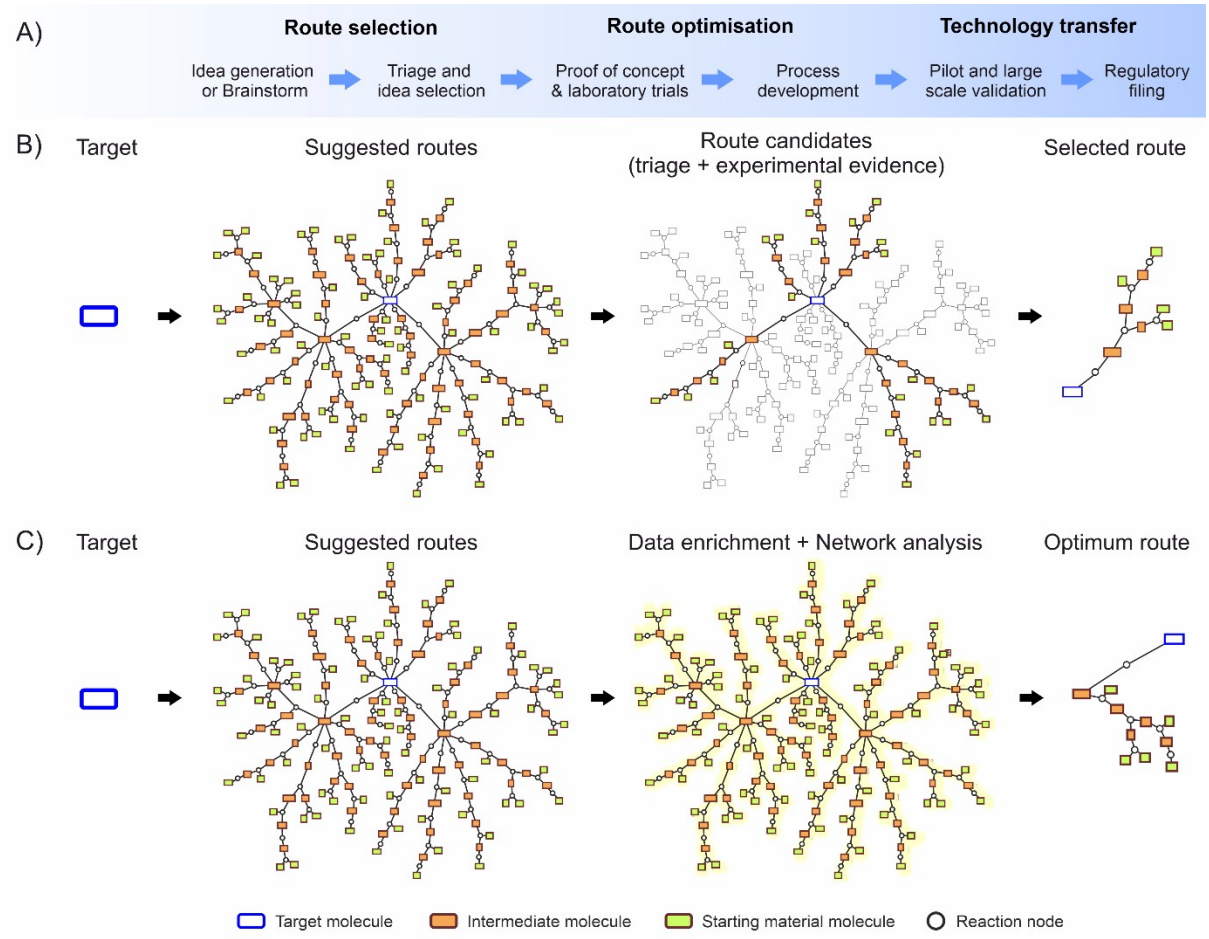


*Figure 1: (A) Traditional workflow from initial concept up to the commercial production of a pharmaceutical API; (B) The current methods usually start from the target molecule, gathering a large number of ideas (retrosynthetic analysis), narrowing this down to fewer feasible routes, up to selecting the most feasible by a panel of experts; (C) Envisioned process: a data rich approach collating many synthesis ideas from different sources, subsequently enriching the entire network with experimental and modelling data, and using an algorithmic approach to identify the optimum route for commercial synthesis.*

