Predicting Novel Metabolic Pathways through Subgraph Mining

Supplementary Material

Aravind Sankar
Dept. of CSE
IIT Madras
Chennai, India 600036
aravindsankar28@gmail.com

Sayan Ranu[†][*]
Dept. of CSE
IIT Madras
Chennai, India 600036
sayanranu@cse.iitd.ac.in

Karthik Raman[†]
Dept. of Biotechnology, Bhupat and
Jyoti Mehta School of Biosciences
IIT Madras
Chennai, India 600036
kraman@iitm.ac.in

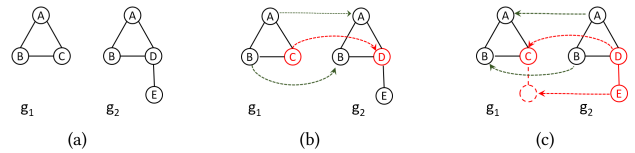


Figure 1: Illustration of subgraph edit distance. (a) Two sample graphs, g_1 and g_2. (b) and (c) illustrate the mappings corresponding to $sed(g_1, g_2)$ and $sed(g_2, g_1)$, respectively. Dashed arrows indicate the mapping between different vertices, while dashed lines denote the dummy vertices and edges. A mapping in red indicates a mapping between vertices (or edges) of unequal labels.

A SUPPLEMENTARY METHODS

A.1 Graph Isomorphism

Definition A.1. Graph Isomorphism. Graph $g(V, E)$ is isomorphic to $g'(V', E')$ if there exists a bijection ϕ such that for every vertex $v \in V$, $\phi(v) \in V'$ and $l(v) = l(\phi(v))$, and for every edge $e = (v_1, v_2) \in E$, $\phi(e) = (\phi(v_1), \phi(v_2)) \in E'$, and $l(e) = l(\phi(e))$.

The concept of *subgraph isomorphism* is defined analogously by using an injection instead of a bijection. We use the notation $s \subseteq g$ to denote the relationship that s is subgraph isomorphic to g.

A.2 Reactant–Product Mapping

A.2.1 Subgraph Edit Distance. **Illustration of sed and RPM.** Let us consider the graphs in Fig 1a. Intuitively, $sed(g_1, g_2)$ should be 1 since g_1 can be converted to a subgraph of g_2 (the triangle *ABD*) by changing the label of vertex *C* in g_1 to *D*. One possible mapping from g_1 to g_2 for $sed(g_1, g_2)$ is shown in Fig 1b. From the definition of subgraph mapping, it is clear that the operation is asymmetric. More specifically, Fig 1c shows one possible mapping from g_2 to g_1 for $sed(g_2, g_1)$. The mapping steps are illustrated in Figs. 1b and 1c.

We now define the formal approach to compute $sed(g, g')$ for any two graphs $g(V, E)$ and $g'(V', E')$. If $|V'| < |V|$ or $|E'| < |E|$, then we extend g' by *dummy vertices* or *dummy edges* such that both graphs are of equal sizes. Specifically, we create a graph $g'^*(V'^*, E'^*)$ where $|V| = |V'^*|$, $|E| = |E'^*|$. A dummy vertex or edge has the label ϵ. Adding dummy vertices and edges to g' when it is smaller than g allows us to define an injection from g to g'^* and construct a *subgraph mapping*.

To illustrate, let us revisit Fig. 1a. If we need to compute $sed(g_1, g_2)$, we do not need to add any dummy vertices or edges to g_1 since it is smaller than g_2. On the other hand, if we are to compute $sed(g_2, g_1)$ then dummy vertices need to be added to g_1. The dashed vertex and edge in Fig. 1c show the dummy additions for $sed(g_2, g_1)$.

Definition A.2. Subgraph Mapping. A mapping ϕ between graphs g and g' is an injection $g \longrightarrow g'^*$ where $\forall v \in V, \phi(v) \in V'^*$ and $\forall e = (v_1, v_2) \in E, \phi(e) = (\phi(v_1), \phi(v_2)) \in E'^*$.

One possible mapping from g_1 to g_2 is shown in Fig 1b. From the definition of subgraph mapping, it is clear that the operation is asymmetric. More specifically, Fig 1c shows one possible mapping from g_2 to g_1. Since g_2 contains more edges and vertices, it is necessary to add dummy vertices and edges to g_1.

