Terray Therapeutics • 2026
Source: https://www.terraytx.ai/news-and-research/terrasynth

TerraSynth

Accelerating the Synthesis of AI-designed Small Molecules

Jun 11, 2026
Authors
Miles Wang-Henderson
Zack Strater
John Parkhill
Matthew Katcher
Narbe Mardirossian

Small-molecule drug discovery is bottlenecked by how fast chemists can actually make virtually designed molecules. TerraSynth is a generative synthesis planner that proposes new molecules together with a recipe for making them, so designs don't get stuck waiting on synthesis. Its reconstruction of drug-like molecules is 67% higher relative to the rate of the next best model, while being 1000x faster. This efficiency allows it to be used inside an optimizer, ensuring that AI-generated compounds are synthetically feasible. On real drug discovery projects at Terray, designs proposed this way were synthesized up to 2-4x faster relative to unconstrained counterparts, driving faster iteration cycles that speed up the progression of our internal and partnered programs.

Introduction

In small-molecule drug discovery, the latency of a DMTA (design-make-test-analyze) cycle[1] dictates how quickly teams can iterate from hit to clinical candidate. At Terray, we have built our EMMI platform at the intersection of experimentation and AI, creating an efficient and integrated drug discovery engine to advance our internal and partnered therapeutics programs.

The "Make" step in DMTA is the slowest step, as it involves realizing a virtual design as an experimentally-synthesized molecule. Synthesis times for designs can take anywhere from weeks to a few months for complex AI-generated molecules. Models that incorporate synthetic considerations into the design process have historically been either too slow or too limited in their coverage of chemical space. To realize the promise of AI speeding up small-molecule discovery timelines, we need efficient generative models that are constrained by the design language of synthetic chemistry[2].

This is the gap TerraSynth closes. It reconstructs more drug-like chemistry than any prior synthesis planner, and does so roughly 1000x faster than the next best baseline on reconstruction rate. That speed lets us use TerraSynth inside the inner-loop of an optimizer, to decode and score candidates fast enough that synthesis is a native part of the design loop, ensuring that every proposed compound has been optimized for synthesizability.

Across internal projects at Terray, designs proposed through this loop have reduced synthesis latency by up to 75%.