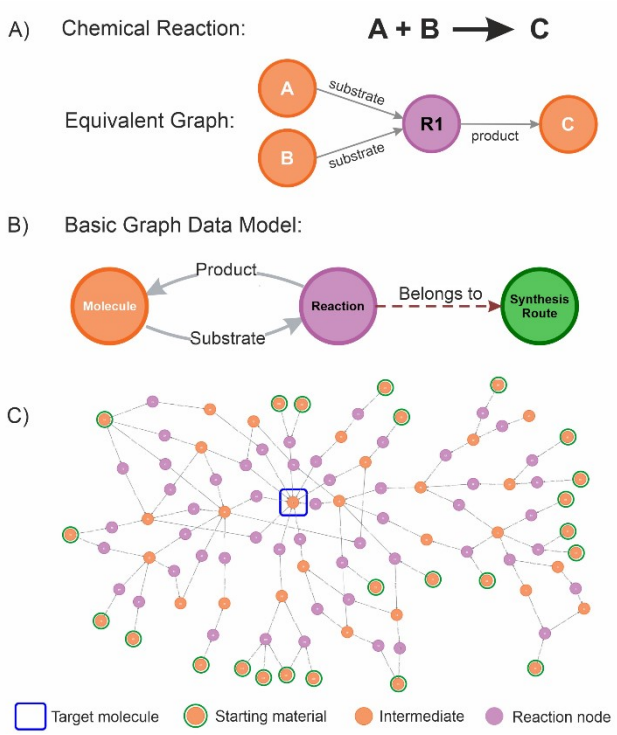


*Figure 2: A) Chemical representation of a single reaction (substrates A and B react to produce C) and its equivalent graph representation in the context of graph databases; B) Basic data model showing the relationships between molecule, reaction, and route nodes. In the graph database, nodes can be used to store specific domain information, creating a multidimensional data structure; C) Example of large network of transformations (omitting route nodes), illustrating possible paths from the starting materials to the target molecule, as observed from a graph database.*