Definition A.3. Subgraph edit distance under ϕ. The distance $sed_\phi(g, g')$ with respect to mapping ϕ is as follows:

$$sed_\phi(g, g') = \sum_{\forall v \in V} d(v, \phi(v)) + \sum_{\forall e \in E} d(e, \phi(e)) \qquad (1)$$

where $d(v, \phi(v)) = 0$ if their labels are identical, i.e., $l(v) = l(\phi(v))$. Otherwise, $d(v, \phi(v)) = 1$. $d(e, \phi(e))$ is defined analogously.

Since subgraph mapping is asymmetric, $sed_\phi(g, g')$ is asymmetric as well. For the mapping in Figure 1b, $sed_\phi(g_1, g_2) = 1$. On the other hand, $sed_\phi(g_1, g_2) = 3$ (Fig 1c). The mappings between vertices and edges of unequal labels are highlighted in red. Each of these red mappings incur a cost of 1.

Definition A.4. Subgraph Edit Distance. The *subgraph edit distance* $sed(g, g')$ is the minimum distance under all possible mappings. Mathematically,

$$sed(g, g') = \min_{\forall \phi} \{sed_\phi(g, g')\} \qquad (2)$$

$sed(g, g')$ is asymmetric. For example, $sed(g_1, g_2) = 1$ since the mapping in Example 1b minimises the distance. Similarly, $sed(g_2, g_1) = 3$.

Algorithm 1 presents the pseudocode to perform RPM using *sed*. The algorithm proceeds in a greedy manner: first, we identify the pair in a reaction $\mathcal{R}$ that minimises the following function (line 4).

[*] Present address: Department of Computer Science and Engineering, IIT Delhi, New Delhi – 110016, India

[†] Corresponding author.

```
1: $matchedPairs \leftarrow \emptyset$
2: Discard simple molecules from $PS(\mathcal{R})$
3: while $PS(\mathcal{R}) \neq \emptyset$ do
4:     $(A, B) \leftarrow \min_{\forall A \in RS(\mathcal{R}), B \in PS(\mathcal{R})} \{min\{sed(A, B), sed(B, A)\}\}$
5:     $matchedPairs \leftarrow matchedPairs \cup (A, B)$
6:     $PS(\mathcal{R}) \leftarrow PS(\mathcal{R}) - B$
7: return $matchedPairs$
```

Algorithm 1: *RPM*($\mathcal{R}$). The algorithm for computing RPM for a given reaction, $\mathcal{R}$.

$$(A, B) = \min_{\forall A \in RS(\mathcal{R}), B \in PS(\mathcal{R})} \{min\{sed(A, B), sed(B, A)\}\} \quad (3)$$

More simply, we choose the pair that matches best. Since $sed(g, g')$ is asymmetric, we explore mapping in both directions and choose the one that minimises the distance. In case of a tie, we choose the pair that is closer in size. We assign the (A, B) pair as matched (line 5) and then retrieve the next best pair containing an unmatched product (line 6). We continue this iteration till all products are matched (line 3). Notice that RPM may not necessarily be one-to-one. In a decomposition reaction $AB \longrightarrow A + B$, we would have one-to-many mappings of $RPM(AB, A)$ and $RPM(AB, B)$. Similarly, many-to-one mappings are possible when two reactants combine to form a single product. For practical purposes, we do not match molecules such as water, oxygen, ammonia, etc. even if they appear as products since they cannot be used as primary reactants in a pathway (line 2). The complete list of unmatched metabolites is given in S1 Table.

Note that Algorithm 1 iterates till all products are matched and hence it is possible for some reactants to remain unmatched. This does not hurt our ultimate goal of predicting pathways. Any target molecule that we want to synthesize would be a product of some reaction. Thus, we only need to store structural changes corresponding to products.

Revisiting Fig 1a (main manuscript), it is easy to see that ethanal and propanal would get mapped to ethanol and propanol. A slightly more complex reaction is shown in Fig 3a (main manuscript). We refer to each molecule by their KEGG compound IDs (CIDs) shown in the image. There are three products in this reaction, out of which Ammonia (C00014) is discarded. Among the remaining two, C00002 matches equally well with C00020 and C00013, with a distance of 1. However, since it is closer in size to C00020, we pick the pair (C00020,C00002). C00049 matches best with C00152 with a distance of 1. Thus, the pair (C00049,C00152) is added and the RPM process completes, since there are no more products left to match.

