [1].

4.4. Synthesis: Constraint-Shaped Organisation of Major Classes

Viewed collectively, these examples show how major natural product classes occupy different positions shaped by intersecting constraints.

Terpenoids tend to combine high structural feasibility with broad ecological applicability, explaining their numerical dominance. Polyketides occupy a middle position, where modular biosynthesis enables wide structural reach at manageable metabolic cost. Alkaloids sit at a high-cost edge of chemical space, persisting only where ecological payoff, specifically defence, signalling, or targeting, is sufficiently strong to compensate.

These patterns do not require new explanatory constructs. They reflect how chemistry, metabolism, and ecology jointly shape which molecules are repeatedly made, which remain rare, and which never appear at all.

5. Implications for Discovery

Understanding natural product chemistry as structured by intersecting constraints has direct consequences for how discovery efforts are designed, evaluated, and refined. Rather than treating unexplored chemical space as uniformly promising, this perspective highlights where discovery is most likely to succeed—and why some pipelines consistently underperform.

5.1. Predicting Where New Scaffolds Are Likely to Emerge

New natural product scaffolds are most likely to arise in systems where three conditions coincide.

Firstly, structural feasibility must be high. Genome-enabled discovery has shown that biosynthetic gene clusters encode a limited set of enzymatic logics that are repeatedly recombined and adapted. New scaffolds most often emerge not from entirely new chemistries, but from novel rearrangements, fusions, or modifications of existing biosynthetic modules, particularly in polyketide, terpene, and peptide pathways [3,25].

Secondly, metabolic cost must be manageable. Reviews of discovery success consistently note that productive systems leverage abundant precursors and modular enzymology, allowing structural diversification without prohibitive energetic or regulatory burden. This is particularly evident in microbes with large biosynthetic repertoires constrained by genome size and metabolic economy [26].

Thirdly, ecological pressure must be strong. Chemical ecology and comparative genomics show that novel natural products most often arise in competitive, resource-limited, or symbiotic environments, where small chemical advantages translate into large fitness effects. Targeting such systems, rather than randomly sampling biodiversity, has been repeatedly identified as a high-yield strategy [27].

Together, these criteria help explain why certain environments (e.g., actinomycetes, symbiotic microbes, stressed plants) remain disproportionately productive sources of new chemistry.

5.2. Understanding Why Some Discovery Pipelines Stagnate

Many natural product discovery pipelines stall not because novelty is exhausted, but because they implicitly pursue chemically implausible regions of space. Industrial and academic reviews identify rediscovery as a dominant failure mode in traditional pipelines, particularly when screening strategies repeatedly target the same taxa, culture conditions, or bioactivity endpoints. In these cases, effort concentrates on regions of chemical space already saturated by common scaffolds, yielding diminishing returns [28].

Other pipelines fail by targeting regions that violate metabolic or ecological constraints, for example, highly complex structures that are biosynthetically rare, unstable intermediates unlikely to accumulate, or molecules without clear ecological function. Retrospective analyses show that such regions are sparsely populated by natural products despite being synthetically conceivable, explaining why brute-force search strategies frequently underperform [3].

Recognizing stagnation as a consequence of constraint mismatch reframes failure from a technical to a strategic problem.

5.3. Designing more Efficient Search Strategies

Efficient natural product discovery increasingly relies on constraint-aware search, focusing effort on chemically and biologically probable space rather than the full theoretical universe.

Modern strategies, including genome mining, biosynthetic gene cluster prioritization, and machine-learning-assisted prediction, implicitly operationalize this approach by ranking candidates based on biosynthetic plausibility, precursor availability, and ecological context. Tools that integrate genetic, chemical, and ecological information consistently reduce rediscovery rates and improve hit quality compared to untargeted screening [29,30].

From this perspective, successful discovery does not maximize novelty in the abstract. It maximizes alignment with the constraints that shape natural product chemistry in the first place. Discovery becomes a matter

of navigating probable chemical space—where feasibility, economy, and function intersect—rather than attempting to exhaust what is merely possible.

These implications suggest that discovery is most productive when guided by an explicit understanding of chemical constraint. Making these constraints visible allows researchers to predict where novelty is likely to arise, diagnose why pipelines stagnate, and design search strategies that are both more efficient and more empirically grounded.

6. The Conceptual Framework

Natural product research has generated a large and diverse empirical literature describing recurring patterns in structure, biosynthesis, and function. However, these patterns are often treated as separate observations rather than as manifestations of a small number of organizing principles. This section articulates three such principles that are already implicit in the field and supported by a broad range of existing studies.

