Supporting Information
iMet: a network-based computational tool to assist in the annotation of metabolites from tandem mass spectra

Antoni Aguilar-Mogas[1], Marta Sales-Pardo[1], Miriam Navarro[2,3], Roger Guimerà[1,4], and Oscar Yanes[2,3]

[1]Departament d'Enginyeria Química, Universitat Rovira i Virgili, Av. Països Catalans 26, 43007 Tarragona, Catalonia, Spain
[2]Metabolomics Platform, Department of Electronic Engineering (DEEEA), Universitat Rovira i Virgili, Av. Països Catalans 26, 43007 Tarragona, Catalonia, Spain
[3]Biomedical Research Centre in Diabetes and Associated Metabolic Disorders (CIBERDEM), Monforte de Lemos 35, 28029 Madrid, Spain
[4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Lluís Companys 23, 08010 Barcelona, Catalonia, Spain

Contents

iMet step by step

Given an unknown metabolite χ, a reference set $R = \{R_1, R_2 \ldots, R_n\}$ comprising n metabolites, and a trained random forest (see "Construction of the classifier" in the

Main Text), the algorithm builds a list of candidate neighbors of χ as follows (Figure S1):

1. Obtaining spectral similarities and mass differences: Given an unknown metabolite χ, the algorithm computes the cosine similarity between the unknown metabolite's ESI Q-TOF MS/MS spectra and every MS/MS spectra in the database obtained under the same experimental conditions (i.e., ionization mode and collision energy). Spectral similarity is calculated for each collision energy separately, that is, we do not combine or merge spectra from different collisions energies. Our algorithm can work using just one or two collision energies, but it performs better if all possible collision energies are provided as input data. The algorithm also computes the exact mass of χ, as the precursor ion mass plus or minus the hydrogen atom mass, depending on the ionization mode ($[M+H]^+$ or $[M-H]^-$, for positive or negative ionization, respectively). Although other possible adducts can be formed, the protonated adducts are by far the most abundant in ESI-MS and their MS/MS spectra are prevalent in databases[1], and so iMet only takes into account protonated adducts (although application to other adducts would be immediate as soon as a significant number of spectra from other adducts are incorporated in databases).

   With the exact mass M_S of the unknown metabolite, the algorithm computes the mass difference Δm between χ and every metabolite in the reference set R. In each case, the algorithm obtains the fraction of all metabolite pairs with that Δm that are actually reactant pairs. We assume that this ratio reflects the probability of two metabolites being neighbors given only their mass difference (Fig. 3A in the Main Text).

2. Classification: Using the spectral similarity obtained from the comparison of MS/MS spectra and the ratio of RPs from the calculated mass difference, the classification algorithm computes the likelihood for each metabolite in the database of being neighbors with the unknown metabolite (score v). Those metabolites in the database with a likelihood lower than 0.5 are discarded; the remaining metabolites are the most likely to form a RP with the unknown metabolite. These are the candidate neighbors of the unknown metabolite.

3. Determination of the chemical transformation: From the mass difference of each candidate neighbor, the algorithm proposes the most probable chemical transformations linking each candidate to the unknown metabolite. Specifically, the algorithm considers all possible molecular formulas that are compatible with the observed mass difference. Note that, in general, these molecular formulas coincide with one of the 202 unique chemical transformations derived from RPs listed in Table S4.

4. Formula consensus: Given that the molecular formulas of the candidate neighbor and its associated chemical transformation are known (from step (3)), the algorithm postulates a final formula for the unknown metabolite based on the sum of these two formulas. The algorithm then computes the frequency of each

potential final formula as the proportion of candidate neighbors and their chemical transformation that yield that same molecular formula. Then, in order to resolve possible ties, the algorithm favors simple chemical transformations over complex ones by dividing the frequency of each final formula by the square root of the number of atoms associated with the chemical transformation. This final number is the formula consensus score (score f).

5. Isotope pattern comparison: The theoretical MS isotope pattern of every final formula is computed (as described in "Elucidation of the theoretical isotope pattern" above), and compared to the experimental isotope pattern of the unknown metabolite. As before, we use the spectral similarity previously defined. Consequently, the algorithm associates every candidate neighbor with a spectral similarity (score p) based on its isotopic distribution.

6. Output: for each unknown metabolite-candidate neighbor pair, we have an overall score (s)

$$s = v \times f \times p \tag{S1}$$

   where, again, v is the score of the classifier, f the formula consensus score, and p is the spectral similarity between theoretical and empirical isotopic patterns.

All in all, the outcome of our algorithm is a sorted list of candidate neighbors of the unknown sample χ, ranked by their score s.

Elucidation of the theoretical isotope pattern

Consider a molecular formula F composed by n different chemical elements A_k ($k \in \{1 \ldots n\}$) each of them appearing ν_k times in the formula. The isotopic pattern of F can be calculated taking into account the exact masses and abundances of each isotope of each element present in the formula. In particular, all possible combinations of isotopes of all elements present in the formula have to be calculated, as each different combination of isotopes yields a separate peak in the spectrum.

The mass M_S of each different isotopic composition S is the sum over the contributions of the n chemical elements present in the molecule:

$$M_S = \sum_{k=1}^{n} \sum_{i=1}^{N} x_i m_i \tag{S2}$$

where N represents the number of different possible isotopes of each element, m_i is the exact mass of isotope i and x_i is the number of isotopes of type i that contribute to the signal in the specific isotopic composition under consideration.

The intensity of the peak I_S corresponding to a specific isotopic composition of mass M_S, is proportional to the abundance of that isotopic composition, or in other words, to the probability of picking that precise isotopic composition given the molecular formula

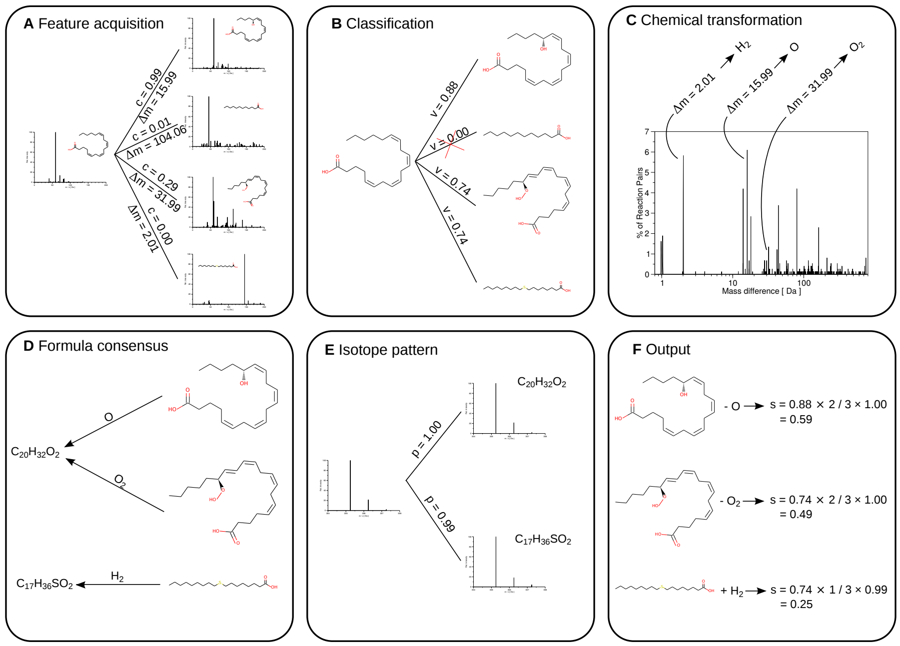


Figure S1: General procedure followed by the iMet algorithm. In this simplified example the "unknown" test metabolite is arachidonic acid, and the database consists only of four metabolites (16-hydroxyeicosatetraenoic acid, dodecanoic acid, 15-hydroperoxyeicosatetraenoic acid and 9-thiastearic acid) with their corresponding MS/MS spectra obtained from the same experimental conditions. All mass values are expressed in Da. (A) Feature acquisition. The MS/MS spectrum of the test metabolite is compared to those in the database, obtaining a spectral similarity (c) and a value of mass difference. (B) Classification. Using the above-mentioned features, the classifier computes the confidence of the classifier for every compound in the database of being a neighbor of the sample metabolite (v). Those metabolites with a classifier confidence higher than 0.5 are candidate neighbors of the test metabolite. (C) Determination of chemical transformation. For each candidate, a moiety is proposed according to the mass difference between the candidate neighbor and the test metabolite. (D) Formula consensus. The molecular formula of the test metabolite is then calculated as the sum of the candidate neighbor's formula plus the chemical transformation. The frequency of each formula amongst the formulas of all the candidate neighbors is computed (f). (E) Isotope pattern. If the experimental isotope pattern of the test metabolite is available, the algorithm compares it to the theoretical isotope pattern of the proposed formula for the sample metabolite (p). (F) Output. The final output is a sorted list of candidates, ranked by the final algorithm score (s).

$$I_S = \lambda p(S|F) = \lambda \prod_{k=1}^{n} N^{-\nu_k} \nu_k! \prod_{i=1}^{N} \frac{p_k(i)^{x_i}}{x_i!} \tag{S3}$$

where λ is a normalizing factor that sets the maximum intensity to 100, $p_k(i)$ is the relative abundance in nature of isotope i, and ν_k is the number of atoms of an element in the molecular formula.

