

SHORT COMMUNICATION

Validation of Pharos3D by Recovery Testing in Enamine REAL Space

A structure-agnostic 3D search benchmark across approximately 100 billion accessible molecules

Modest v. Korff

Alipheron AG, Basel, Switzerland

Release Date: 2026-08-26. Correspondence to contact@alipheron.com.

Abstract

Pharos3D is a fully three-dimensional, structure-agnostic method designed to search ultra-large combinatorial chemical spaces from the spatial arrangement of pharmacophore points. Because public benchmarks tailored to pharmacophore searches in combinatorial spaces are lacking, we assessed whether Pharos3D could recover known compounds from Enamine REAL Space. Twenty-four randomly selected accessible products, each originating from a different reaction, were used as queries against the April 2026 REAL Space release (approximately 100 billion accessible molecules). Conformations were generated with the OpenChemLib ConformerGenerator and optimized with the MM94 force field. Searches used the database selection as the only search parameter, and the top 100 hits were evaluated. The exact query structure was recovered in 21 of 24 searches (87.5%). In two additional searches, the best hit had maximum-common-substructure (MCS) similarity above 0.9, yielding 23 successful tests under the prespecified criterion (96%). The unsuccessful query had an MCS similarity of 0.58 and exhibited sparse pharmacophore descriptions. These results provide an initial recovery benchmark supporting the ability of Pharos3D to locate exact or closely related molecules in an ultra-large combinatorial space.

Keywords: virtual screening; pharmacophore search; ultra-large chemical space; recovery test; Enamine REAL Space; Pharos3D

1. Introduction

Virtual-screening methods require empirical characterization to establish whether their ranking procedures can retrieve relevant molecules from large candidate collections. Pharos3D was developed for fully three-dimensional searches of ultra-large combinatorial libraries. It is structure agnostic: molecular comparison is based on shape matching and the spatial arrangement of pharmacophore points rather than on two-dimensional structural identity.

No established public test set specifically addresses pharmacophore searching across combinatorial spaces of this scale. A fundamental first challenge is therefore a recovery test: a molecule known to be accessible in the searched space is submitted as the query, and success is measured by whether that molecule, or a close structural match, is returned among the highest-ranked results. Here, we report a 24-query recovery experiment in Enamine REAL Space.

2. Materials and methods

2.1 Query selection and conformer generation

Twenty-four accessible products were selected at random from Enamine REAL Space, with each query drawn from a different reaction. The searched collection was the April 2026 release, comprising approximately 100 billion accessible molecules. Query conformations were generated using the OpenChemLib ConformerGenerator and optimized with the MM94 force field.