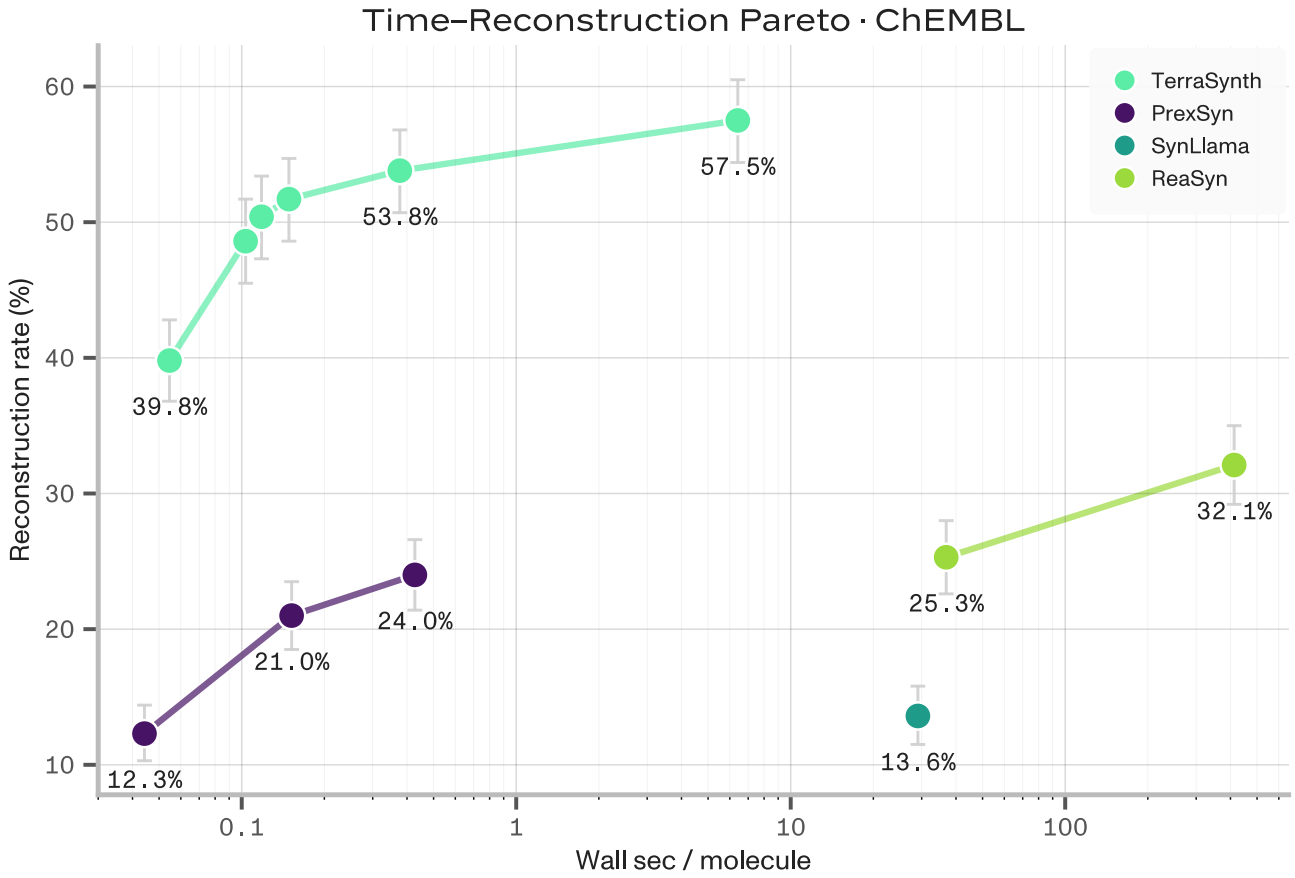

Figure 1. Reconstruction rate (%) versus inference wall-time on a random N=1,000 subset of ChEMBL with 95% bootstrapped CIs for TerraSynth and three established baselines. Times reported using single H100. TerraSynth strikes a good balance between chemical space coverage and speed.

Generative Synthesis Planning

Conventionally, a synthesis planner takes a target molecule and returns a route, which is a sequence of reactions over commercially available building blocks. For our purposes, "planner" is shorthand for a generative model over routes, not a retrosynthesis search engine that recurses backwards from the target molecule until purchasable building blocks are reached[3].

We want a planner to play two roles at once: (1) To *reconstruct* known chemistry, and (2) to *generate* new molecules whose routes are feasible by construction. This ensures both that the chemical space the model operates in is vast and diverse, and that every novel design it proposes is practically realizable at the bench. To do so requires that a planner is:

- **Expressive.** The model must span the chemistries that medicinal chemists tend to explore. If it cannot reconstruct internal leads or ChEMBL-like molecules, then it cannot propose them when coupled with an optimizer.
- **Feasible.** Every route the model emits has to use robust reactions and building blocks that are purchasable, with estimable costs.

- **Optimizable.** Designs come from optimizing in a latent space, so the planner has to expose one. We use COATI[4] to encode a molecule to a latent and decode to a route that yields the molecule.

This is challenging because, unlike commonly-used databases of small molecules that are used to pre-train small molecule foundation models, there is little public data that covers full multi-step routes annotated end-to-end.

Recent methods get around this by pre-training on routes generated from synthetic data engines. These routes are obtained by sampling from mutually compatible reaction templates and building-block catalogs. The resulting (product molecule, route) tuples serve as pre-training data[5]. This recipe relies on reaction templates as an inductive bias for feasibility and realism.

TerraSynth also leverages this recipe but significantly advances along the two axes that matter for practical use in small-molecule drug discovery workflows: spanning all relevant chemistry needed to make small-molecule drugs, and preserving the realism of sampled routes.

Developing TerraSynth

TerraSynth is a 2B-parameter autoregressive decoder over route tokens that leverages molecular representations from COATI. A target molecule can be encoded via COATI, TerraSynth then emits a sequence of building-block and reaction tokens in post-fix notation[6]. A key effort was to scale and curate the reaction templates supported to 260 of the most reliable chemical transformations in modern synthesis.

We pre-train on ~30B high-quality route tokens sampled from our data engine, where each route has a depth of up to 6 reactions. We then perform reinforcement learning (RL) against an interpretable reward function that penalizes inconsistencies which violate common synthetic chemistry rules.

The result is a planner that is fast to inference (>0.05 sec/mol) while emitting routes that are deemed likely to succeed by expert medicinal chemists.

Assessing Chemical Space Coverage

The first criterion[7] for any practical synthesis planner is expressiveness, which we measure by reconstruction. Can the planner emit a route whose product matches any drug-like molecule that we would like to evaluate?