Fingerprints and similarity coefficients

Fingerprints are algebraic objects that encompass relevant sctructural features of molecules [2]. They are usually defined as strings or vectors, where each component represents a specific structural feature. These strings or vectors can be divided in two cathegories: bit vectors in which each element indicates the presence or abscence of a given feature, and integer vectors in which each elemenet counts the number of ocurrences of each structural feature.

In the Molecular Access System structural keys or MACCS fingerprints [3], each element in the fingerprint vector represents the presence or abscence of a predefined substructure or fragment. These are very general purpose fingerprints, with a very low computational cost.

Atom-pair fingerprints (AP) [4] describe the molecular structure of a compound by establishing all possible pairs of atoms in the molecule, and then describing their connectivity. When done throghout all possible atom pairs in the molecule, the fingerprint is highly specific of that molecule. An extension to this concept are the Daylight fingerprints (DL) [5]. These fingerprints are constructed by representing a large enough number of paths along the molecule's bonds. Then the fingerprint is *folded* (divided and combined using logical operators) increasing its bit density. Both AP and DL fingeprints represent molecular topology through connectivity pathways.

Circular fingerprints allow for the representation of complex molecules. They take into account the local environment of each atom of a molecule, so that each feature describes the connectivity of the neighbors of each atom. This type of fingerprints are able to represent a larger set of sctructural features and capture the complexity in the molecular structure. One of the most widely used fingerprint of this type is the Extended-connectivity Fingerprint with diameter four or ECFP4 [6].

One of the advantages of representing molecular structures with fingerprints is that numerical similarities can be established between them. By reducing the question of similarity to algebraic operations, we can establish similarity relations in a consistent manner. One of the most popular similarity coefficients is the Tanimoto coefficient T_c [7, 8], which is equivalent to the Jaccard index calculated between bit vectors of molecular structure features. It is defined as the number of shared features of the two fingerprints divided by the total number of features:

$$T_c = \frac{|A \cap B|}{|A \cup B|} \tag{S4}$$

where A and B are molecular fingerprints of the bit vector type.

One similarity coefficient that can be calculated over all types of fingerprints is the Dice coefficient D_c [9] (also named F1 score). It is defined as the number of shared features between the fingerprints over the arithmetic mean of the number of features in A and B.

$$D_c = \frac{2|A \cap B|}{|A \cup B| + |A \cap B|} \tag{S5}$$

It can be seen that, given that $|A \cup B| + |A \cap B| \leq 2|A \cap B|$, $D_c \geq T_c$ when calculated over the same fingerprints A and B. In fact, the relation between these two similarity coefficients is

$$D_c = \frac{2T_c}{1 + T_c} \tag{S6}$$

These two similarity coefficients depend on the sizes of both fingerprints equally. To control more finely over these dependencies, the Tversky coefficient T_v can be used [10]. It provides two parameters that allow the shifting of the balance between the sizes of the fingerprints compared.

$$T_v = \frac{|A \cap B|}{\alpha(|A| - |A \cap B|) + \beta(|B| - |A \cap B|) + |A \cap B|} \tag{S7}$$

Setting $\alpha = \beta = 0.5$ we recover the expression of D_c, while setting them to $\alpha = \beta = 1$ we obtain the T_c.

Because these fingerprints and similarity coefficients have different characteristics, we investigated what particular combination best suited our problem. We tested every combination of fingerprints and similarity coefficients, and conducted an AUC [11] analysis using a dataset with 5,074 metabolites, including 834 RPs. To assess the discriminatory power of each combination, we calculated the area under the curve (Table S1). Tanimoto coefficients were calculated using the bit vector variant of each fingerprint, while the rest of the methods used the integer vector version. For T_v we set α to 0.01 and β to 0.99 to ensure that this coefficient produces results different enough from the other coefficients (note that in our case, the assignment of a fingerprint in a pair as A or B in Equation S7 is made randomly). The results obtained show that ECFP4 type fingerprints have a higher AUC value with all similarity coefficients, indicating that they can capture the structural features needed to distinguish between RPs and non RPs. As for the similarity coefficients, both D_c and T_v produce higher AUCs than T_c, meaning that these two coefficients are more adequate for the present problem. From all the combinations, the ECFP4-D_c was the combination with a higher AUC. This result, together with the fact that the D_c is parameter-free, make it the combination of choice for the current problem.

Similarity threshold

Despite several attemps to find a similarity value that could be applied in different areas of research to differentiate between similar and non similar molecules, such a universal threshold has not yet been found. In fact, those values that are adequate for

Table S1: AUCs for fingerprints and similarity coefficients.

	D_c	T_c	T_v
MACCS	0.9472	0.8897	0.8680
AP	0.9497	0.8823	0.9381
DL	0.9264	0.8687	0.7993
ECFP4	0.9697	0.9081	0.9630

some particular application, fail to provide relevant structural information in others. Therefore, the most adequate procedure is to find the value that best suits a particular problem, which in this case is separating RPs from non RPs according to their structural similarity from our database.

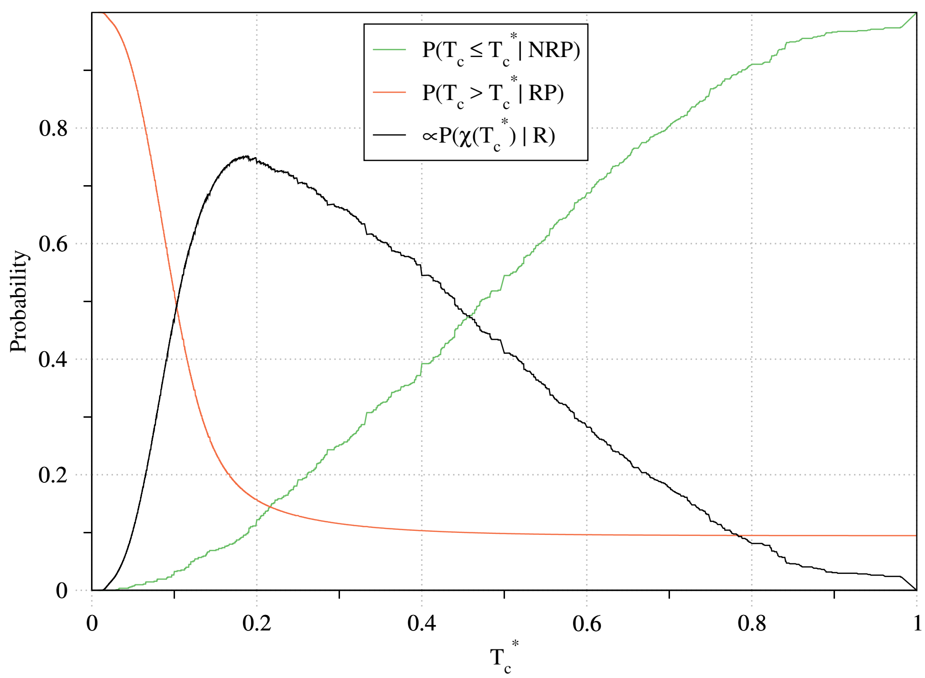


Figure S2: Cumulative distribution functions for the Tanimoto coefficient T_c using ECFP4. The red line depicts the probability of a non RP having a T_c value higher than T_c^*, and the green line represents the probability of a RP having a T_c lower. The black line is proportional to the probability that a pair of metabolites has a T_c value higher than T_c^* if they ar RP, and lower if they are not RPs..

We can plot the FPR associated to each value x of a similarity coefficient when comparing each pair of metabolites in our database (see Figure S2 for $x = T_c$, see Main Text for $x = D_c$). In this case, this FPR(x) corresponds to the ratio of non RPs that have a similarity higher than the given vale. Note that this curve is equivalent to the complementary cumulative distribution function, as for each value of similarity x

the curve depicts the probability of a non RP having a similarity higher than that value $P(X > x|NRP)$. The same can be done with the FNR, which in this case represents the proportion of RPs with a similarity below a given value. In this case, the curve corresponds to the cumulative distribution function representing the probability of two metabolites having a similarity equal to or lower than a given value, when they are RPs $P(X \leq x|RP)$. Of course we are interested in the opposite, which is the probability that a RP has a higher value than the given similarity. The same happens with the non RPs. In fact, we are interested in the joint probability, that is, the probability that, given a similarity value, the RPs have a higher similarity and that the non RPs have a lower similarity value. This can be expressed with the equation:

$$
\begin{aligned}
P(\chi|R) &\propto P(X \leq x|NRP)P(X > x|RP) \\
&= [1 - P(X > x|NRP)][1 - P(X \leq x|RP)] \\
&= [1 - \mathrm{FPR}(x)][1 - \mathrm{FNR}(x)]
\end{aligned} \tag{S8}
$$

where R represents the relation between two metabolites (whether they are RPs or non RPs), and χ is the interval to which their structural similarity value x belongs (from 0 to x or from x to 1). This is, in fact, proportional to the probability that two metabolites share a similarity that is higher or lower than a given threshold, depending on whether they are a RP or not.