Reactant	Product	sed	Reactant	Product	sed
C00020	C00002	1	C00152	C00049	1
C00013	C00002	1	C00013	C00049	>1
C00152	C00002	>1	C00020	C00049	>1

A.3 Reaction Centres

The reaction centre for a pair (A, B) is the set of vertices in the product B to which new edges are added or existing edges are removed during its transformation from A. The reaction centre can easily be determined from the mapping ϕ corresponding to $sed(A, B)$. Specifically, it is a vertex v in the product B, such that $l(v) = l(\phi(v))$, but there exists an edge (v, v'), where $l(v') \neq l(\phi(v'))$. Recall, $l(v)$ denotes the label of v.

A.4 Computing Reaction Signatures

The reaction signatures involve computations of added and removed subgraphs, as we explain below.

A.4.1 *Addition of subgraph.*

This contains the subgraph that got added to the product during the reaction. For example, in the $(C00152, C00049)$ pair, OH gets added to C00049.

Formally, this added subgraph D can be computed using the mapping function ϕ. A vertex $v \in V_B$ is also in the added subgraph $D(V_D, E_D)$ if it satisfies one of the following conditions:

(1) If ϕ is from A to B, either $\nexists v' \in V_A$, such that $\phi(v') = v$, or $\exists v' \in V_A$, such that $\phi(v') = v$ and $l(\phi(v')) \neq l(v)$
(2) If ϕ is from B to A, either $\nexists v' \in V_A$, such that $\phi(v) = v'$, or $\exists v' \in V_A$, such that $\phi(v) = v'$ and $l(\phi(v)) \neq l(v')$
(3) $v \in V_c$, V_c is the set of reaction centres

We include the reaction centres in this subgraph since it will contain at least one connecting edge that got added. The edge set $E_D = \{e = (v_1, v_2) \in E_B | \; v_1, v_2 \in V_D\}$.

A.4.2 *Removal of subgraph.*

This information encodes all subgraphs that got removed from the reactant. For example, in C00152, NH_2 gets removed. We compare the structures of the product and the reactant using the mapping ϕ and compute the subgraph $R(V_R, E_R)$ that was removed. A vertex $v \in V_A$ is also in V_R if it satisfies one of the following conditions:

(1) If ϕ is from A to B, either $\nexists v' \in V_B$, such that $\phi(v) = v'$, or $\exists v' \in V_B$, such that $\phi(v) = v'$ and $l(\phi(v)) \neq l(v')$
(2) If ϕ is from B to A, either $\nexists v' \in V_A$, such that $\phi(v') = v$, or $\exists v' \in V_A$, such that $\phi(v') = v$ and $l(\phi(v')) \neq l(v)$
(3) $v \in V_c$, V_c is the set of reaction centres

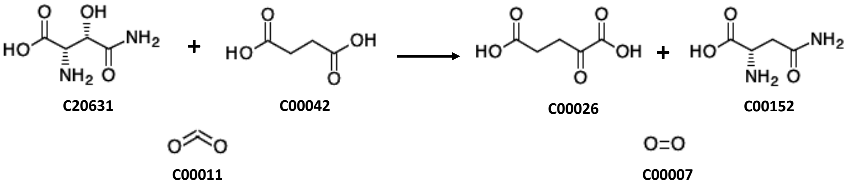
The edge set $E_R = \{e = (v_1, v_2) \in E_A | \; v_1, v_2 \in V_R\}$.

A.5 Extending the KEGG Dataset

We expanded an initial seed set of 10,065 known KEGG biochemical reactions to form a synthetic set of 150,000 reactions, to examine the scalability of our approach. We first extract the reaction product pairs from our basic dataset. In a given pair, we randomly replace one or more hydrogen atoms with different functional groups to create multiple new pairs. We aggregate all these pairs to obtain our expanded synthetic compound and reaction databases. The reaction rules are finally mined on this synthetic dataset. The final reaction database contained a total of 188,604 unique molecules.

A.6 Metabolites unmatched in RPM

Table S1 provides a list of inorganic metabolites unmatched in RPM. These are small metabolites that routinely occur in reactions but are not important in the context of the *main backbone transformation* happening in a biosynthetic pathway.