By benchmarking reconstruction on common virtual catalogs like WuXi GalaXi and Enamine REAL, we can ensure that when paired with an optimizer, TerraSynth rarely precludes desirable molecules from being reached.

Speed matters as much as reconstruction here. Recent works like PrexSyn (Luo & Coley, 2025) have made progress along this axis. TerraSynth sits in the same speed regime while reconstructing significantly more relevant chemical matter to span a wider chemical space.

We evaluate synthesis planners on random subsets of virtual catalogs and chemical repositories, as well as Terray's own internal program leads. Enamine REAL molecules use a limited set of reactions and building blocks, so are easily reconstructed by most baselines. On collections with more complex molecules, like ChEMBL or Wuxi GalaXi, TerraSynth achieves almost double the reconstruction rate of the next best baseline.

Notably, TerraSynth is able to expend additional compute at test-time by re-sampling routes or searching over a conditional latent to improve reconstruction.

Reconstruction Benchmark

Dataset	PrexSyn	SynLlama	ReaSyn	TerraSynth		
				M=256	Greedy	TTS
ChEMBL	24.0±2.70.43 s/mol	13.6±2.129 s/mol	32.1±2.9413 s/mol	53.8±3.10.41 s/mol	39.8±2.90.05 s/mol	57.5±3.17.4 s/mol
Enamine REAL	88.6±2.00.30 s/mol	27.1±2.737 s/mol	95.2±1.446 s/mol	99.0±0.60.38 s/mol	83.0±2.40.06 s/mol	98.3±0.85.9 s/mol
Terray Program Leads	43.7±3.00.40 s/mol	19.9±2.526 s/mol	43.2±3.1416 s/mol	85.8±2.20.42 s/mol	71.5±2.80.06 s/mol	86.9±2.17.0 s/mol
WuXi GalaXi	41.6±3.00.41 s/mol	27.7±2.834 s/mol	55.1±3.0277 s/mol	96.5±1.20.48 s/mol	79.3±2.50.07 s/mol	96.0±1.37.6 s/mol

Table 1. Reconstruction rate (%) on N=1,000 subsets of each dataset (with 95% bootstrapped CIs) for TerraSynth and three established baselines, where a target counts as reconstructed if any sampled route's product matches the target by exact canonical SMILES. TerraSynth outperforms all baselines across the four considered datasets. Baselines run on author-recommended configs: PrexSyn samples 256 routes per target; SynLlama uses greedy sampling; ReaSyn uses BU+TD+EB with num_cycles=12. We show three scaling regimes for TerraSynth where speed can be traded for performance. (1) greedy M=1 sampling, (2) M=256 routes per target, (3) test-time search (TTS) over latent. We use M=256 as the standard variant for TerraSynth applications. The smaller grey value beneath each reconstruction rate is the inference wall-time (s/mol), reported on a single H100.

Balancing Potency and Synthesizability

A natural concern with synthesis-constrained optimization is that there is a strong trade-off between potency and synthesizability. In simulation, we find that TerraSynth is able to retrieve potent molecules that are also easy to synthesize.

To check this, we ran a simple experiment to verify that TerraSynth could reach potent chemical matter compared to an unconstrained baseline. We used TerraBind, our universal potency model, to predict pIC50. We then used a genetic algorithm (GA) over COATI latents as the optimizer, where the objective was to maximize predicted pIC50. We then ablated the decoder used in this design loop, using either COATI (unconstrained by synthesis) or TerraSynth.

As an independent proxy for synthesizability, we used the retrosynthesis engine AiZynthFinder (AZF)[8] to infer the route length of generated designs. To be comparable, we also provided AZF with the same TerraSynth building-block stock. Since AZF-predicted routes are independent of TerraSynth routes, AZF route lengths served as an unbiased proxy for synthetic difficulty.