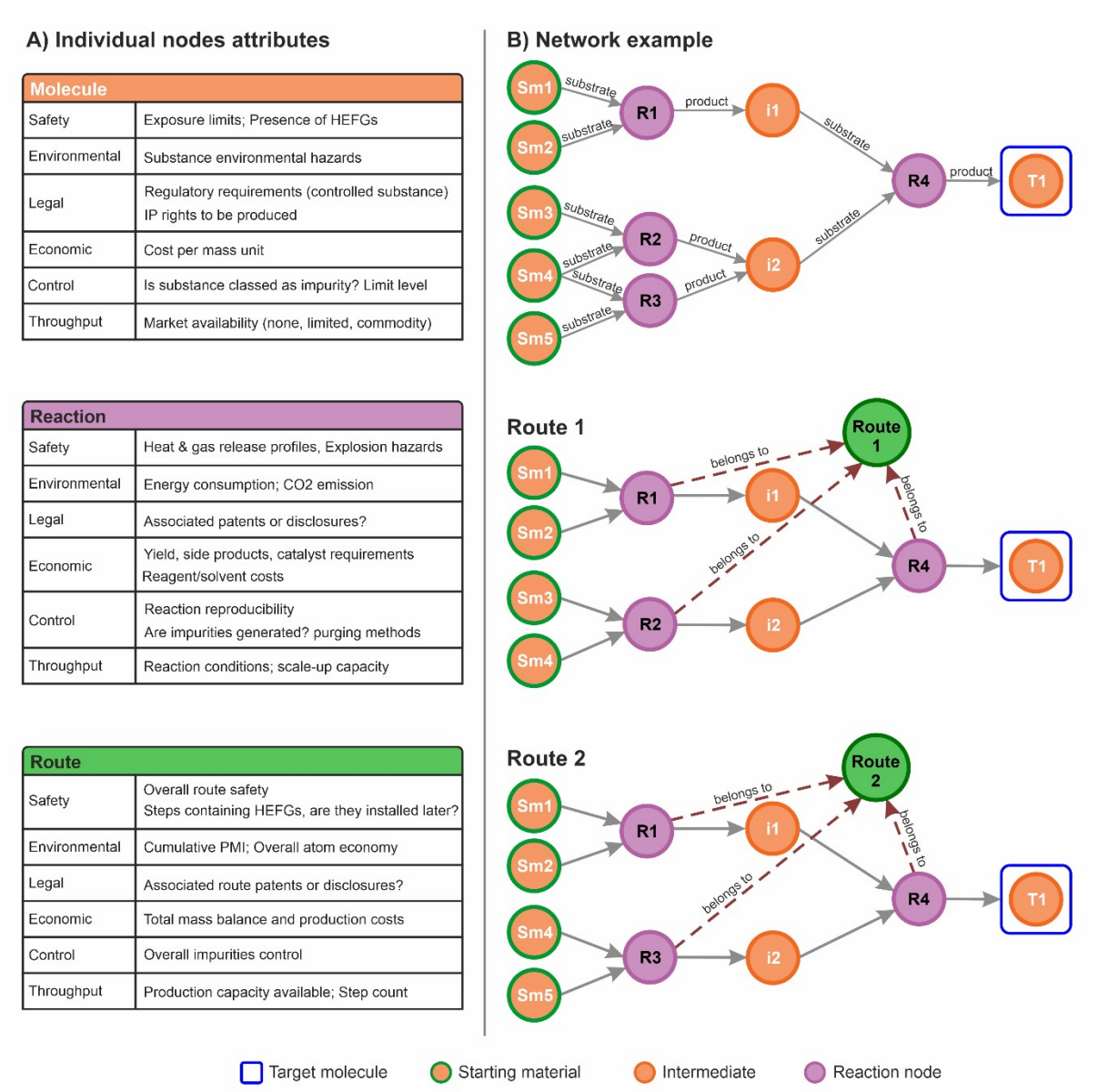


*Figure 3: A) Example of nodes attributes for molecules, reactions, and routes, aligned with the SELECT criteria. Other attributes can also be stored such as names, identifiers, etc. without interfering with the data model and other elements already stored; B) Example of a small graph database network, illustrating the route identification with two possible paths linked to a route node (green). Some edges labels (arrows) were removed for clarity.*

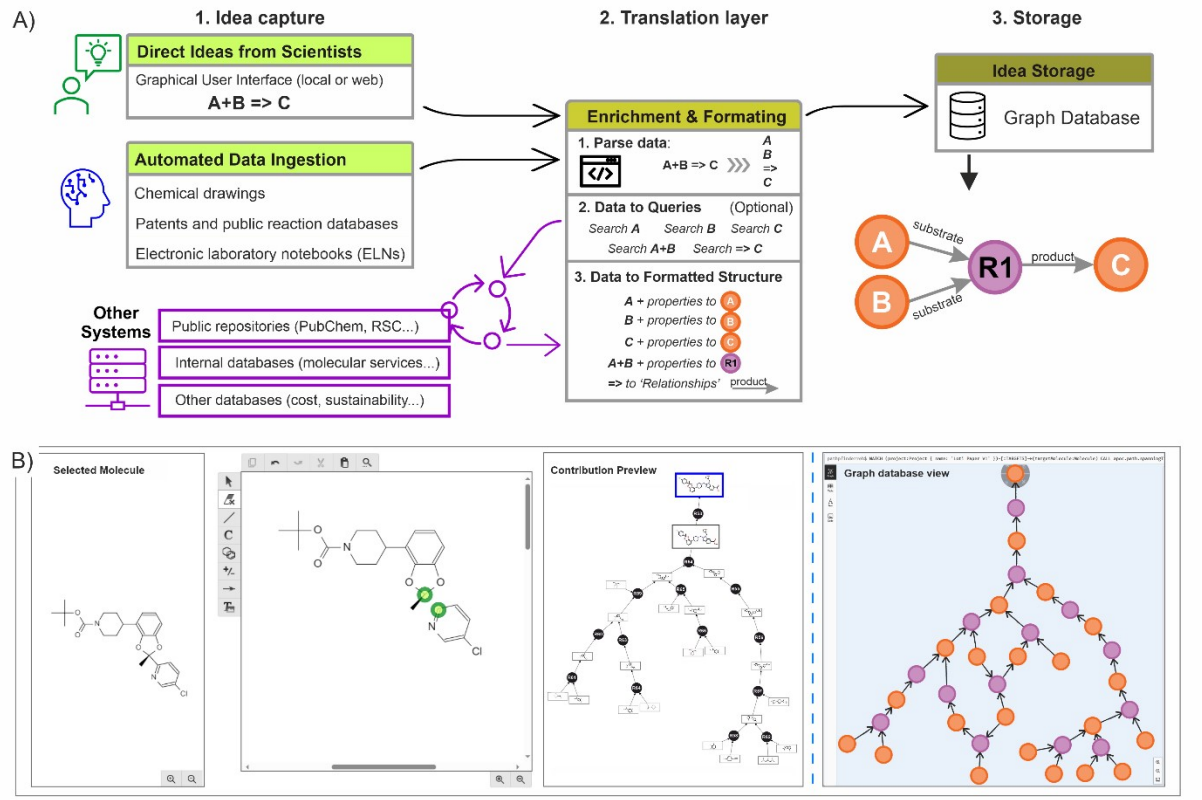


*Figure 4: A) Idea capture process starts with ideas obtained directly from the scientist or ingested from other data systems. Next, a translation layer parses and query the data in other databases to capture additional metadata (enrichment), with the results categorised to fit the data model. Finally, the data is queried for consistency into the graph database before being stored permanently. B) Example of graphical user interface (GUI), with an embedded drawing canvas to introduce chemical ideas directly from the scientist (left). The resulting fragments are extracted and queried programmatically into the graph database, building a dynamic graph visible to the scientist (contribution preview). Once the idea is submitted, the data is registered into the graph database (Neo4J backend representation on right side).*