Reactant	Product	sed
C20631	C00152	0
C00042	C00152	2
C20631	C00026	6
C00042	C00026	2

Figure 2: Another example reaction to illustrate RPM using subgraph edit distance. The final matched pairs are highlighted blue.

S1 Table. A list of metabolites unmatched in RPM. The table provides a list of (mostly) inorganic metabolites unmatched in RPM. These are small metabolites that routinely occur in reactions but are not important in the context of the *main backbone transformation* happening in a biosynthetic pathway.

KEGG CID	Metabolite Name
C00001	H_2O
C00007	Oxygen
C00009	Orthophosphate
C00010	CoA
C00011	CO_2
C00013	Diphosphate
C00014	Ammonia
C00027	H_2O_2
C00080	H^+
C00237	CO
C01327	HCl

A.7 Pre-processing of MOL files

Each `.mol` file in the compound database was processed using OpenBabel [1] to obtain the graph structure. As explained earlier, the atoms and bonds constitute the vertices and edges respectively and the edge labels are defined using the bond order and stereochemistry information available.

A.8 Details on the different pathways used

We consider shorter sub-pathways in cases the original pathway is long, such as glycolysis. The RPM identified by our algorithm is indicated by underlining the matching metabolites.

S2 Table. Details on the different pathways used. We consider shorter sub-pathways in cases the original pathway is long, such as glycolysis. The RPM identified by our algorithm is indicated by underlining the matching metabolites.

Pathway 1: Glycolysis sub-pathway 1, α-D-Glucose to D-Glyceraldehyde	
C00267 ⟶ C00668 ⟶ C05345 ⟶ C05378 ⟶ C00118	
R01786	C00002 + $\underline{\text{C00267}}$ ⇔ C00008 + $\underline{\text{C00668}}$ or
or **R02189**	C00404 + $\underline{\text{C00267}}$ ⇔ C00404 + $\underline{\text{C00668}}$ or
or **R09085**	$\underline{\text{C00267}}$ + C00008 ⇔ $\underline{\text{C00668}}$ + C00020
R02740	$\underline{\text{C00668}}$ ⇔ $\underline{\text{C05345}}$
R04779	C00002 + $\underline{\text{C05345}}$ ⇔ C00008 + $\underline{\text{C05378}}$ or
or **R09084**	$\underline{\text{C05345}}$ + C00008 ⇔ $\underline{\text{C05378}}$ + C00020
R01070	$\underline{\text{C05378}}$ ⇔ C00111 + $\underline{\text{C00118}}$
Pathway 2: Glycolysis sub-pathway 2, D-Glyceraldehyde to Pyruvate	
C00118 ⟶ C00197 ⟶ C00631 ⟶ C00074 ⟶ C00022	
R07159	$\underline{\text{C00118}}$ + C00001 + 2 C00139 ⇔ $\underline{\text{C00197}}$ + 2 C00080 + 2 C00138
R01518	$\underline{\text{C00197}}$ ⇔ $\underline{\text{C00631}}$
R00658	$\underline{\text{C00631}}$ ⇔ $\underline{\text{C00074}}$ + C00001
R00200	C00008 + $\underline{\text{C00074}}$ ⇔ C00002 + $\underline{\text{C00022}}$
Pathway 3: L-Histidine Biosynthesis full pathway, from 5-Phospho-α-D-ribose	
C00119 ⟶ C02739 ⟶ C02741 ⟶ C04896 ⟶ C04916 ⟶ C04666 ⟶ C01267 ⟶ C01100 ⟶ C00860 ⟶ C00135	
R01071	C00002 + $\underline{\text{C00119}}$ ⇔ $\underline{\text{C02739}}$ + C00013
R04035	$\underline{\text{C02739}}$ + C00001 ⇔ $\underline{\text{C02741}}$ + C00013
R04037	$\underline{\text{C02741}}$ + C00001 ⇔ $\underline{\text{C04896}}$
R04640	$\underline{\text{C04896}}$ ⇔ $\underline{\text{C04916}}$
R04558	$\underline{\text{C04916}}$ + C00064 ⇔ $\underline{\text{C04666}}$ + C04677 + C00025
R03457	$\underline{\text{C04666}}$ ⇔ $\underline{\text{C01267}}$ + C00001
R03243	$\underline{\text{C01267}}$ + C00025 ⇔ $\underline{\text{C01100}}$ + C00026
R03013	$\underline{\text{C01100}}$ + C00001 ⇔ $\underline{\text{C00860}}$ + C00009
R01158	$\underline{\text{C00860}}$ + 2 C00003 + C00001 ⇔ $\underline{\text{C00135}}$ + 2 C00004 + 2 C00080
Pathway 4: L-Histidine Biosynthesis sub-pathway 1	
C04916 ⟶ C04666 ⟶ C01267 ⟶ C01100 ⟶ C00860 ⟶ C00135	
R04558	$\underline{\text{C04916}}$ + C00064 ⇔ $\underline{\text{C04666}}$ + C04677 + C00025
R03457	$\underline{\text{C04666}}$ ⇔ $\underline{\text{C01267}}$ + C00001
R03243	$\underline{\text{C01267}}$ + C00025 ⇔ $\underline{\text{C01100}}$ + C00026
R03013	$\underline{\text{C01100}}$ + C00001 ⇔ $\underline{\text{C00860}}$ + C00009
R01158	$\underline{\text{C00860}}$ + 2 C00003 + C00001 ⇔ $\underline{\text{C00135}}$ + 2 C00004 + 2 C00080
Pathway 5: D-Galacturonate degradation to Pyruvate	
C00333 ⟶ C00558 ⟶ C00817 ⟶ C00204 ⟶ C04442 ⟶ C00022	
R01983	$\underline{\text{C00333}}$ ⇔ $\underline{\text{C00558}}$
R02555	$\underline{\text{C00558}}$ + C00004 + C00080 ⇔ $\underline{\text{C00817}}$ + C00003
R01540	$\underline{\text{C00817}}$ ⇔ $\underline{\text{C00204}}$ + C00001
R01541	C00002 + $\underline{\text{C00204}}$ ⇔ C00008 + $\underline{\text{C04442}}$
R05605	$\underline{\text{C04442}}$ ⇔ $\underline{\text{C00022}}$ + C00118