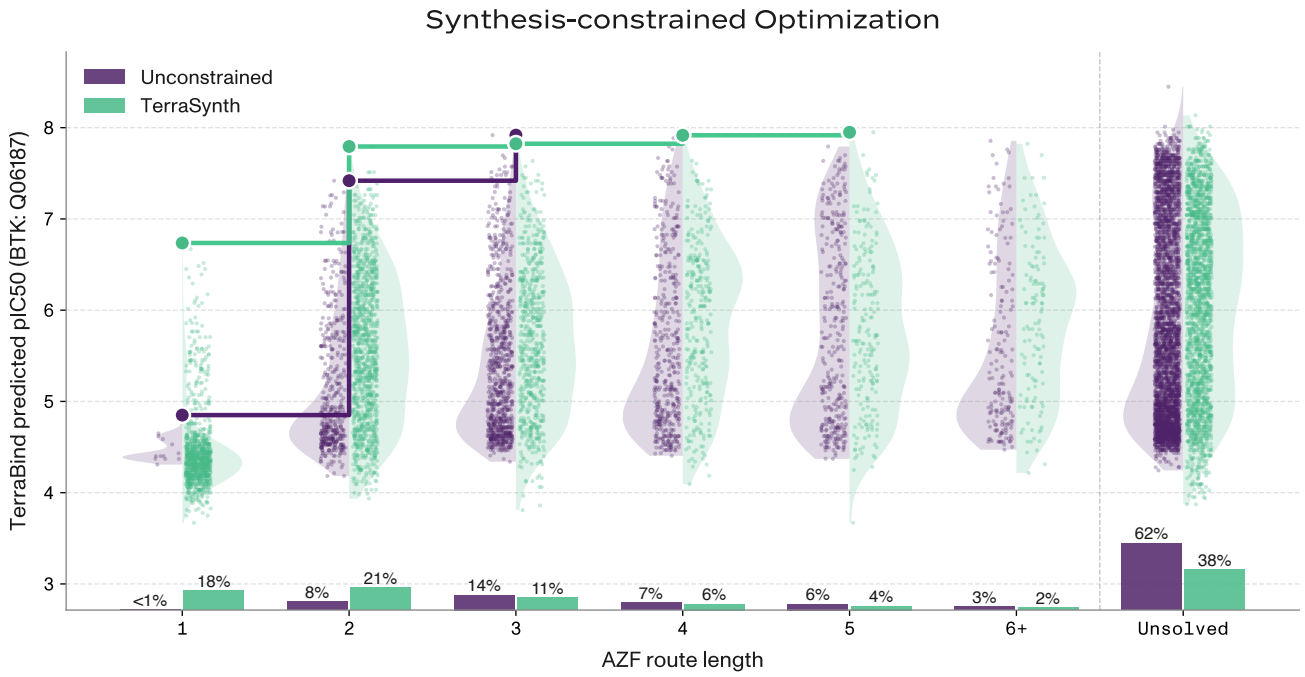

Figure 3. TerraBind predicted pIC50 of designs against a kinase target (BTK: UniProt Q06187) versus AZF route length. TerraSynth is able to retrieve potent molecules that are also easy to synthesize. Designs are obtained by genetic algorithm over COATI latent. TerraSynth decodes molecules with routes. Unconstrained baseline uses COATI to decode molecules and does not produce routes. Solid line traces the step-wise Pareto front (max pIC50 achievable in ≤ k AZF reactions). AZF-unsolved designs are in the rightmost bin. 62% of molecules from the unconstrained baseline are unsolved, versus 38% from TerraSynth.

TerraSynth proposes more potent molecules that take fewer reactions to make (median 2 vs. 4), and the majority of them are also AZF-solvable. Most molecules produced by the unconstrained baseline, however, are not AZF-solved[9].

As seen in this setup, using TerraSynth in a loop to constrain designs to synthesizable chemistry does not hinder our ability to reach potent compounds.

Accelerating DMTA Cycles

The ultimate test of a synthesis planner is whether it actually shortens the "Make" step in practice. For this prospective assessment, we looked at 17 internal projects requiring custom synthesis at Terray.

Eight of these projects contained compounds generated via an unconstrained decoder that does not jointly plan a synthetic route (before TerraSynth was developed). Five of these projects contained compounds proposed by our initial rollout of TerraSynth, and four of these projects contained compounds proposed by an optimized TerraSynth workflow that incorporates our latest learnings.

Compared to the unconstrained projects, using TerraSynth yielded a meaningful shift in cumulative synthesis rate. Time to reach 50% completion is halved from 8 weeks to 4 weeks.

Notably, the 2-week synthesis rate jumps from 13% to 33% (roughly 2.5×), and by 16 weeks the gap widens to 66% vs 90%. This represents a nearly +24% lift on complex compounds that would otherwise stall on a missing intermediate or a low-yielding step. In aggregate, this is months recovered per design cycle.

With the optimized TerraSynth workflow, the set of molecules proposed are optimized to share building-blocks in order to reduce their complexity and share intermediates. In addition, out of 22 initial designs per project, 15 are selected by a chemist based on synthetic difficulty as the final set.

These projects yielded a 90% synthesis rate within 4 weeks, which is 4x faster compared to the initial TerraSynth workflow.