Plotting the curves for the three similarity coefficients proposed in this paper, namely the Tanimoto coefficient T_c, the Dice coefficient D_c and the Tversky coefficient T_v, we can observe that these probabilities have a maximum at a specific similarity value (Figure S3). Note that we can compare the probabilitites because the proportionality is the same for the three coefficients, as it does not depend on them, but rather on the dataset used (which in this case is the same dataset for the three coefficients). At this point, the probability that a RP has a higher similarity and that a non RP has a lower similarity is maximum. This implies that the probability of correct classification using only the similarity coefficient value is maximum at this point. This is, therefore, the optimal point to set a threshold to discriminate between RPs and non RPs. In our case, the similarity coefficient with a higher maximum probability of correct classification is the D_c, with a threshold value of $D_c = 0.32$.

Comparison of cosine and spectral alignment methods

Methods of spectral comparison that rely in a dot-product of intensities outperform other methods [12]. In the simplest implementation, the algorithm discretizes each spectrum to compare in equal intervals of m/z (or bins). From this it constructs a vector, with each of its components corresponding to the relative intensity of m/z values in the interval. Then the cosine similarity is just the normalized dot-product of the two vectors. An extension of this simple cosine comparison is the spectral alignment method [13]. The main difference between the spectral alignment and the cosine similarity lies in the way peaks are compared. While the cosine similarity is done by binning in the m/z dimension and matching bins, the spectral alignment compares each

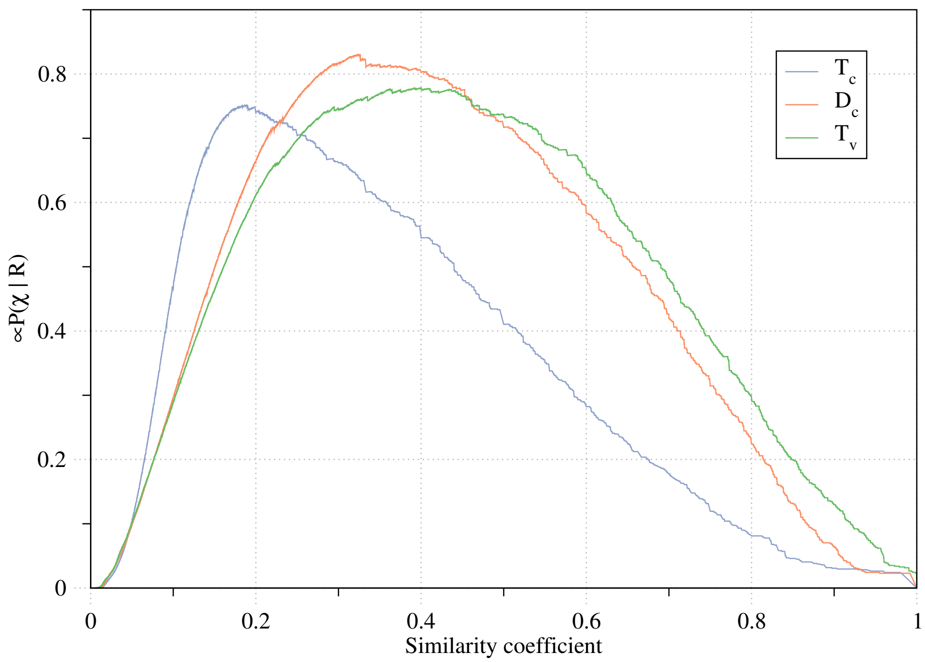


Figure S3: Probability of discrimination of the three similarity coefficients using ECFP4. In green, the Tanimoto coefficient T_c, in red the Dice coefficient D_c, and in blue the Tversky coefficient T_v. The three maxima occur at $T_c = 0.17$, $D_c = 0.32$ and $T_v = 0.40$.

peak in one spectrum with every peak in the other spectrum and then chooses the optimal peak-to-peak matching consistent with certain constraints (e.g. only certain mass difference between matched peaks are allowed). As discussed in [13], this process involves solving a bipartite matching problem, which current implementations can only do in $O(n^3)$ computational time, where n is the number of peaks (see, for instance, [14]). On the other hand, the cosine similarity can be implemented in such a way that its computational cost scales linearly with the number of peaks, that is, it is $O(n)$. As to train the algorithm we need tens of thousands of spectrum comparisons (and to test an unknown metabolite against the current database we need approximately 5,000 comparison), the cosine similarity is more suitable from this perspective. To confirm the difference in computational time, we benchmarked both methods of spectral comparison. We compared the same 100 pairs of metabolites taken randomly from our database with both methods in the same machine, and measured their execution time. The mean computational time was 0.0008s per comparison for the cosine similarity, and 2s per comparison for the spectral alignment method (in this case, spectra obtained at high collision energies took much longer to align than at low ones due to the cubic scaling in the number of peaks).

To further investigate possible advantages and disadvantages of both methods, we analyzed their discriminatory power at separating RPs from non RPs. In particular, we performed an AUC analysis using the same subset of metabolites for both methods (the subset of metabolites consisted of 500 metabolites and included 25 reactant pairs). We obtained the following values of the AUC statistic, summarized in the Table S2. Even though both methods lie in the same range of classification power, the AUCs for the cosine similarity are consistently higher than those obtained with the spectral alignment.

This, together with the fact that the cosine similarity is cheaper to compute than the spectral alignment (specially if the database grows), make it the method of choice for the iMet algorithm.

Table S2: AUCs for spectral alignment and cosine methods

	Neg. Ion. Mode			Pos. Ion. Mode		
	10V	20V	40V	10V	20V	40V
Spec. Alignment	0.7148	0.6425	0.6778	0.7500	0.6512	0.7366
Cosine	0.7432	0.7864	0.7091	0.7636	0.7936	0.7873

References

[1] M. Vinaixa, E. L. Schymanski, S. Neumann, M. Navarro, R. M. Salek, and O. Yanes. Mass spectral databases for LC/MS and GC/MS-based metabolomics: state of the field and future prospects. *Trends Anal. Chem.*, 78:23–35, 2016.

[2] D. Stumpfe and J. Bajorath. Similarity searching. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 1(2):260–282, 2011.

[3] J. L. Durant, B. A. Leland, D. R. Henry, and J. G. Nourse. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci. 2002*, 42:1273–1280, 2002.

[4] R. E. Carhart, D. H. Smith, and Venkataraghavan R. Atom pairs as molecular features in structure-activity dtudies: definition and applications. *J. Chem. Inf. Comp. Sci.*, 25:64–73, 1985.

[5] C. A. James and D. Weininger. *Daylight theory manual*. Daylight Chemical Information Systems, Inc., 2011.

[6] D. Rogers and M. Hahn. Extended-connectivity fingerprints. *J. Chem. Inf. and Model.*, 50:742–754, 2010.

[7] T. Tanimoto. An elementary mathematical theory of classification and prediction. IBM internal report, IBM Corporation, Armonk, NY, 1957.

[8] P. Willett, J. M. Barnard, and G. M. Downs. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.*, 38:983–996, 1998.

[9] L. R. Dice. Measures of the amount of ecological association between species. *Ecology*, 26:297–302, 1945.

[10] A. Tversky. Features of similarity. *Psychol. Rev.*, 84:327–352, 1977.

[11] J. A. Hanley and B. J. McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143:29–36, 1982.

[12] S. E. Stein and D. R. Scott. Optimization and testing of mass spectral library search algorithms for compound identification. *J. Am. Soc. Mass Spectrom.*, 5(9):859 – 866, 1994.

[13] J. Watrous, P. Roach, T. Alexandrov, B. S. Heath, J. Y. Yang, R. D. Kersten, M. van der Voort, K. Pogliano, H. Gross, J. M. Raaijmakers, B. S. Moore, J. Laskin, N. Bandeira, and P. C. Dorrestein. Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. U. S. A.*, 109(26):E1743–E1752, 2012.

[14] R. E. Burkard, M. Dell'Amico, and S. Martello. Linear sum assignment problem. In *Assignment Problems*, pages 73–144. Society for Industrial and Applied Mathematics, 2012.