(S2 Table contd.) Details on the different pathways used.

Pathway 6: Pyridoxal biosynthesis, D-Erythrose to Pyridoxal phosphate

C00279 ⟶ C03393 ⟶ C06054 ⟶ C06055 ⟶ C07335 ⟶ C11638 ⟶ C00627 ⟶ C00018

R01825	C00279 + C00003 + C00001 ⇔ C03393 + C00004 + C00080
R04210	C03393 + C00003 ⇔ C06054 + C00004 + C00080
R05085	C06054 + C00025 ⇔ C06055 + C00026
R05681	C06055 + C00003 ⇔ C07335 + C00004 + C00080
R07406	C07335 ⇔ C11638 + C00011
R05838	C11638 + C11437 ⇔ C00627 + C00009 + 2 C00001
R00278	C00627 + C00007 ⇔ C00027 + C00018

Pathway 7: L-Threonine to L-Isoleucine

C00188 ⟶ C00109 ⟶ C06006 ⟶ C14463 ⟶ C06007 ⟶ C00671 ⟶ C00407

R00996	C00188 ⇔ C00109 + C00014
R08648	C00022 + C00109 ⇔ C06006 + C00011
R05069	C06006 ⇔ C14463
R05068	C14463 + C00005 + C00080 ⇔ C06007 + C00006
R05070	C06007 ⇔ C00671 + C00001
R02199	C00671 + C00025 ⇔ C00407 + C00026

Pathway 8: Tetrahydrofolate biosynthesis sub-pathway 1, GTP to 7,8-dihydropteridine

C00044 ⟶ C05922 ⟶ C05923 ⟶ C06148 ⟶ C04895 ⟶ C04874

R00428	C00044 + C00001 ⇔ C05922
R05046	C05922 + C00001 ⇔ C05923 + C00058
R05048	C05923 ⇔ C06148
R04639	C06148 ⇔ C04895 + C00001
R04620	C04895 + 3 C00001 ⇔ C04874 + 3 C00009

Pathway 9: Tetrahydrofolate biosynthesis sub-pathway 2, 7,8-Dihydroneopterin triphosphate to Dihydrofolate