6.1. Feasible Structures Appear more often than Possible Structures

Although the theoretical space of small organic molecules is vast, natural products populate only a restricted and highly structured subset of this space. Large-scale scaffold analyses consistently show that a relatively small number of molecular frameworks account for a disproportionate fraction of known natural products, while many synthetically accessible structures are absent from nature [31].

This pattern reflects the distinction between chemical *possibility* and biological feasibility. Feasible structures are those that enzymatic systems can reliably construct, fold, and stabilize under physiological conditions. Comparative cheminformatics demonstrates that natural product scaffolds cluster into chemically robust motifs optimized for interaction with biological macromolecules, rather than exploring chemical space uniformly [32].

Rare natural products containing unusual bonds or strained architectures—such as N–O heterocycles—illustrate this principle by their scarcity, despite being chemically synthesizable. Reviews of such scaffold classes explicitly note their limited natural distribution and biosynthetic specialization, underscoring the selective pressure favouring feasible over merely possible structures [33].

6.2. Affordable Pathways Persist Longer than Expensive Ones

Biosynthetic pathways differ substantially in energetic cost, enzymatic complexity, and regulatory burden. Comparative studies in metabolic biology show that biological systems are subject to strong selective pressure to balance functional gain against the energetic cost of biosynthesis, a principle observed across levels of biological organization [34].

In natural product chemistry, this is reflected in the persistence of pathways built on abundant precursors and modular enzymology, such as terpene cyclases and polyketide synthases. These systems enable extensive diversification through domain reuse and late-stage modification, minimizing the need for energetically costly de novo innovation.

Conversely, pathways requiring rare precursors, extensive tailoring, or multi-step oxidative transformations tend to be taxonomically restricted and phylogenetically fragile. The metabolic-cost framework, though often discussed in primary metabolism, provides a useful explanatory backdrop for why certain secondary metabolite families remain sparse despite high functional potency [35]. However, although plants can access abundant energy through photosynthesis, the production of secondary metabolites is constrained by trade-offs in resource allocation, particularly for limiting nutrients such as nitrogen, meaning that metabolic cost still shapes the distribution of compound classes.

6.3. Ecologically Coherent Metabolites Dominate Functional Space

Natural products are not distributed randomly with respect to biological function. Chemical ecology research consistently shows that secondary metabolites cluster around ecological roles such as defence, competition, signalling, and resource acquisition. Compounds fulfilling these roles efficiently recur across unrelated lineages, often with conserved or convergent structural motifs [36].

Recent work further demonstrates that many secondary metabolites are multifunctional, acting as toxins at high concentrations and as signals or regulators at lower doses. This functional versatility contributes to their persistence, as it allows a single chemical scaffold to operate across multiple ecological contexts [37].

Environmental stress studies in plants and microbes also reveal predictable shifts toward particular metabolite classes under competitive or resource-limited conditions, reinforcing the idea that ecological coherence strongly shapes which molecules are retained and elaborated in nature [16].

These three principles capture a shared empirical insight: natural product chemistry is structured by what biological systems can feasibly build, affordably maintain, and ecologically deploy. Framing discovery and analysis around these principles does not require abandoning existing models or introducing new theory. Rather, it makes explicit the constraints that already underpin observed regularities across natural product structure, biosynthesis, and function.

7. Conclusions

Natural product biosynthesis is not a random walk through chemical possibility. Decades of empirical work in natural product chemistry, cheminformatics, and chemical ecology demonstrate that biological systems populate chemical space unevenly, repeatedly returning to particular structural motifs while leaving vast regions unexplored. Large-scale analyses of scaffold distributions show that a small number of frameworks dominate natural product databases, reflecting patterns of biological feasibility rather than exhaustive chemical exploration [31].

These patterns arise because biosynthesis operates within a geometry of constraints. Enzymatic mechanisms delimit which chemical transformations are reliable, metabolic economics shape which pathways are maintainable, and ecological pressures determine which compounds persist and matter over evolutionary time. Comparative studies of biosynthetic machinery illustrate how modular enzymatic systems repeatedly generate diverse compounds within bounded chemical architectures, reinforcing recurrence rather than unlimited divergence [22,38].

Ecological coherence further structures this landscape. Chemical ecology has shown that secondary metabolites recur in association with specific functional roles—defence, signalling, competition, and resource acquisition—and that these roles exert strong selective pressure on metabolite persistence. Metabolites that efficiently mediate ecologically relevant interactions dominate functional chemical space, even when they arise from taxonomically distant organisms [39,40].

This framing provides a unifying vocabulary for interpreting past discoveries and a pragmatic guide for future ones, without redefining the field, but by clarifying the deeper architecture that has shaped it all along.