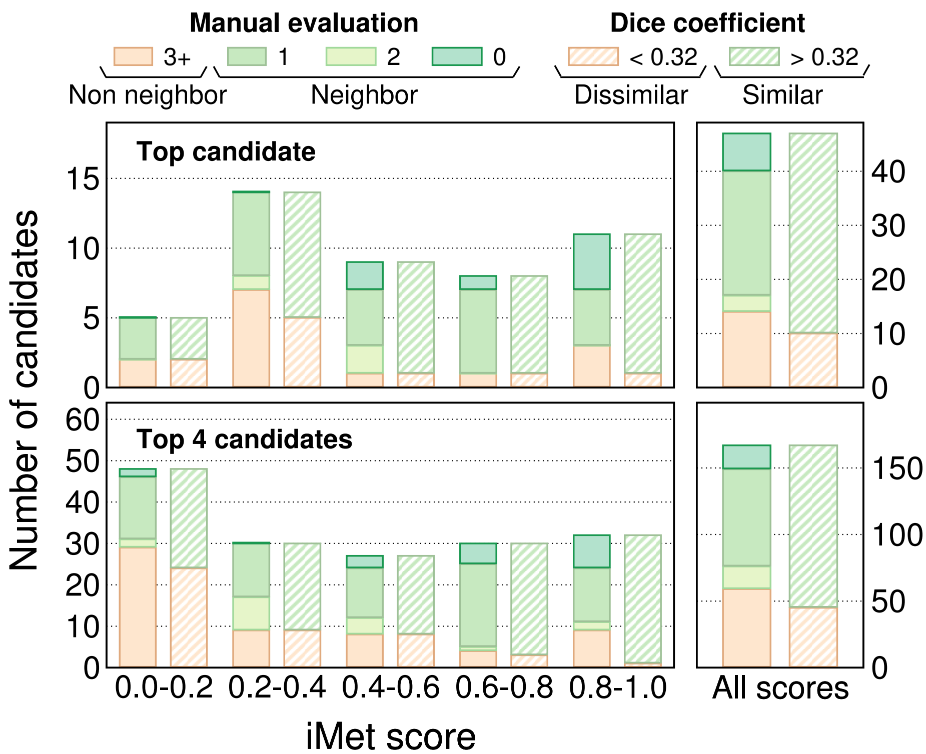


Figure S4: Scores of the experimental validation. Summary of validation for 48 metabolites whose MS/MS spectra were acquired in our laboratory. We display the number of chemical transformation required to reach the target metabolite, as well as the top candidate neighbors that are structurally similar to the test metabolite as a function of the iMet score.

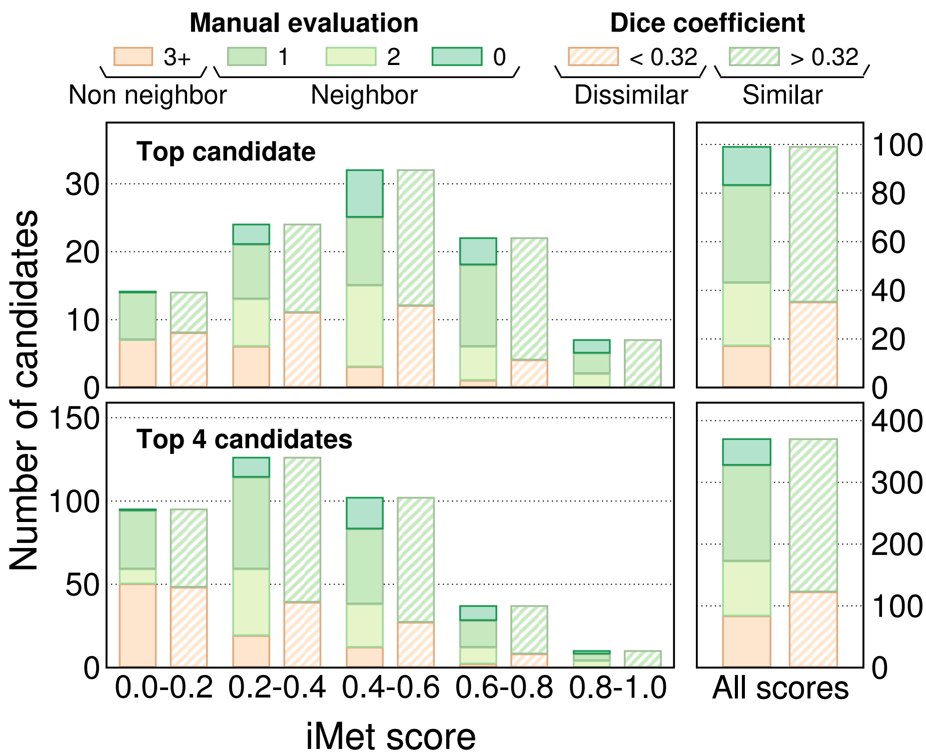


Figure S5: Scores of the leave-one-out cross-validation. Summary of validation for 100 metabolites that were taken directly from our database. We display the number of chemical transformation required to reach the target metabolite, as well as the top candidate neighbors that are structurally similar to the test metabolite as a function of the iMet score.

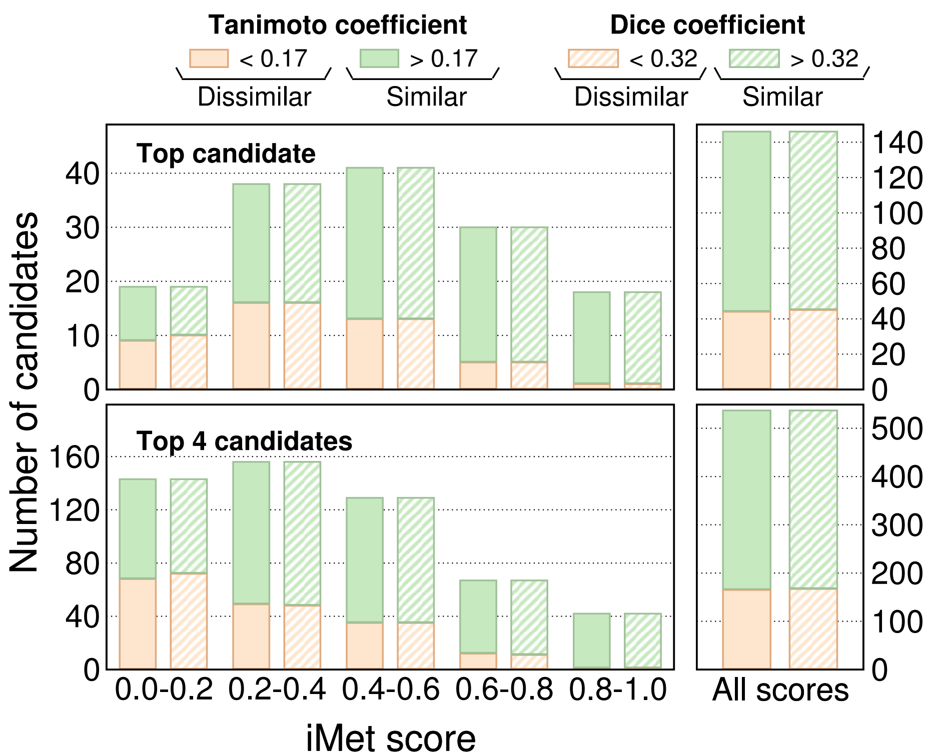


Figure S6: Score comparison between Dice and Tanimoto coefficients. Summary of validation for 148 metabolites (the combination of the two cross-validations), displaying the number of structurally similar candidates according to the Dice coefficient and to the Tanimoto coefficient criterion.