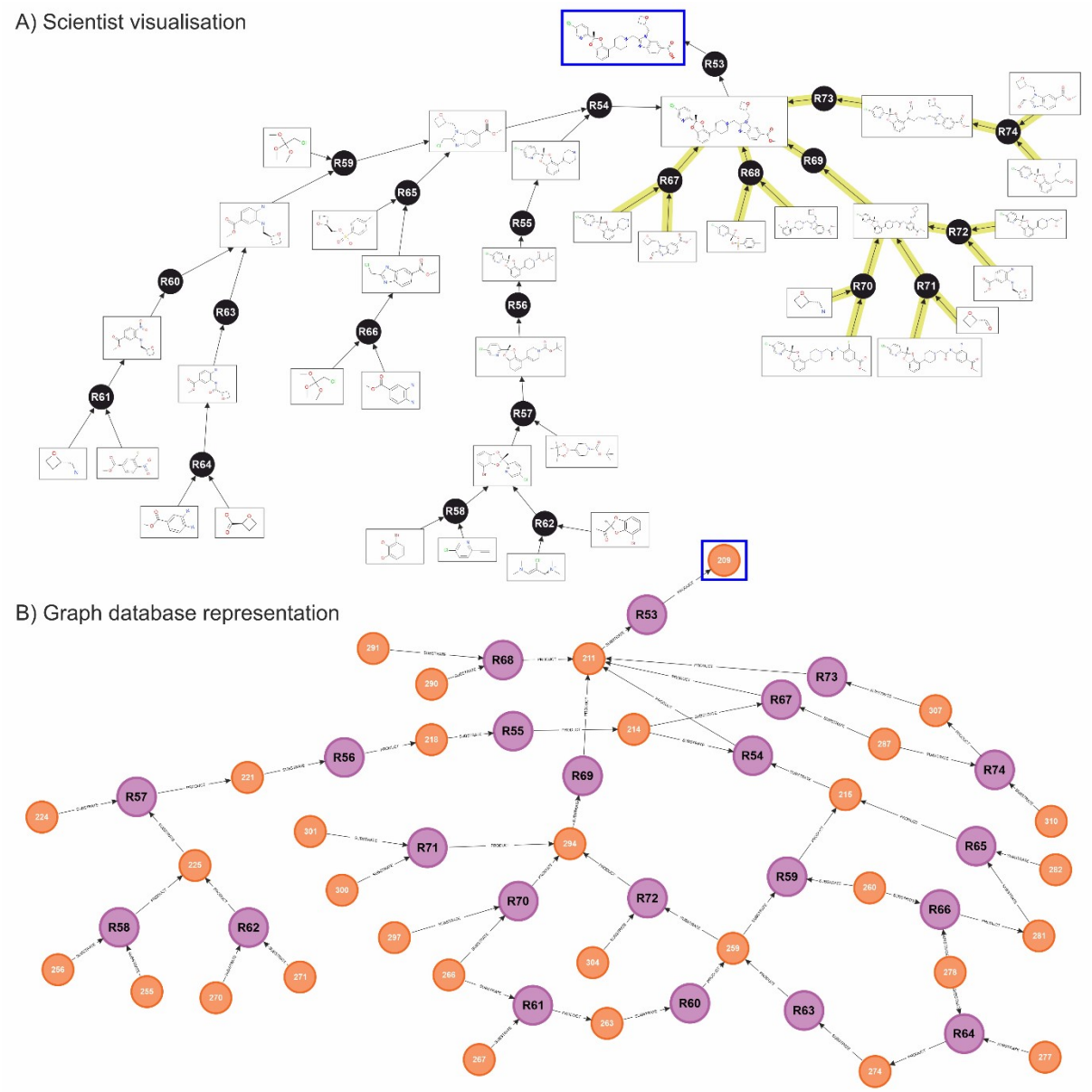


*Figure 5: GLP1 (Lotiglipron) Network – A) scientist view, combining human ideas (left side, no overlaying colour), with synthetic ideas from predictive software ASKCOS (right side, yellow overlayed). The resulting network observed from the scientist interface shows linear branches with duplication of substrates and products for user clarity. B) Equivalent back-end graph database representation obtained from native graph database (Neo4J), showing unique molecule nodes. In both cases, the target molecule is indicated by a square.*