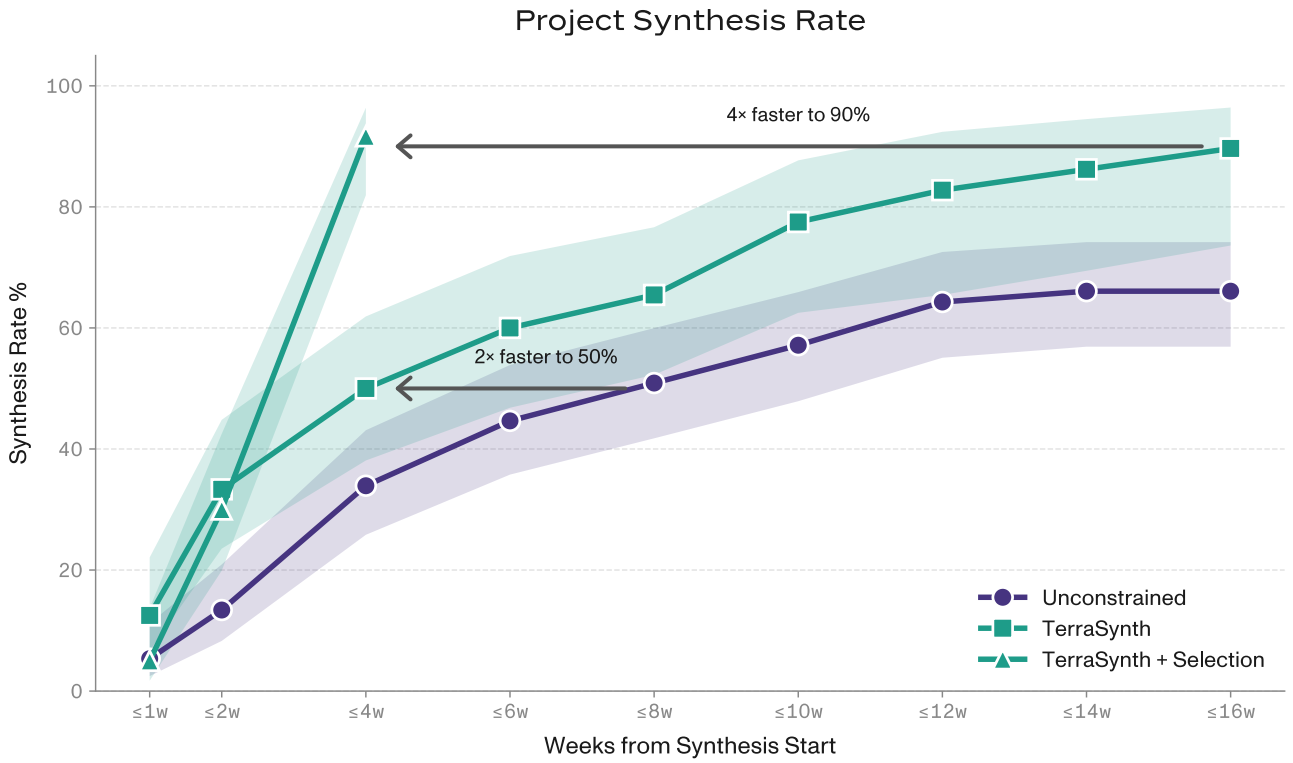

Figure 4. Cumulative synthesis rate by week from synthesis start, pooled across projects, for internal projects requiring singleton synthesis. Designs proposed by TerraSynth were synthesized up to 2-4x faster relative to unconstrained counterparts. Unconstrained: 8 projects, N=112 compounds. TerraSynth: 5 projects, N=72 compounds. TerraSynth + Selection: 4 projects, N=60 compounds. Bands are 95% Wilson binomial CIs. TerraSynth projects are assigned by date cutoff. Projects have overlapping target classes across cutoff, and comparable molecular weight + complexity.

The Data Engine Advantage

During pre-training, the design of a route sampler strongly shapes the prior chemical space the model is able to recapitulate. Compared to other works, the TerraSynth data engine samples molecules that are closer to drug-like chemistry.