Table S3: RPs in the training set

KEGG IDs	KEGG IDs	KEGG IDs	KEGG IDs	KEGG IDs
C08300 - C02765	C00103 - C00394	C00079 - C00082	C00498 - C00008	C02727 - C00047
C05831 - C00632	C00588 - C00157	C00121 - C00212	C00214 - C00364	C00004 - C00003
C00180 - C00156	C00133 - C00022	C05956 - C00427	C00256 - C00022	C01083 - C00267
C00006 - C00003	C01157 - C03440	C00637 - C00955	C00049 - C00152	C00124 - C00492
C05926 - C04874	C00482 - C00933	C00366 - C00385	C02666 - C01494	C00655 - C00144
C02273 - C00333	C05335 - C00109	C00074 - C00022	C00417 - C00490	C00106 - C00105
C01179 - C00544	C02050 - C01595	C00978 - C01598	C00122 - C00049	C05302 - C05301
C00509 - C06561	C05660 - C05635	C00135 - C07474	C00037 - C05519	C00212 - C00020
C00025 - C01047	C14749 - C00219	C00954 - C02693	C00239 - C00705	C00092 - C00103
C00055 - C00307	C05965 - C00219	C11555 - C01284	C00794 - C00267	C00398 - C06213
C00483 - C00082	C02515 - C00082	C00157 - C00219	C00140 - C00329	C00811 - C00082
C05944 - C00864	C04608 - C12627	C05122 - C00695	C00062 - C04137	C03582 - C05901
C08591 - C06098	C00117 - C00455	C00386 - C00135	C00783 - C00729	C14762 - C14765
C00544 - C00642	C05402 - C00492	C00257 - C00198	C00037 - C05951	C07649 - C11736
C00144 - C00044	C00376 - C00473	C05962 - C05961	C00249 - C00154	C00166 - C05852
C00639 - C00427	C00100 - C00163	C00696 - C00427	C00292 - C00108	C00628 - C05585
C09815 - C00180	C03758 - C05587	C00105 - C00029	C00267 - C00089	C00010 - C02050
C00166 - C00079	C00010 - C01944	C00020 - C00008	C00095 - C00492	C00167 - C00191
C00155 - C00109	C01575 - C16719	C00954 - C00463	C12989 - C00047	C00882 - C00020
C00105 - C00052	C05512 - C00559	C00239 - C00881	C00077 - C00327	C00010 - C00683
C02985 - C01019	C00149 - C00048	C05356 - C00219	C05625 - C05623	C00387 - C00144
C00141 - C00183	C00043 - C00105	C04858 - C04608	C00498 - C00103	C16358 - C16359
C02710 - C00123	C00231 - C00199	C01179 - C00642	C04741 - C04654	C00780 - C00643
C00020 - C00016	C01026 - C00213	C00037 - C00048	C00061 - C00255	C00164 - C00332
C00010 - C02249	C00542 - C00065	C00319 - C06124	C00581 - C00300	C05442 - C01753
C00048 - C00213	C00090 - C00196	C00090 - C00230	C00389 - C05623	C04579 - C01204
C00073 - C00155	C05488 - C00735	C01672 - C10773	C14770 - C00219	C00099 - C02642
C06772 - C00189	C02140 - C05490	C01094 - C00111	C01204 - C04563	C00637 - C00954
C00355 - C03758	C00179 - C00134	C00331 - C00078	C01902 - C01054	C02939 - C00010
C07130 - C16358	C08491 - C11512	C10028 - C09827	C00331 - C02043	C00547 - C03758
C00251 - C00022	C00793 - C00022	C01801 - C00673	C09099 - C09789	C00250 - C00314
C00636 - C00096	C01204 - C01284	C00474 - C00255	C00440 - C00101	C00780 - C00978
C00092 - C00085	C07130 - C07481	C01179 - C00082	C01595 - C04056	C06606 - C00870
C00584 - C00427	C00128 - C00055	C08734 - C06554	C00072 - C05422	C05903 - C12249
C05593 - C01161	C05961 - C01312	C05904 - C12137	C07201 - C07204	C02625 - C03664
C03546 - C01220	C00140 - C01674	C02166 - C00051	C00811 - C00156	C00117 - C00020
C00740 - C00065	C00750 - C00315	C00137 - C01177	C03077 - C01161	C14768 - C14772
C00122 - C01384	C01468 - C00642	C00199 - C00345	C00530 - C00156	C00508 - C00259
C02990 - C00318	C00006 - C00153	C00055 - C00112	C10521 - C06563	C00042 - C00311
C00262 - C00147	C05958 - C05957	C05138 - C01953	C00331 - C00637	C00482 - C05610
C00498 - C00020	C00025 - C00051	C14769 - C14773	C01108 - C01424	C00003 - C00455
C00243 - C00124	C00078 - C02693	C05580 - C05584	C00526 - C00365	C00106 - C00526
C01479 - C00729	C00319 - C01747	C00794 - C00247	C00153 - C02918	C02291 - C00022

KEGG IDs	KEGG IDs	KEGG IDs	KEGG IDs	KEGG IDs
C00519 - C01678	C00245 - C05122	C06044 - C06046	C00025 - C00026	C01607 - C02060
C00136 - C00630	C04137 - C00022	C00036 - C00049	C01161 - C00642	C00144 - C00096
C00036 - C00402	C00257 - C00345	C00135 - C00388	C00134 - C00077	C01921 - C00695
C05138 - C01176	C01144 - C00877	C00263 - C01118	C01419 - C00051	C00129 - C00341
C00267 - C00208	C13747 - C07481	C00683 - C02170	C00041 - C00022	C00117 - C00199
C05966 - C00219	C01724 - C01054	C00155 - C00021	C14771 - C00219	C00315 - C01029
C00065 - C00022	C00398 - C00637	C00491 - C00022	C02814 - C01751	C00491 - C00097
C03794 - C00130	C00090 - C02480	C00099 - C00049	C00159 - C00392	C00055 - C00117
C00251 - C00156	C01118 - C00109	C00049 - C02512	C00735 - C02821	C13747 - C16358
C08650 - C08578	C00121 - C00117	C00814 - C06563	C00148 - C00077	C00049 - C03794
C00042 - C01118	C05284 - C05285	C00387 - C00242	C00601 - C05332	C01035 - C00334
C00189 - C03557	C00106 - C00380	C00696 - C00639	C00020 - C00130	C00575 - C00002
C00065 - C02737	C00178 - C03088	C02985 - C00325	C00010 - C00100	C00010 - C03069
C01620 - C00072	C00483 - C03758	C16614 - C02380	C00025 - C05829	C08604 - C08639
C00078 - C00022	C00136 - C00877	C00489 - C00527	C10509 - C00858	C00090 - C00805
C02249 - C00219	C01477 - C04608	C03005 - C00108	C00310 - C00181	C00121 - C00299
C00805 - C01198	C02140 - C03205	C06427 - C16321	C00250 - C00534	C02291 - C00097
C00544 - C05585	C05952 - C06462	C03626 - C00327	C00047 - C00449	C02515 - C01060
C00559 - C00360	C00559 - C00147	C11700 - C11900	C00037 - C01419	C14748 - C00219
C14769 - C00219	C14825 - C01595	C05565 - C05829	C05631 - C00509	C04483 - C00695
C08725 - C12137	C00780 - C06212	C01227 - C04555	C00504 - C00101	C05956 - C00219
C14827 - C01595	C01094 - C00095	C00095 - C00089	C00715 - C03975	C00090 - C00108
C00423 - C00079	C00073 - C00109	C00010 - C00136	C00006 - C00005	C00843 - C09893
C00497 - C01384	C00385 - C00655	C00275 - C00636	C00230 - C02814	C00042 - C00091
C00041 - C00133	C00354 - C00111	C00010 - C00630	C00208 - C00103	C00526 - C00299
C00055 - C00475	C00214 - C00178	C00047 - C00739	C00025 - C06462	C00061 - C00016
C00100 - C00479	C00025 - C00334	C14536 - C16195	C01595 - C00157	C10107 - C05903
C02305 - C00300	C08737 - C08733	C02519 - C00156	C00766 - C15608	C09827 - C16419
C00108 - C05653	C00164 - C01089	C00385 - C00262	C00082 - C00022	C00055 - C00380
C06538 - C06536	C00423 - C00180	C00106 - C00429	C01727 - C00255	C00788 - C05588
C00062 - C00327	C00295 - C00337	C00262 - C00294	C00037 - C00213	C00330 - C00362
C00329 - C03752	C00942 - C00144	C00398 - C06535	C00680 - C00047	C00475 - C00299
C00253 - C01004	C00689 - C01083	C03990 - C02528	C02868 - C01839	C00311 - C00417
C00415 - C00101	C00253 - C01020	C02465 - C01060	C00026 - C00449	C00261 - C02137
C00427 - C00219	C00152 - C02512	C00002 - C00008	C00036 - C00022	C00123 - C02486
C01672 - C00047	C02165 - C04853	C00294 - C00130	C06124 - C00346	C00029 - C00052
C00355 - C00082	C00986 - C00315	C02291 - C00155	C02465 - C01829	C07649 - C12739
C00134 - C00315	C00385 - C16357	C00819 - C00064	C00037 - C00065	C10208 - C10216
C00095 - C01613	C00242 - C00144	C00763 - C00431	C00099 - C00386	C00805 - C00180
C00085 - C00354	C00178 - C02376	C00010 - C02060	C00385 - C00242	C04051 - C04677
C00186 - C00256	C02814 - C00530	C00099 - C03065	C00261 - C00180	C05011 - C07108
C00018 - C00647	C00951 - C05301	C01312 - C00427	C00025 - C00064	C01477 - C01460
C00376 - C00777	C00188 - C00109	C00575 - C00020	C06427 - C04785	C01456 - C01479
C02483 - C02477	C00186 - C00022	C00135 - C00785	C01750 - C00389	C07054 - C07446
C00793 - C00097	C05576 - C05594	C00025 - C00669	C00010 - C05274	C00224 - C00053

