

Enamine REAL Input Data (RID): Reference Guide

Document version: 1.1 | Date: October 2026 | Prepared by: Enamine Ltd.

*Reference on accessing Enamine **REAL Input Data (RID)** — the fragment-level representation of REAL Space used for synthon-based and fragment-based virtual screening.*

*This document covers the synthon-level data that REAL Space is built from. Pre-enumerated flat files of complete structures are a separate product, obtained on **different terms** — see Section 6.*

1. What is RID

[REAL \(“REadily Accessible”\)](#) is Enamine's make-on-demand chemical space of synthetically feasible compounds built from validated reactions and qualified building blocks. Synthesis success rate >80%. Typical synthesis time is 3–4 weeks.

REAL exists in two forms. The **enumerated** form is a flat list of complete product structures. The **fragment-based** form is the set of synthons plus the rules for combining them — this is **REAL Input Data (RID)**. Both describe the same chemical space. RID is what the enumerated files are generated from.

A **synthon** is a building block fragment — the part of a reagent that ends up in the product, with the leaving group removed. Synthons are not building blocks: one building block can give rise to several synthons, so the set of synthon identifiers is substantially larger than the set of building blocks it derives from. RID is therefore not the same thing as the REAL building-block list, and not the same thing as a reagent catalogue.

RID is the data layer that synthon-based virtual screening methods consume — for example **V-SYNTHES**, which docks synthon-level representations before elaborating to full products. If a workflow is described as “fragment-based screening followed by elaboration in REAL”, RID is the data it needs.

2. What RID Contains

RID is delivered in machine-readable form and comprises:

- **Synthon lists** — synthon structures with identifiers, organized by reaction and by role within that reaction.
- **Reaction rules** — delivered as reaction SMARTS, linked to the synthon lists by a shared reaction_id, defining which synthon roles combine and how.

Two points on scope that are commonly misread:

- **RID covers REAL Space (~94.5 billion structures in release 2026.1, April 2026).** It does not cover xREAL, LegacyREAL, or unREAL.
- **Coverage is uneven across reactions.** Larger reactions are sampled and filtered during data preparation rather than delivered in full, so the fraction of a reaction that reaches RID varies from one reaction to another.

Note also that Enamine does **not** provide reaction rules as classical reagents→products SMIRKS transforms. The REAL space is built on a synthon-based model, so the rules are expressed in those terms.

3. Licensing and Update Schedule

RID for REAL Space is licensed to clients **exclusively through BioSolveIT GmbH, Enamine's broker for access to RID**, which manages the license paperwork and delivers the data package. This is a data licensing arrangement, not a software purchase — no BioSolveIT software license is required to obtain RID.

Standard terms are a **one-year license**. Renewal requires a new license agreement. A license does not roll over automatically. This differs from access to the enumerated flat files, where credentials once issued remain valid for subsequent releases.

Enamine publishes two REAL releases per year, and RID packages are prepared in step with them.

Eligibility. Enamine is a Ukrainian company. RID licenses are not granted to applicants established in, or acting for parties established in, states engaged in or materially supporting the armed aggression against Ukraine. Enamine reserves the right to decline a request at its discretion.

A **sample RID package** is available for groups that would like to see how the data is structured before committing. BioSolveIT delivers it at no charge. It is not a required step, and a group that already knows what it needs from RID can go straight to a license request.

4. Practical Notes for Screening Groups

- **Synthon identifiers are versioned and not stable across releases.** A mapping built against one release should not be assumed to hold against the next.
- **Keep synthon names with your hits.** A synthon name encodes the compound, which lets Enamine resolve a hit back to one specific structure without ambiguity.
- **Stereochemistry and macrocycles are constrained in the delivered data.** One stereoisomer per chiral synthon is included, and macrocycles are excluded from the distributed set. These are constraints on the delivered data package, not on what Enamine can synthesize.
- **Outdated copies of RID may contain compounds Enamine can no longer produce.** REAL is a dynamic space — reactions and building blocks are added and retired between releases. Screening against a stale copy wastes docking effort on structures that cannot be delivered. Use the latest official release files as the structures source.
- **Third-party copies of RID may not be used.** RID is not open source. It may be used only under a valid license, and a copy obtained from anyone else is unlicensed use, even if it is current.
- **Derivatives of RID may be used internally, but not distributed.** A licensee may build things from RID and use them freely in its own workflows and research, including predictive models trained on the data. What is not permitted is publishing such a derivative, or selling or otherwise supplying it to parties that do not hold a RID license. This applies equally to models, custom chemical spaces and other derived datasets. The license text sets the position out in more detail.

5. Free Alternatives — What You May Not Need RID For

Not every fragment-based workflow requires a RID license. The following are available at no cost, and they can be requested through the customer support portal:

- **REAL building block lists** — reagent structures with Enamine IDs, without reaction-role mapping.

- **REAL Reaction Descriptions** — a human-readable document describing the REAL reactions. It is not machine-readable and is not a substitute for RID. Supplied at no cost, but proprietary: it must not be redistributed without Enamine's prior written consent.

6. How to Obtain RID — and How It Differs from Enumerated Data

RID is not free. This is the principal difference from the enumerated flat files: enumerated REAL structures are available at no cost, whereas RID requires a commercial license for a fee.

To start: **submit a request at [Enamine's customer support portal](#)**. Describe the intended workflow and the reactions or datasets of interest. Enamine will route the request and BioSolveIT will handle the quote and license agreement. If you would like to see a sample package before deciding, say so in the same request.

7. Intellectual Property — Key Points

The IP position for RID is **materially different** from that for enumerated REAL structures, and the difference matters:

- **Compound structures are free of restriction.** Enamine claims no IP over REAL compound structures. SMILES and structural data may be used, shared and redistributed without restriction.
- **RID is not.** Synthon lists and reaction rules constitute Enamine's confidential information, proprietary know-how and trade secret. RID may be used within the licensed scope. It may not be shared, redistributed or disclosed to third parties.
- **REAL Reaction Descriptions are likewise proprietary** and must not be redistributed without prior written consent, notwithstanding that the document itself is supplied at no cost.
- **All research outputs belong to you.** Hits, SAR, docking scores, ADMET predictions, experimental data and any other derived insight are yours exclusively. Enamine has no access to, claim on, or interest in them. This holds regardless of whether the work used enumerated data or RID. Research results are distinct from derivatives of RID itself, which are covered in Section 4.
- **Supply is non-exclusive** — Enamine reserves the right to synthesize any REAL compound for any client.