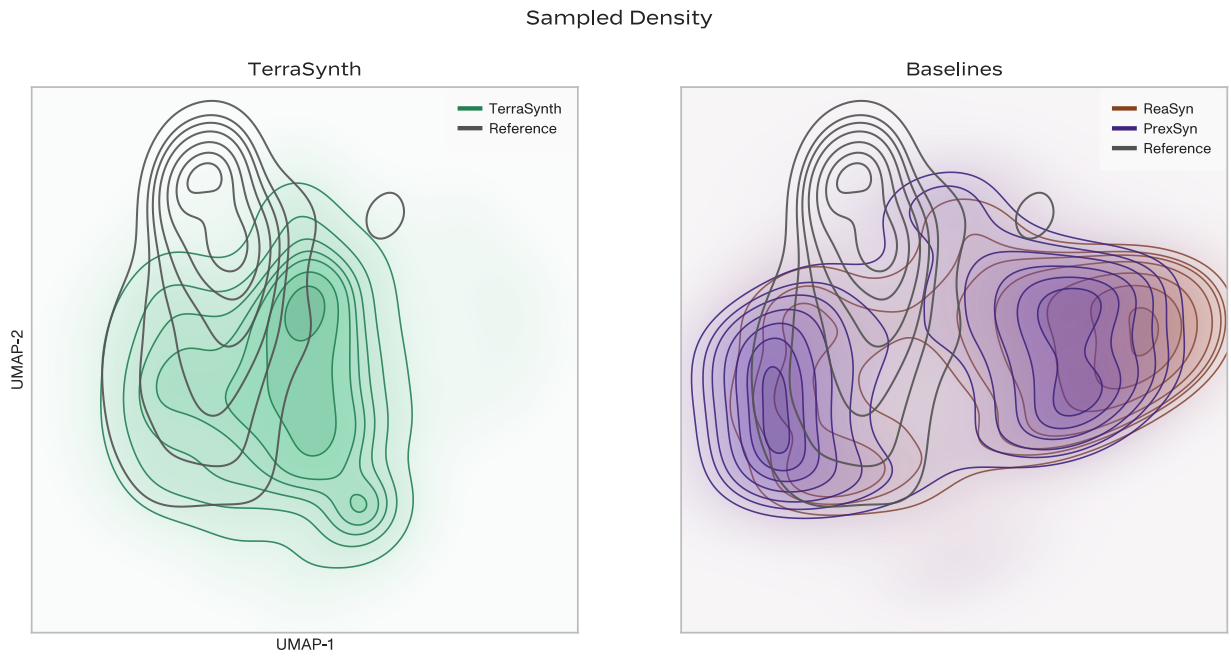

Figure 5. UMAP of COATI embeddings, with per-sampler density overlaid on samples from reference catalogs. TerraSynth's data engine has good mutual support and concentrates on the reference catalog distribution tightly. ReaSyn's data engine inherits from ChemProjector.

We can examine where the TerraSynth data engine's products actually land by embedding them with COATI, projecting to 2D with UMAP, and overlaying their density against a reference that consists of ChEMBL, Enamine REAL, and WuXi GalaXi molecules. Here, linear routes are sampled with a max-depth of 6 reactions (convergent routes may have more).

As shown above, TerraSynth's sampled molecules are closer in distribution to the reference, covering similar modes.

Reference	PrexSyn	ReaSyn / ChemProjector	TerraSynth
ChEMBL	0.028FID 10.5	0.050FID 15.5	0.023FID 9.5
Enamine REAL	0.067FID 23.8	0.095FID 29.6	0.071FID 24.3
WuXi GalaXi	0.058FID 22.0	0.071FID 24.1	0.051FID 19.3

Table 2. MMD (top, unbiased, RBF σ from median heuristic) and FID (below) of each data engine's COATI embeddings against the reference; lower is better, best per row in bold. MMD computed on 5k samples from each generator vs 5k from each reference; FID computed on 50k generated vs 10k reference.

The same gap shows up on standard physicochemical descriptors. Samples from TerraSynth's data engine have marginals that track ChEMBL and land in drug-like chemical space.