C04895 ⟶ C04874 ⟶ C01300 ⟶ C04807 ⟶ C00921 ⟶ C00415

R04620	C04895 + 3 C00001 ⇔ C04874 + 3 C00009
R03504	C04874 ⇔ C00266 + C01300
R03503	C00002 + C01300 ⇔ C00020 + C04807
R03067	C04807 + C00568 ⇔ C00013 + C00921
R02237	C00002 + C00921 + C00025 ⇔ C00008 + C00009 + C00415

Pathway 10: L-Aspartate to 2,3,4,5-Tetrahydrodipicolinate, part of Lysine biosynthesis

C00049 ⟶ C03082 ⟶ C00441 ⟶ C20258 ⟶ C03972

R00480	C00002 + C00049 ⇔ C00008 + C03082
R02291	C03082 + C00005 + C00080 ⇔ C00441 + C00009 + C00006
R10147	C00441 + C00022 ⇔ C20258 + C00001
R04198	C20258 + C00004 + C00080 ⇔ C03972 + C00003 + C00001
or **R04199**	C20258 + C00005 + C00080 ⇔ C03972 + C00006 + C00001

(S2 Table contd.) Details on the different pathways used.

Pathway 11: Threonine biosynthesis, L-Aspartate to L-Threonine

C00049 ⟶ C03082 ⟶ C00441 ⟶ C00263 ⟶ C01102 ⟶ C00188

R00480	C00002 + $\underline{\text{C00049}}$ ⇔ C00008 + $\underline{\text{C03082}}$
R02291	$\underline{\text{C03082}}$ + C00005 + C00080 ⇔ $\underline{\text{C00441}}$ + C00009 + C00006
R01773	$\underline{\text{C00441}}$ + C00004 + C00080 ⇔ $\underline{\text{C00263}}$ + C00003
or **R01775**	$\underline{\text{C00441}}$ + C00005 + C00080 ⇔ $\underline{\text{C00263}}$ + C00006
R01771	C00002 + $\underline{\text{C00263}}$ ⇔ C00008 + $\underline{\text{C01102}}$
R01466	$\underline{\text{C01102}}$ + C00001 ⇔ $\underline{\text{C00188}}$ + C00009

Pathway 12: Oxaloacetate to L-Glutamate

C00036 ⟶ C00158 ⟶ C00311 ⟶ C00026 ⟶ C00025

R00351	C00024 + C00001 + $\underline{\text{C00036}}$ ⇔ $\underline{\text{C00158}}$ + C00010
R01324	$\underline{\text{C00158}}$ ⇔ $\underline{\text{C00311}}$
R00267	$\underline{\text{C00311}}$ + C00006 ⇔ $\underline{\text{C00026}}$ + C00011 + C00005 + C00080
R00355	C00049 + $\underline{\text{C00026}}$ ⇔ C00036 + $\underline{\text{C00025}}$

Pathway 13: Entner-Doudoroff pathway, β-D-Glucose to D-Glyceraldehyde

C01172 ⟶ C01236 ⟶ C00345 ⟶ C04442 ⟶ C00118

R02736	$\underline{\text{C01172}}$ + C00006 ⇔ $\underline{\text{C01236}}$ + C00005 + C00080
R02035	$\underline{\text{C01236}}$ + C00001 ⇔ $\underline{\text{C00345}}$
R02036	$\underline{\text{C00345}}$ ⇔ $\underline{\text{C04442}}$ + C00001
R05605	$\underline{\text{C04442}}$ ⇔ $\underline{\text{C00118}}$ + C00022

Pathway 14: 2-Oxobutanoate to L-Isoleucine

C00109 ⟶ C06006 ⟶ C14463 ⟶ C06007 ⟶ C00671 ⟶ C00407

R08648	C00022 + $\underline{\text{C00109}}$ ⇔ $\underline{\text{C06006}}$ + C00011
R05069	$\underline{\text{C06006}}$ ⇔ $\underline{\text{C14463}}$
R05068	$\underline{\text{C14463}}$ + C00005 + C00080 ⇔ $\underline{\text{C06007}}$ + C00006
R05070	$\underline{\text{C06007}}$ ⇔ $\underline{\text{C00671}}$ + C00001
R02199	$\underline{\text{C00671}}$ + C00025 ⇔ $\underline{\text{C00407}}$ + C00026