KEGG IDs	KEGG IDs	KEGG IDs	KEGG IDs	KEGG IDs
C00122 - C00149	C00121 - C00294	C00078 - C00643	C04368 - C00082	C05488 - C01176
C00519 - C00245	C13636 - C00530	C05643 - C01598	C00429 - C02642	C00224 - C00020
C00330 - C00242	C12147 - C00188	C08620 - C08604	C00197 - C00258	C05963 - C05964
C00144 - C00394	C06563 - C09126	C00037 - C01586	C00588 - C00307	C00794 - C00095
C00576 - C00719	C02512 - C00097	C00942 - C00044	C00099 - C00864	C01879 - C00669
C01586 - C00180	C00346 - C00189	C00475 - C00380	C00836 - C02934	C00106 - C00299
C14717 - C05958	C01198 - C05629	C00253 - C00153	C00122 - C03794	C00300 - C00213
C00129 - C00235	C01514 - C03951	C01697 - C00124	C04443 - C00389	C01477 - C10192
C04520 - C01284	C00018 - C00250	C14825 - C14828	C00065 - C00097	C00402 - C00049
C01367 - C00212	C00986 - C01029	C00601 - C05853	C00062 - C00077	C00137 - C03546
C06426 - C01595	C05905 - C08604	C00230 - C00156	C00117 - C00130	C03661 - C00089
C00121 - C00475	C00763 - C00148	C01477 - C00509	C05297 - C05290	C00632 - C00108
C00639 - C00584	C00310 - C00379	C05610 - C02325	C00121 - C00387	C00811 - C00423
C00328 - C00108	C00233 - C00123	C00010 - C00091	C00423 - C05629	C00341 - C11388
C05512 - C00262	C00805 - C00628	C02291 - C00109	C10447 - C00355	C00149 - C00022
C00002 - C00020	C00647 - C00534	C05402 - C00124	C01083 - C00208	C14774 - C14770
C00791 - C02565	C02538 - C00468	C00078 - C00398	C00410 - C01176	C05138 - C01227
C00091 - C00232	C00105 - C00299	C00389 - C05903	C00385 - C16353	C00212 - C00021
C00128 - C00270	C12621 - C00423	C00085 - C00279	C00025 - C00624	C06427 - C00157
C00133 - C00993	C14775 - C14771	C00576 - C00114	C00036 - C00158	C00159 - C00095
C00951 - C00468	C01983 - C02137	C17742 - C10443	C05822 - C00475	C00780 - C05659
C00089 - C00492	C00109 - C01234	C00530 - C06186	C00294 - C00212	C00083 - C00383
C00986 - C03375	C00258 - C00168	C00010 - C00882	C00476 - C00310	C00041 - C00097
C06427 - C16300	C00341 - C01500	C00232 - C00334	C00168 - C00065	C01516 - C06174
C00114 - C00719	C05589 - C00547	C05582 - C01161	C05297 - C00468	C00136 - C00246
C00251 - C00108	C00669 - C00097	C08262 - C00123	C00051 - C00669	C13747 - C16353
C00870 - C02235	C00705 - C00112	C14768 - C00219	C00300 - C00791	C01617 - C05631
C00159 - C00275	C00025 - C00449	C00042 - C00122	C00526 - C00881	C07130 - C16357
C00124 - C01613	C00010 - C00527	C00037 - C00581	C01752 - C01494	C00329 - C00352
C02166 - C05951	C00296 - C00852	C00079 - C05332	C02094 - C00376	C00583 - C00479
C00415 - C00504	C00141 - C02504	C01468 - C06730	C00735 - C05284	C00153 - C00455
C00167 - C00029	C00037 - C00051	C00114 - C00933	C05952 - C05951	C00632 - C01987
C00267 - C00095	C00408 - C00047	C02814 - C02235	C16681 - C15493	C03461 - C01126
C03882 - C01746	C00430 - C00931	C03205 - C00410	C05953 - C00584	C05631 - C01514
C01477 - C01514	C01419 - C00097	C05593 - C00544	C02824 - C00571	C00328 - C00041
C01595 - C04717	C02291 - C00065	C00114 - C00157	C00025 - C01879	C14810 - C05965
C00135 - C01152	C00065 - C01005	C06337 - C00180	C00003 - C00020	C00037 - C05598
C00010 - C00083	C00134 - C02714	C01697 - C00795	C01530 - C00412	C00230 - C00568
C00542 - C00097	C03479 - C00101	C00270 - C00022	C00085 - C00095	C00073 - C02989
C00179 - C00062	C00898 - C00258	C08734 - C08733	C00114 - C00588	C00655 - C00130
C00010 - C00154	C01103 - C00295	C07073 - C16561	C00544 - C05852	C01613 - C00492
C00568 - C00292	C00392 - C00095	C00530 - C00628	C00898 - C00036	C00010 - C01832
C07100 - C00048	C00078 - C00065	C02939 - C03069	C00048 - C03664	C05497 - C01176
C00153 - C00003	C00527 - C00877	C01477 - C01470	C00147 - C00020	C00568 - C02372
C01026 - C00719	C00191 - C02670	C07585 - C07446	C00385 - C16358	C00170 - C00147

KEGG IDs	KEGG IDs	KEGG IDs	KEGG IDs	KEGG IDs
C00043 - C00203	C00158 - C00311	C05356 - C04805	C00037 - C00188	C06108 - C00160
C05497 - C00735	C00029 - C00103	C00049 - C01042	C00788 - C00547	C00042 - C00232
C02954 - C05551	C00364 - C00365	C00762 - C00735	C00048 - C00160	C00010 - C00412
C00166 - C05607	C00243 - C00267	C01157 - C00148	C00311 - C00026	C14778 - C00219
C00463 - C12312	C00410 - C01953	C00114 - C01996	C00542 - C00022	C00155 - C00542
C01617 - C00389	C00020 - C03794	C00117 - C00119	C00078 - C00463	C00005 - C00004
C10906 - C00089	C00041 - C00049	C00683 - C00100	C05905 - C09727	C00010 - C00332
C00120 - C01909	C00239 - C00365	C00085 - C00352	C10208 - C00858	C01405 - C00805
C00497 - C00022	C00097 - C00022	C00196 - C00108	C00951 - C08357	C02305 - C00791
C00333 - C00879	C04062 - C01177	C01060 - C01829	C00311 - C00048	C00601 - C00166
C00696 - C05957	C00689 - C00092	C05959 - C00696	C01198 - C02274	C00437 - C00077
C00250 - C00847	C08591 - C02094	C00036 - C00149	C00231 - C00310	C07054 - C07585
C00379 - C00181	C02140 - C01124	C00137 - C00191	C00232 - C00026	C00870 - C03360
C06325 - C01717	C00275 - C00085	C01144 - C00332	C00509 - C09099	C00144 - C00130
C00090 - C00292	C00021 - C00147	C00037 - C01921	C00267 - C00103	C00212 - C00147
C00158 - C00417	C00216 - C00652	C01103 - C00105	C00836 - C00319	C00262 - C00130
C01040 - C02670	C00051 - C00127	C05642 - C01598	C00025 - C00415	C00740 - C00022

Table S4: Net atomic additions

Mass difference [Da]	Chemical transformation	Mass difference [Da]	Chemical transformation
0.980	$-HN+O$	42.030	$+C_6H_2-O_2$
1.035	$+H_3N-O$	42.050	$+C_3H_6$
1.970	$+Cl-HO_2$	43.005	$+CHNO$
1.990	$+F-HO$	43.050	$+C_2H_5N$
2.020	$+H_2$	43.990	$+CO_2$
2.995	$+HNS-CO_2$	44.025	$+C_2H_4O$
4.035	$+H_4$	46.005	$+CH_2O_2$
6.970	$+HS-CN$	47.030	$+CH_5NS-O$
12.000	$+C$	47.990	$+O_3$
13.985	$-H_2+O$	53.050	$+C_5H_7-N$
14.015	$+CH_2$	54.020	$+C_3H_2O$
14.020	$+H_2N_2-O$	56.025	$+C_3H_4O$
15.010	$+HN$	56.040	$+C_2H_4N_2$
15.050	$+CH_5N-O$	57.025	$+C_2H_3NO$
15.980	$+S-O$	57.055	$+C_3H_7N$
15.990	$+O$	58.010	$+C_2H_2O_2$
16.980	$-HN+O_2$	58.040	$+C_3H_6O$
17.000	$+C_3N-HO_2$	60.025	$+C_2H_4O_2$
17.025	$+NH_3$	63.970	$+HPO_2$
18.010	$+H_2O$	63.980	$+O_4$
18.990	$+HO_2-N$	63.990	$+HPO_4-S$
26.015	$+C_2H_2$	67.030	$+C_5H_5O-N$
27.000	$+CHN$	70.040	$+C_4H_6O$
27.040	$+C_2H_5N-O$	71.040	$+C_3H_5NO$
27.995	$+CO$	71.070	$+C_4H_9N$
28.025	$+C_2H_4$	72.020	$+C_3H_4O_2$
28.990	$+NO-H$	74.000	$+C_2H_2O_3$
29.020	$+CH_3N$	79.955	$+SO_3$
29.950	$+S-H_2$	79.965	$+HPO_3$
29.970	$+O_2-H_2$	82.040	$+C_5H_6O$
30.010	$+CH_2O$	82.090	$+C_6H_9O_2-P$
31.970	$+S$	84.020	$+C_4H_4O_2$
31.990	$+O_2$	84.040	$+C_4H_4O_4-S$
33.005	$+H_3NS-O$	84.060	$+C_5H_8O$
35.980	$+HPO-C$	86.000	$+C_3H_2O_3$
39.050	$+C_3H_5N-O$	87.035	$+C_3H_5NO_2$
39.080	$+C_4H_9N-O_2$	88.020	$+C_3H_4O_3$
39.990	$+C_2O$	89.010	$+C_3H_3O_4-N$
42.005	$+C_2H_2O$	90.030	$+C_3H_6O_3$
42.020	$+CH_2N_2$	93.040	$+C_4H_3N_3$