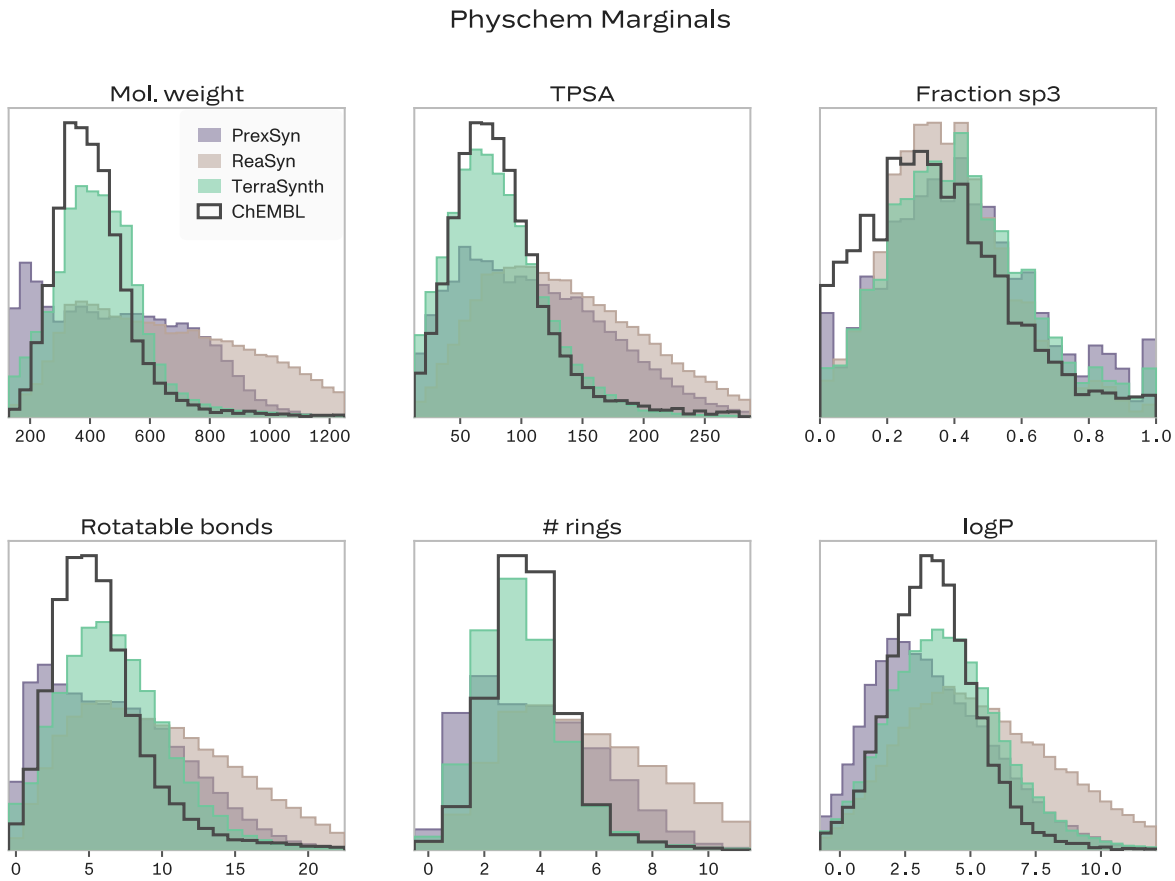

Figure 6. Physchem descriptor marginals of 10K randomly sampled ChEMBL molecules against 50K sampled molecules from the data engines of TerraSynth, PrexSyn, and ReaSyn. Distributions that overlap with ChEMBL are desirable, as it is representative of drug-like molecules.

Reasoning for Synthetic Route Realism

While exact reconstruction rate is a useful metric for assessing breadth of chemical space coverage, it does not evaluate the realism of routes sampled by synthesis planners, which is an equally important direction of development. One can take many routes to reach the same target molecule, many of which would not be feasible.

Analogous to LLM-RL for mathematical reasoning, we want to ensure that routes are consistent under the logic of synthetic chemistry. This is the next frontier for synthesis planning in the "Era of Validity"[10].

To address this, we perform RL on a reward function that penalizes some of the most common issues observed in inferred routes. These failures are coarsely categorized here either as route economy, chemoselectivity, or site-selectivity.

We also leverage the test-time gains shown previously and use a "lightly off-policy" setup where the behavior policy performs test-time search (TTS) to sample high quality routes that reconstruct the target. We then use masked importance sampling (seq-MIS) to discard trajectories whose likelihood ratios indicate excessive off-policy mismatch. This lets us absorb the quality gains from search while keeping updates close enough to the learner's distribution to remain stable.

On a random N=1,000 subset of Enamine REAL, the post-training stage reduces mean occurrence across all failure categories.

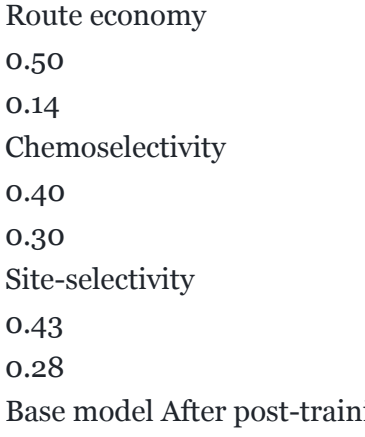

Figure 7. Mean per-route occurrences of each failure category on a random N=1,000 subset of Enamine REAL; lower is better. Post-training shortens route lengths and reduces redundant intermediate construction (route economy). Selectivity errors which depend on more subtle context like protecting-group state shrink more modestly.