Pathway 15: Chorismate to L-Tryptophan

C00251 ⟶ C00108 ⟶ C04302 ⟶ C01302 ⟶ C03506 ⟶ C00078

R00985	$\underline{\text{C00251}}$ + C00014 ⇔ $\underline{\text{C00108}}$ + C00022 + C00001
or **R00986**	$\underline{\text{C00251}}$ + C00064 ⇔ $\underline{\text{C00108}}$ + C00022 + C00025
R01073	$\underline{\text{C00108}}$ + C00119 ⇔ $\underline{\text{C04302}}$ + C00013
R03509	$\underline{\text{C04302}}$ ⇔ $\underline{\text{C01302}}$
R03508	$\underline{\text{C01302}}$ ⇔ $\underline{\text{C03506}}$ + C00011 + C00001
R02722	C00065 + $\underline{\text{C03506}}$ ⇔ $\underline{\text{C00078}}$ + C00118 + C00001

Pathway 16: Shikimate to L-Tyrosine

C00493 ⟶ C03175 ⟶ C01269 ⟶ C00251 ⟶ C00254 ⟶ C01179 ⟶ C00082

R02412	C00002 + $\underline{\text{C00493}}$ ⇔ C00008 + $\underline{\text{C03175}}$
R03460	C00074 + $\underline{\text{C03175}}$ ⇔ C00009 + $\underline{\text{C01269}}$
R01714	$\underline{\text{C01269}}$ ⇔ $\underline{\text{C00251}}$ + C00009
R01715	$\underline{\text{C00251}}$ ⇔ $\underline{\text{C00254}}$
R01728	$\underline{\text{C00254}}$ + C00003 ⇔ $\underline{\text{C01179}}$ + C00011 + C00004 + C00080
R00734	$\underline{\text{C01179}}$ + C00025 ⇔ $\underline{\text{C00082}}$ + C00026

(S2 Table contd.) Details on the different pathways used.

Pathway 17: Ornithine biosynthesis, L-Glutamate ⇒ L-Ornithine	
C00025 ⟶ C00624 ⟶ C04133 ⟶ C01250 ⟶ C00437 ⟶ C00077	
R00259	C00024 + C00025 ⇔ C00010 + C00624
R02649	C00002 + C00624 ⇔ C00008 + C04133
R03443	C04133 + C00005 + C00080 ⇔ C01250 + C00009 + C00006
R02283	C01250 + C00025 ⇔ C00437 + C00026
R00669	C00437 + C00001 ⇔ C00033 + C00077
or **R02282**	C00437 + C00025 ⇔ C00077 + C00624
Pathway 18: Phosphoenolpyruvate to L-Aspartate	
C00074 ⟶ C00022 ⟶ C00041 ⟶ C00049	
R00200	C00008 + C00074 ⇔ C00002 + C00022
R00258	C00022 + C00025 ⇔ C00041 + C00026
R00397	C00041 + C00011 ⇔ C00049
Pathway 19: Phosphoenolpyruvate to L-Asparagine	
C00074 ⟶ C00022 ⟶ C00041 ⟶ C00049 ⟶ C00152	
R00200	C00008 + C00074 ⇔ C00002 + C00022
R00258	C00022 + C00025 ⇔ C00041 + C00026
R00397	C00041 + C00011 ⇔ C00049
R00483	C00002 + C00049 + C00014 ⇔ C00020 + C00013 + C00152
Pathway 20: L-Glutamate to L-Proline	
C00025 ⟶ C03287 ⟶ C01165 ⟶ C03912 ⟶ C00148	
R00239	C00002 + C00025 ⇔ C00008 + C03287
R03313	C03287 + C00005 + C00080 ⇔ C01165 + C00009 + C00006
R03314	C01165 ⇔ C03912 + C00001
R01251	C03912 + C00005 + C00080 ⇔ C00148 + C00006

REFERENCES

[1] Noel M. O'Boyle, Michael Banck, Craig A. James, Chris Morley, Tim Vandermeersch, and Geoffrey R. Hutchison. 2011. Open Babel: An open chemical toolbox. *Journal of Cheminformatics* 3, 1 (2011), 33–14. DOI: https://doi.org/10.1186/1758-2946-3-33