Mass difference [Da]	Chemical transformation
93.050	$+C_6H_7N$
94.020	$+C_4H_2N_2O$
94.970	$+CH_5NSe-O$
95.040	$+C_5H_5NO$
97.980	$+H_3PO_4$
99.050	$+C_8H_5N-O$
99.070	$+C_5H_9NO$
100.015	$+C_4H_4O_3$
101.050	$+C_4H_7NO_2$
103.010	$+C_3H_5NOS$
104.030	$+C_7H_4O$
104.030	$+C_6H_4N_2$
105.020	$+C_6H_3NO$
107.010	$+C_2H_5NO_2S$
112.970	$+CH_4O_5P-N$
113.080	$+C_6H_{11}NO$
114.030	$+C_5H_6O_3$
115.030	$+C_4H_5NO_3$
115.060	$+C_5H_9NO_2$
116.010	$+C_4H_4O_4$
116.045	$+C_5H_8O_3$
116.070	$+C_8H_8N_2-O$
117.030	$+C_4H_7NOS$
117.045	$+C_5H_3N_5-O$
117.050	$+C_4H_7NO_3$
118.025	$+C_5H_2N_4$
118.030	$+C_4H_6O_4$
118.040	$+C_8H_6O$
119.010	$+C_3H_5NO_2S$
119.040	$+C_7H_5NO$
120.040	$+C_3H_8N_2OS$
121.020	$+C_3H_7NO_2S$
123.105	$+C_8H_{13}N$
125.900	$+I-H$
126.110	$+C_8H_{14}O$
129.045	$+C_5H_7NO_3$
129.080	$+C_6H_{11}NO_2$
130.030	$+C_5H_6O_4$
130.060	$+C_6H_{10}O_3$
130.110	$+C_6H_{14}N_2O$
132.045	$+C_5H_8O_4$
133.040	$+C_5H_3N_5$

Mass difference [Da]	Chemical transformation
134.050	$+C_4H_{10}N_2OS$
134.060	$+C_5H_{10}O_4$
136.030	$+C_4H_9O_3P$
137.050	$+C_6H_7N_3O$
138.030	$+C_7H_6O_3$
143.970	$+C_3HN_2O_3P$
145.970	$+C_6H_4Cl_2$
146.050	$+C_6H_{10}O_4$
148.055	$+C_9H_8O_2$
150.040	$+C_8H_6O_3$
152.010	$+C_3H_8N_2OS2$
154.000	$+C_3H_7O_5P$
154.140	$+C_{10}H_{18}O$
158.110	$+C_6H_{14}N_4O$
159.930	$+H_2O_6P_2$
160.030	$+C_5H_8N_2O_2S$
162.030	$+C_9H_6O_3$
162.040	$+C_6H_{10}O_3S$
162.055	$+C_6H_{10}O_5$
169.990	$+C_3H_7O_6P$
176.030	$+C_6H_8O_6$
177.945	$+H_4O_7P_2$
179.930	$+C_6H_3Cl_3$
182.170	$+C_{12}H_{22}O$
194.240	$+C_{15}H_{30}-O$
203.080	$+C_8H_{13}NO_5$
204.190	$+C_{15}H_{24}$
212.005	$+C_5H_9O_7P$
217.920	$+C_6H_3OI$
221.090	$+C_8H_{15}NO_6$
223.935	$+C_3H_2N_2O_6P_2$
224.070	$+C_{12}H_8N_4O$
230.050	$+C_8H_{10}N_2O_6$
232.050	$+C_8H_{12}N_2O_4S$
238.230	$+C_{16}H_{30}O$
242.020	$+C_6H_{11}O_8P$
249.070	$+C_9H_{15}NO_5S$
249.090	$+C_{10}H_{11}N_5O_3$
251.800	$+I_2-H_2$
252.040	$+C_{15}H_8O_4$
252.220	$+C_{20}H_{28}-O$
262.230	$+C_{18}H_{30}O$

Mass difference [Da]	Chemical transformation
266.260	$+C_{18}H_{34}O$
267.100	$+C_{16}H_{13}NO_3$
286.235	$+C_{20}H_{30}O$
294.290	$+C_{20}H_{38}O$
296.055	$+C_{10}H_{18}O_6P_2$
296.100	$+C_{14}H_{12}N_6O_2$
305.040	$+C_9H_{12}N_3O_7P$
305.070	$+C_{10}H_{15}N_3O_6S$
306.030	$+C_9H_{11}N_2O_8P$
314.070	$+C_{10}H_{20}O_7P_2$
316.050	$+C_{11}H_{13}N_2O_7P$
318.220	$+C_{20}H_{30}O_3$
324.110	$+C_{12}H_{20}O_{10}$
329.055	$+C_{10}H_{12}N_5O_6P$
330.040	$+C_{10}H_{11}N_4O_7P$
340.090	$+C_{11}H_{21}N_2O_6PS$
343.820	$+C_6H_2I_2O$
345.050	$+C_{10}H_{12}N_5O_7P$
347.060	$+C_{10}H_{14}N_5O_7P$
371.940	$+C_5H_{11}O_{13}P_3$
386.000	$+C_9H_{12}N_2O_{11}P_2$
390.275	$+C_{24}H_{38}O_4$
421.230	$+C_{23}H_{35}NO_4S$
438.090	$+C_{17}H_{19}N_4O_8P$
473.300	$+C_{24}H_{44}NO_6P$
478.250	$+C_{25}H_{38}N_2O_5S$
486.160	$+C_{18}H_{30}O_{15}$
497.290	$+C_{26}H_{44}NO_6P$
499.310	$+C_{26}H_{46}NO_6P$
541.060	$+C_{15}H_{21}N_5O_{13}P_2$
594.470	$+C_{39}H_{62}O_4$
606.010	$+C_{14}H_{21}N_6O_{13}P_3S$
621.030	$+C_{15}H_{22}N_5O_{16}P_3$
630.470	$+C_{35}H_{67}O_7P$
674.440	$+C_{39}H_{63}O_7P$
706.070	$+C_{19}H_{29}N_6O_{15}P_3S$
721.115	$+C_{20}H_{34}N_7O_{14}P_3S$
749.105	$+C_{21}H_{34}N_7O_{15}P_3S$

Table S5: Cross-references of test metabolites

Name	PubChem CID	KEGG ID
8-Hydroxyguanosine	65131	
N-Amidinoglutamate	135313	C03140
alpha-CEHC	9943542	
Curcumin	5281767	C10443
Cystathionine	834	C00542
N,N-dimethylsphingosine	5282309	C13914
delta-Tocotrienol	5282350	C14156
NADPH	5884	C00005
N8-Acetylspermidine	123689	C01029
3-Nitrotyrosine	65124	
Ornithine	389	C01602
Phosphoenolpyruvate	1005	C00074
9-cis-Retinal	6436082	C16681
Thymidine	5789	C00214
Calcifediol	5283731	C01561
12-Methyltridecanoic acid	520298	
13-cis-Retinal	6436079	
15-Ketoprostaglandin F2alpha	5280887	
16-Hydroxyhexadecanoic acid	10466	
1-Stearoyl-2-docosahexaenoyl-sn-glycero-3-phospho-(1'-sn-glycerol)	46891853	
Phosphatidylglycerol	24779551	C00344
Arachidonic acid	444899	C00219
Chimyl alcohol	72733	C13859
Cycloartenol	92110	C01902
Prostaglandin E2	5280360	C00584
Docosanoyl ethanolamide	3023585	
Docosapentaenoic acid	5497182	C16513
Eicosapentaenoic acid	446284	C06428
Elaidic acid	637517	C01712
Hexadecanedioic acid	10459	C19615
Isap acid	6439355	
Isobutyryl-carnitine	168379	
Isodesmosine	13811	
Isopimaric acid	442048	C09118
Lauric acid	3893	C02679
Leukotriene B4	5280492	C02165
Linoleic acid	5280450	C01595
Epi-lipoxin A4	42607306	
Palmitoylethanolamide	4671	C16512
Prostaglandin A1	5281912	C04685
Prostaglandin D3	5282260	C13802
Prostaglandin E3	5280937	C06439