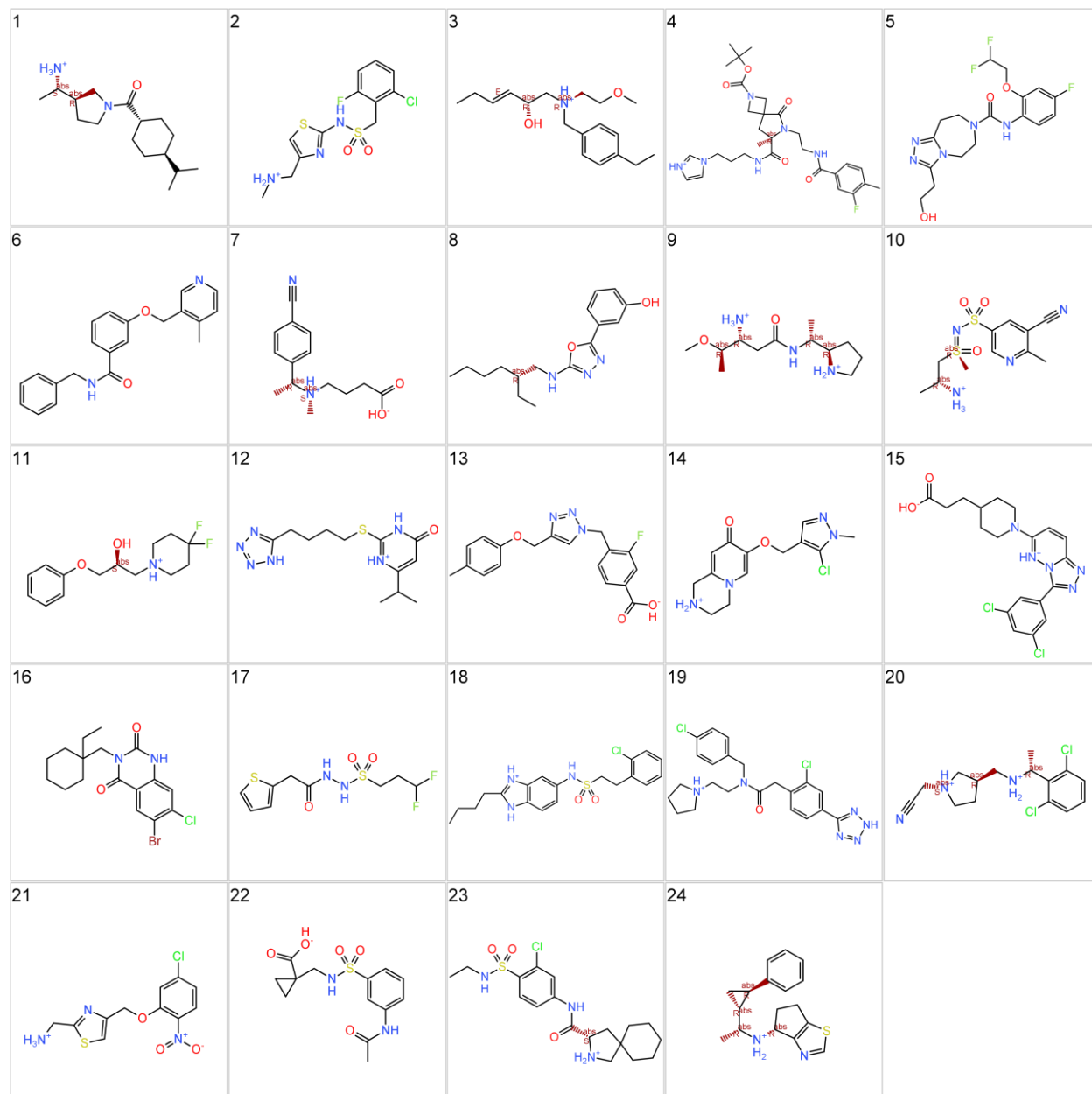


Figure 1. The 24 randomly selected Enamine REAL Space query molecules, labelled with their Pharos3D result identifiers.

2.2 Search protocol and evaluation

Each query was submitted to Pharos3D through the application programming interface. The only search parameter was selection of Enamine REAL Space. The median search time was 3.8 hours, while the mean search time was around 4.8 hours. For each query, the top 100 hits ranked by Pharos3D score were considered.

The returned structures were compared with the query using maximum-common-substructure (MCS) overlap. A test was classified as successful when the exact query occurred within the top 100. If the exact query was absent, the test was also classified as successful when the highest MCS similarity among the top 100 exceeded 0.9. The recorded fields comprised the best-MCS hit and its rank, MCS similarity, Pharos3D score at that rank, the highest Pharos3D score, and an indicator of exact query recovery.

3. Results

All 24 searches returned evaluable results. The exact query structure was present among the top 100 hits in 21 searches (87.5%). Five exact queries were ranked first (rank 0 in the result files), and five were ranked second (rank 1). Among the four searches without exact recovery, two produced close structural matches with MCS similarities of 0.93 and 0.96. Applying the prespecified success criterion, 23 of 24 recovery tests were successful (96%; Table 1).

3.1 Table column definitions

Query ID: The Pharos3D search-result identifier (the 'Name' field in the DataWarrior source file). This identifier can be used to retrieve the corresponding Pharos3D result from the server.

Best-MCS rank: The row in the Pharos3D result file at which the hit with the highest maximum-common-substructure similarity to the query was found. Rank 0 is the top-ranked row.

MCS similarity: The similarity value of the top-100 hit having the greatest maximum-common-substructure overlap with the query molecule; it corresponds to the reported best-MCS rank.

Pharos3D score at best MCS: The Pharos3D score of the hit located at the best-MCS rank.

Top Pharos3D score: The Pharos3D score at rank 0, corresponding to the highest-ranked molecule in the Pharos3D result file.

Exact query recovered: Indicates whether the query structure itself was found among the top 100 Pharos3D hits. This is the 'MetQuery' field in the DataWarrior source file.

Table 1. Complete results for all 24 Pharos3D recovery searches.