By highlighting constraint rather than novelty, this perspective offers a compact and integrative way to understand why natural product chemistry is simultaneously diverse, patterned, and incomplete. It does not replace existing models of biosynthesis, evolution, or chemical ecology, nor does it introduce new theoretical constructs. Instead, it makes explicit the shared assumptions already embedded across these fields: that what emerges, what endures, and what proves significant in natural product chemistry is governed less by what is chemically imaginable than by what is biologically feasible, affordable, and ecologically coherent.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

Conflicts of Interest

The author declares no conflict of interest.

Use of AI and AI-Assisted Technologies

This text of this paper was written by SG and then smoothed by Microsoft Co-Pilot.

References

1. Bogart, J.W.; Cabezas, M.D.; Vögeli, B.; et al. Cell-Free Exploration of the Natural Product Chemical Space. *Chembiochem* **2021**, *22*, 84–91.

2. Angeles Flores, G.; Cusumano, G.; Venanzoni, R.; et al. Biodiversity-Driven Natural Products and Bioactive Metabolites. *Plants* **2025**, *15*, 104
3. Pye, C.R.; Bertin, M.J.; Lokey, R.S.; et al. Retrospective analysis of natural products provides insights for future discovery trends. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 5601–5606.
4. Zuk, M.; Pelc, K.; Szperlik, J.; et al. Metabolism of the Cyanogenic Glucosides in Developing Flax: Metabolic Analysis, and Expression Pattern of Genes. *Metabolites* **2020**, *10*, 288.
5. Nelson, A.; Karageorgis, G. Natural product-informed exploration of chemical space to enable bioactive molecular discovery. *RSC Med. Chem.* **2021**, *12*, 353–362.
6. Lipinski, C.; Hopkins, A. Navigating chemical space for biology and medicine. *Nature* **2004**, *432*, 855–861.
7. Anderson, V.M.; Wendt, K.L.; Najar, F.Z.; et al. Building Natural Product Libraries Using Quantitative Clade-Based and Chemical Clustering Strategies. *mSystems* **2021**, *6*, e00644-21. https://doi.org/10.1128/msystems.00644-21.
8. Koch, M.A.; Ansgar, S.; Michael, S.; et al. Charting Biologically Relevant Chemical Space: A Structural Classification of Natural Products (SCONP). *Proc. Natl. Acad. Sci. USA.* **2005**, *102*, 17272–17277.
9. Dairi, T.; Kuzuyama, T.; Nishiyama, M.; et al. Convergent strategies in biosynthesis. *Nat. Prod. Rep.* **2011**, *28*, 1054–1086.
10. Wei, X.; Wang, W.G.; Matsuda, Y. Branching and converging pathways in fungal natural product biosynthesis. *Fungal Biol. Biotechnol.* **2022**, *9*, 6.
11. Alberts, A.W.; Chen, J.; Kuron, G.; et al. Mevinolin: A highly potent competitive inhibitor of hydroxymethylglutaryl-coenzyme A reductase and a cholesterol-lowering agent. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 3957–3961.
12. Gibbons, S.; Gray, A.I.; Waterman, P.G. Clerodane diterpenes from the bark of *Casearia tremula*. *Phytochemistry* **1996**, *41*, 565–570.
13. Riddick, E.W. Potential of Quercetin to Reduce Herbivory without Disrupting Natural Enemies and Pollinators. *Agriculture* **2021**, *11*, 476.
14. Köllner, T.G.; Held, M.; Lenk, C.; et al. A maize (*E*)-beta-caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties. *Plant Cell* **2008**, *20*, 482–494.
15. Vizcaino, M.I.; Guo, X.; Crawford, J.M. Merging chemical ecology with bacterial genome mining for secondary metabolite discovery. *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 285–299.
16. Sana; Aftab, T.; Naeem, M.; et al. Production of secondary metabolites under challenging environments: Understanding functions and mechanisms of signalling molecules. *Front. Plant Sci.* **2025**, *16*, 1569014.
17. Miyao, T.; Reker, D.; Schneider, P.; et al. Chemography of natural product space. *Planta Med.* **2015**, *81*, 429–435.
18. Ali, J.G.; Casteel, C.L.; Mauck, K.E.; et al. Chemical Ecology of Multitrophic Microbial Interactions: Plants, Insects, Microbes and the Metabolites that Connect Them. *J. Chem. Ecol.* **2020**, *46*, 645–648.
19. Jacoby, R.P.; Koprivova, A.; Kopriva, S. Pinpointing secondary metabolites that shape the composition and function of the plant microbiome. *J. Exp. Bot.* **2021**, *72*, 57–69.
20. Duggar, B.M. Aureomycin: A Product of The Continuing Search for New Antibiotics. *Ann. N. Y. Acad. Sci.* **1948**, *51*, 177–181.
21. Sikandar, A.; Popoff, A.; Jumde, R.P.; et al. Revision of the Absolute Configurations of Chelocardin and Amidochelocardin. *Angew. Chem. Int. Ed.* **2023**, *62*, e202306437.
22. Walsh, C.T.; Tang, Y. Major Classes of Natural Product Scaffolds and Enzymatic Biosynthetic Machinery. In *Natural Product Biosynthesis*; The Royal Society of Chemistry: London, UK, 2022; p. 1–21.
23. Tewksbury, J.J.; Nabhan, G.P. Seed dispersal. Directed deterrence by capsaicin in chilies. *Nature* **2001**, *412*, 403–404.
24. Castañeda-Tamez, P.; Ramírez-Peris, J.; Pérez-Velázquez, J.; et al. Pyocyanin Restricts Social Cheating in *Pseudomonas aeruginosa*. *Front. Microbiol.* **2018**, *9*, 1348.
25. Jensen, P.R. Natural Products and the Gene Cluster Revolution. *Trends Microbiol.* **2016**, *24*, 968–977.
26. Atanasov, A.G.; Zotchev, S.B.; Dirsch, V.M.; et al. Natural products in drug discovery: Advances and opportunities. *Nat. Rev. Drug Discov.* **2021**, *20*, 200–216.
27. Barry, S.M. Rethinking natural product discovery to unblock the antibiotic pipeline. *Future Microbiol.* **2025**, *20*, 179–182.
28. Harris, L.A.; Mitchell, D.A. Reactivity-based screening for natural product discovery. *Methods Enzymol.* **2022**, *665*, 177–208.
29. Mei, S. A Multi-Label Learning Framework for Predicting Chemical Classes and Biological Activities of Natural Products from Biosynthetic Gene Clusters. *J. Chem. Ecol.* **2023**, *49*, 681–695.
30. Liu, G.; Li, Y.; Ong, G.; et al. Unified genomic and chemical representations enable bidirectional biosynthetic gene cluster and natural product retrieval. *Sci. Rep.* **2026**, *16*, 49955.
31. Yongye, A.B.; Waddell, J.; Medina-Franco, J.L. Molecular scaffold analysis of natural products databases in the public domain. *Chem. Biol. Drug Des.* **2012**, *80*, 717–724.
32. Ertl, P.; Schuhmann, T. Cheminformatics Analysis of Natural Product Scaffolds: Comparison of Scaffolds Produced by Animals, Plants, Fungi and Bacteria. *Mol. Inform*. **2020**, *39*, 2000017.