The two routes below illustrate an instance of a site-selectivity failure that we address via this post-training stage. The pre-trained model (top) samples a route with an amidation step that has a site-selectivity issue. The presence of two acids on one of the reactants would produce multiple products, requiring a separation step. After post-training (bottom), the model chooses a parsimonious route that avoids producing a reactant with multiple degenerate functional groups.

Future Directions

TerraSynth was built to close the loop between molecular design and synthesis. It delivers a step change in synthetic planning performance by balancing two critical capabilities: being expressive enough to span all relevant chemical space, and sampling realistic routes that adhere to the logic of synthetic chemistry.

On the ChEMBL dataset, its reconstruction rate is 67% higher relative to the rate of the next best synthesis planner, at roughly 1/1000th of the cost. This efficiency allows it to be used inside the inner-loop of an optimizer, ensuring synthetic feasibility of every AI-generated compound. On real drug discovery projects at Terray, designs proposed this way were synthesized up to 2-4x faster relative to unconstrained counterparts.

We are actively pursuing several exciting directions to fully unlock the Solv-3[11] capability of TerraSynth, as well as post-train the model for specific design tasks. In the meantime, we plan to release a forthcoming technical report.

How to cite this

```
@misc{terraytx2026terrasynth,
  title        = {TerraSynth: Closing the Loop from Design to Synthesis},
  author       = {Miles Wang-Henderson, Zack Strater},
```

```
  year         = {2026},
  month        = {May},
  howpublished = {\url{https://www.terraytx.ai/news-and-research/terrasynth}},
}
```

Thank you to Anton Morgunov for insightful discussion on Syntax of Matter and feedback on evaluating forward synthesis planners. Thank you to Shitong Luo for insightful discussion on synthesis planners and development of PrexSyn. Thank you to Ryan Pederson for help with experiments using TerraBind. Thank you to Yoshito Takahashi for reviewing TerraSynth designs and providing feedback.

Footnotes

1. DMTA cycles are the canonical iteration loop in medicinal chemistry; in practice, the "Make" step dominates the overall cycle time.

2. Our work can be categorized as direct sequence generation, where routes are represented as a sequence of tokens. See *The Syntax of Matter. Morgunov, et al.* 2026.

3. Forward synthesis planners can both find routes and design new molecules. They run a synthetic route "forward" from its starting materials. Retrosynthesis engines take a molecule and break it up, recursing backwards until plausible starting materials are reached.

4. COATI is Terray's multimodal foundation model for small molecules, which produces joint embeddings of strings, molecular graphs, and point clouds. See Introducing EMMI: Where Experimentation Meets Machine Intelligence and our JCIM paper.

5. Some recent examples include ReaSyn, PrexSyn, and Synformer. Each is powered by a data engine that uses reaction templates combined with a building-block catalog. ReaSyn inherits the data engine from ChemProjector.

6. Representing routes as a post-fix token sequence was introduced for synthesis planning by ChemProjector (Luo et al., 2024), where each reaction token consumes the preceding building-block/intermediate tokens on the stack.

7. We say that a molecule is reconstructed or "solved" if the planner proposes a sequence of chemically plausible transformations to reach the target. The second criterion is whether the route is realistic enough, and passes checks like selectivity that make it feasible to execute in practice. In the Syntax of Matter, the first and second criteria are referred to as Solv-1 and Solv-2.

8. AiZynthFinder is an open-source retrosynthesis engine that performs MCTS over backward reaction templates; mean CPU time is ~30 sec/mol. Unlike TerraSynth, it cannot be used to generate new designs.

9. Here "solved" refers to the Solv-1 metric for retrosynthesis engines: if the backward search from the target molecule terminates at only purchasable building-blocks in the stock set.

10. Beyond brute-forcing one of many degenerate routes to a molecule, in the Era of Validity we want to ensure that those routes are chemically correct and can be executed in practice. In this work we reach Solv-2C and Solv-2R level of the hierarchy. See ischemist.com/syntax-of-matter/validity.

11. The route is experimentally viable: using practical conditions, acceptable per-step yields, workable purification methods.