ID	Best-MCS Rank	MCS Similarity	Pharos3D Score at Best MCS	Top Pharos3D Score	Exact Query Recovered
1	1	1.00	0.87	0.87	Yes
2	6	1.00	0.90	0.92	Yes
3	66	1.00	0.79	0.83	Yes
4	0	1.00	0.91	0.91	Yes
5	17	0.93	0.90	0.93	No
6	4	1.00	0.96	0.96	Yes
7	3	1.00	0.89	0.90	Yes
8	1	1.00	0.98	0.99	Yes

ID	Best-MCS Rank	MCS Similarity	Pharos3D Score at Best MCS	Top Pharos3D Score	Exact Query Recovered
9	28	1.00	0.83	0.86	Yes
10	7	1.00	0.86	0.88	Yes
11	18	0.58	0.79	0.80	No
12	3	1.00	0.87	0.88	Yes
13	0	1.00	0.97	0.97	Yes
14	0	1.00	0.91	0.91	Yes
15	11	1.00	0.90	0.93	Yes
16	72	1.00	0.96	1.00	Yes
17	4	1.00	0.95	0.96	Yes
18	48	0.96	0.79	0.82	No
19	0	1.00	0.92	0.92	Yes
20	1	1.00	0.77	0.77	Yes
21	0	1.00	0.93	0.93	Yes
22	1	1.00	0.97	0.98	Yes
23	73	1.00	0.88	0.92	Yes
24	1	1.00	0.88	0.88	Yes

Notes: Rank 0 denotes the first row of the Pharos3D result file. The three chemical-structure columns (query, best-MCS hit, and top Pharos3D hit) remain available as machine-readable structures in the accompanying DataWarrior file; Figure 1 displays all query structures. Green rows: Exact Query Found. Yellow rows: Best hit had maximum-common-substructure (MCS) similarity above 0.9. Red row: Unsuccessful queries; MCS similarities below 0.6.

The unsuccessful case, query 11, with best MCS similarities of 0.58, was represented by only three pharmacophore points. The hydroxy and amine pharmacophore features were close in space. These characteristics reduce query specificity and plausibly made discrimination within the approximately 100-billion-molecule search space more difficult.

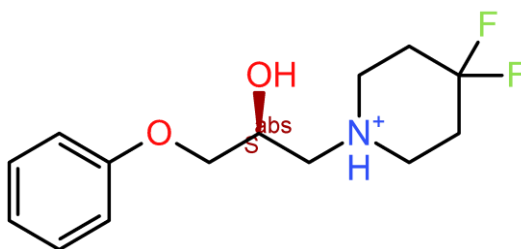


Figure 2. Query structures from the unsuccessful recovery test, described by sparse pharmacophore patterns.

4. Discussion

Recovery of the exact query in 87.5% of searches, and an exact or close match in 96%, indicates that Pharos3D can identify highly relevant molecules within an exceptionally large accessible chemical space using only three-dimensional pharmacophore information. The experiment is particularly demanding because a structure-agnostic representation may assign similar scores to molecules with different two-dimensional structures but comparable pharmacophore arrangements.

The failure also illustrates expected boundary conditions. Sparse pharmacophore patterns provide limited discriminatory information, while proximity of two features in query 11 may further reduce their effective spatial distinctiveness. These observations suggest that recovery performance depends not only on chemical-space size but also on pharmacophore richness.

This study is an initial technical validation rather than a complete comparison of virtual-screening performance. The sample comprised 24 randomly selected accessible products, success was evaluated only within the top 100 hits, and no external benchmark method, decoy set, or prospective activity endpoint was included. Larger stratified test sets should quantify performance by pharmacophore-point count, flexibility, reaction class, and molecular complexity, and should report enrichment and rank-based metrics alongside recovery.

5. Conclusion

In searches of approximately 100 billion accessible Enamine REAL Space molecules, Pharos3D recovered the exact query within the top 100 for 21 of 24 queries and returned a close structural match for two more. The resulting 96% success rate under the prespecified criterion supports Pharos3D as a practical structure-agnostic search method for ultra-large combinatorial spaces, while highlighting sparse pharmacophore descriptions as important challenge for further validation.

Data availability

The underlying 24-search result table is available in the accompanying DataWarrior file, Pharos3D_Data-QueryRecovery_2026-08-26.dwar. Pharos3D result identifiers are included to support retrieval of the corresponding server-side search results, subject to server access and retention.