33. Yan, L.-H.; Li, X.; Wang, B.-G. Natural products with 1,2-oxazine scaffold: Occurrence, chemical diversity, bioactivity, synthesis, and biosynthesis. *Nat. Prod. Rep.* **2023**, *40*, 1874–1900.
34. Hou, C. Energetic cost of biosynthesis is a missing link between growth and longevity in mammals. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2315921121.
35. Wortel, M.T.; Noor, E.; Ferris, M.; et al. Metabolic enzyme cost explains variable trade-offs between microbial growth rate and yield. *PLOS Comput. Biol.* **2018**, *14*, e1006010.
36. Reshi, Z.A.; Ahmad, W.; Lukatkin, A.S.; et al. From Nature to Lab: A Review of Secondary Metabolite Biosynthetic Pathways, Environmental Influences, and In Vitro Approaches. *Metabolites* **2023**, *13*, 895.
37. Erb, M.; Kliebenstein, D.J. Plant Secondary Metabolites as Defenses, Regulators, and Primary Metabolites: The Blurred Functional Trichotomy. *Plant Physiol.* **2020**, *184*, 39–52.
38. Medema, M.H.; Fischbach, M.A. Computational approaches to natural product discovery. *Nat. Chem. Biol.* **2015**, *11*, 639–648.
39. Cheng, J.; Zhao, Z.; Teng, B.; et al. Rethinking Microbial Chemical Ecology: Secondary Metabolites as Concentration-Dependent Signaling Hubs with Implications for Anti-Virulence Intervention. *Microorganisms* **2026**, *14*, 1074.
40. Kessler, A.; Kalske, A. Plant Secondary Metabolite Diversity and Species Interactions. *Annu. Rev. Ecol. Evol. Syst.* **2018**, *49*, 115–138.