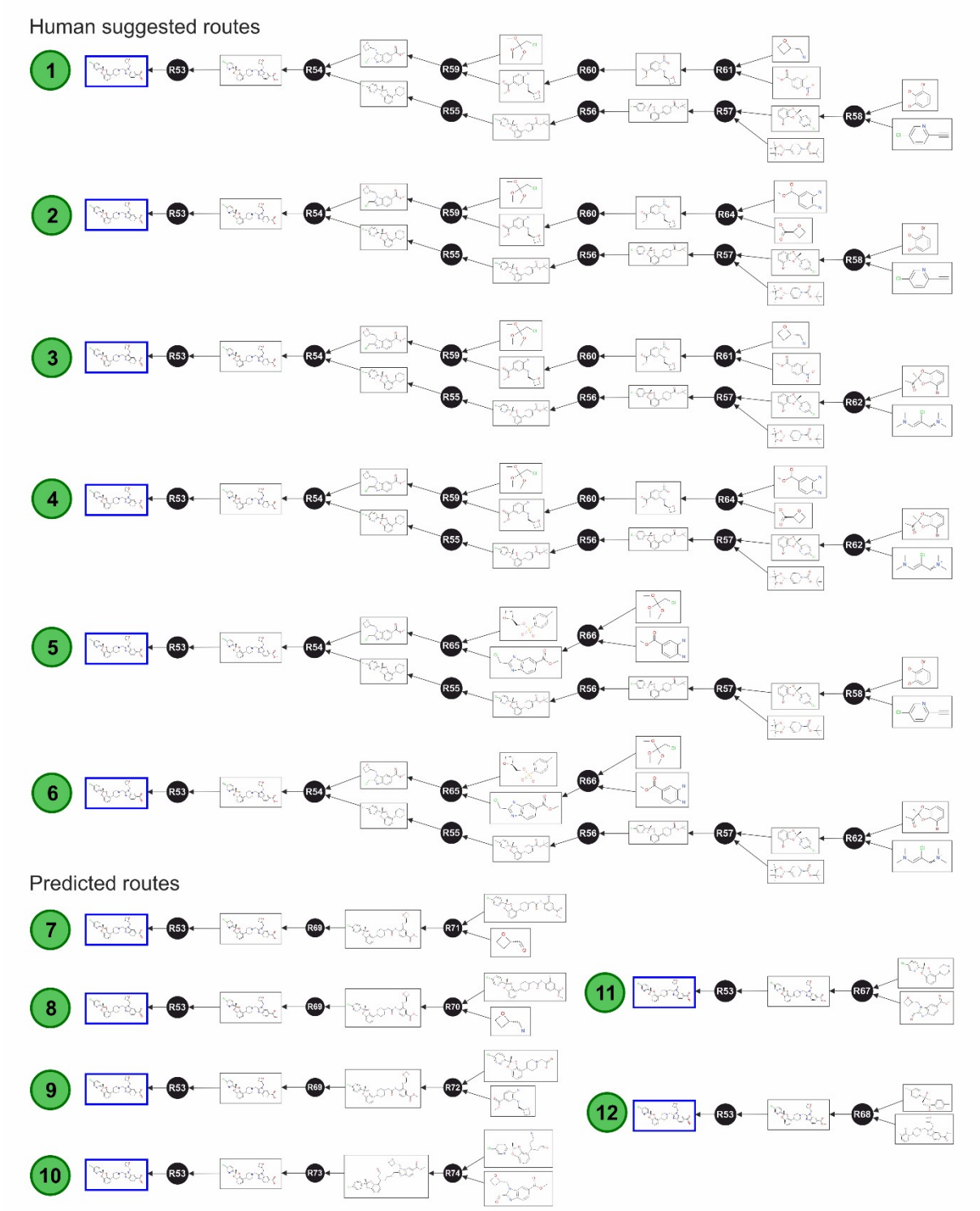


*Figure 6: GLP1 (Lotiglipron) Network – automated route identification algorithm returned twelve routes: 1-6 human generated; 7-12 from predictive algorithm. Route 1 corresponds to the enabling chemistry route. The shortest path query implemented without any constraint returned routes 11 and 12. These two transformations are possible but without real commercial value since the routes are partially developed (the starting materials cannot be purchased).*

*Overlaying additional data, including layers aligned to the SELECT criteria, can help to determine an optimum commercial route.*