Name	PubChem CID	KEGG ID
Prostaglandin H2	445049	C00427
Nicotinate ribonucleotide	121992	C01185
7-Dehydrocholesterol	439423	C01164
Butyrylcarnitine	439829	C02862
17-Oxo-5beta-androstan-3alpha-ol sulfate	9907417	
Androsterone sulfate	159663	
Lyxose	439240	C00476
4-Hydroxyphenylpyruvic acid	979	C01179
Phosphocholine	1014	C00588
Melibiose	440658	C05402
2-[[2-azanylpropanoyl]amino]propanoic acid	20112020	C00993
3-Ureidopropionic acid	111	C02642
Stearidonic acid	5312508	C16300
2,3-Dihydroxybenzoic acid	19	C00196
beta-Mannose	439680	C17207
Leukotriene E4	5280879	C05952
3-Hydroxypyruvic acid	964	C00168
Chitobiose	439544	C01674
2-Phenylethanol	6054	C05853
Lanosterol	246983	C01724
Aldehydogalactose	3037556	C01582
Malic acid	222656	C00149
10-Deacetylbaccatin III	154272	C11700
Selenomethionine	105024	C05335
Calcitriol	5280453	C01673
Porphobilinogen	1021	C00931
Melamine	7955	C08737
4-Isopropylbenzaldehyde	326	C06577
Butyryl-coenzyme A	122283	C00136
1-Palmitoylglycerol 3-phosphate	89566	C04036
Fructose 6-phosphate	69507	C00085
Tropine	8424	C00729
Orotidine	92751	C01103
Sphingosine	5280335	C00319
Stachyose	439531	C01613
Phytanic acid	26840	C01607
8,9-dihydroxyeicosatrienoic acid	5283144	C14773
4-Methyl-2-oxovaleric acid	70	C00233
1-O-phosphono-fructofuranose	439394	C01094
Phosphoribosyl pyrophosphate	7339	C00119
5,6-Epoxyeicosatrienoic acid	5283202	C14768
Isovaleryl-coenzyme A	165435	C02939
Sucrose	5988	C00089
Anabasine	2181	C06180
3-hydroxybutanoyl-coenzyme A	9543037	C01144

Name	PubChem CID	KEGG ID
Corticosterone	5753	C02140
Metanephrine	688084	C05588
5-Aminolevulinic acid	137	C00430
2-Deoxyribose	9828112	C01801
4-Hydroxybenzoic acid	135	C00156
Anthranilic acid	227	C00108
Lutein	5281243	C08601
Tetracycline	54675776	C06570
valine	6287	C00183
Stevioside	442089	C09189
Melatonin	896	C01598
4-Hydroxyphenyl acetate	96009	C13636
3-Methylsalicylic acid	6738	C14088
Pyridoxal	1050	C00250
Cysteinylglycine	439498	C01419
Ribose-5P	439167	C00117
15-oxo-5Z,8Z,11Z,13E-eicosatetraenoic acid	5280701	C04577
Glucosaminic acid	73563	C03752
Phosphonomycin	446987	C06454
Proline	8988	C00763
Cortisone	222786	C00762
2'-Deoxycytidine 5'-monophosphate	13945	C00239
Scopoletin	5280460	C01752
3-amino-4-methylpentanoic acid	2761558	C02486
Orotic acid	967	C00295
Lupanine	442956	C10772
Tryptophan	9060	C00525
Indole	798	C00463
Lipoxin A4	5280914	C06314
Ethylene glycol	174	C01380
2-Oxopentanoic acid	74563	C06255
5-Aminoimidazole-4-carboxamide	9679	C04051
Usnic acid	442614	C10101
Jasmonic acid	5281166	C08491
Raffinose	439242	C00492
Ectoine	126041	C06231
Albendazole	2082	C01779
3-Hydroxybutyric acid	92135	C01089
Lipoamide	863	C00248
2-Isopropyl-malic acid	5280523	C02504
Aspartic acid	83887	C00402
6-(Methylthio)purine	5778	C16614
Pentane-1,5-diamine	273	C01672
P-hydroxy-curucumin	5324476	C17742
Petromyzonol	5284054	C16258

Name	PubChem CID	KEGG ID
Pyroglutamic acid	439685	C02237
Gibberellin A4	443457	C11864
Maltol	8369	C11918
1,3-Dimethylurea	7293	C16364
Uridine diphosphate glucuronic acid	17473	C00167
Glucuronic Acid	94715	C00191
3-Methyl-2-oxobutanoic acid	49	C00141
Succinic acid	1110	C00042
3-(4-hydroxyphenyl)acrylaldehyde	641301	C05608
Quinolinic acid	1066	C03722
3,4-Dihydroxyphenylacetic acid	547	C01161
Dehydroascorbic acid	440667	C05422
Progesterone	5994	C00410
5-Hydroxytryptophan	439280	C00643
15-Ketoprostaglandin E	5280710	C04654

Table S6: Pathways represented in the experimental tests

Pathway name	N.	KEGG ID
Pyruvate metabolism	1	C00074
Proximal tubule bicarbonate reclamation	1	C00074
Inflammatory mediator regulation of TRP channels	3	C00219 C00584 C02165
Glutathione metabolism	1	C00005
Carbon fixation in photosynthetic organisms	1	C00074
Biosynthesis of plant hormones	2	C00074 C01902
Ovarian steroidogenesis	1	C00219
Biosynthesis of unsaturated fatty acids	4	C00219 C16513 C06428 C01595
Stilbenoid, diarylheptanoid and gingerol biosynthesis	1	C10443
Phosphotransferase system (PTS)	1	C00074
Neuroactive ligand-receptor interaction	2	C00584 C02165
Methane metabolism	1	C00074
Serotonergic synapse	4	C00219 C00584 C02165 C00427
African trypanosomiasis	1	C00584
Biosynthesis of secondary metabolites	6	C10443 C14156 C00074 C01902 C09118 C01164
Oxytocin signaling pathway	3	C00219 C00584 C00427
Pyrimidine metabolism	1	C00214
Central carbon metabolism in cancer	1	C00074
Vascular smooth muscle contraction	1	C00219
cAMP signaling pathway	1	C00584
Phototransduction - fly	1	C00219
Biosynthesis of amino acids	1	C00074
Biosynthesis of phenylpropanoids	1	C00074
Platelet activation	2	C00219 C00427
Steroid biosynthesis	2	C01902 C01164
Retinol metabolism	1	C16681
Eicosanoids	3	C00219 C02165 C00427
Nicotinate and nicotinamide metabolism	1	C01185
Linoleic acid metabolism	2	C00219 C01595
Arachidonic acid metabolism	4	C00219 C00584 C02165 C00427
Fatty acid biosynthesis	1	C02679
Glycolysis / Gluconeogenesis	1	C00074
Phosphonate and phosphinate metabolism	1	C00074
Long-term depression	1	C00219
Cutin, suberine and wax biosynthesis	1	C19615
Microbial metabolism in diverse environments	1	C00074
Leishmaniasis	2	C00219 C00584
Thyroid hormone synthesis	1	C00005
Insect hormone biosynthesis	1	C01164

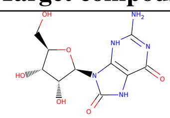
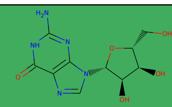
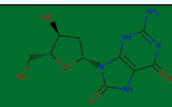
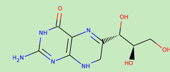
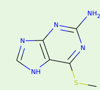
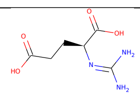
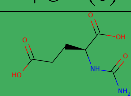
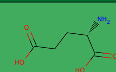
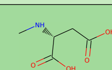
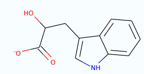
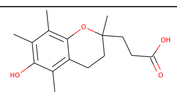
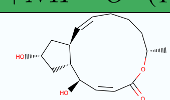
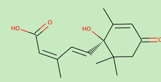
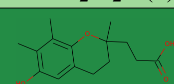
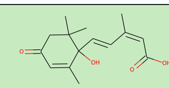
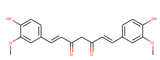
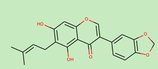
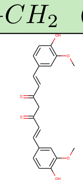
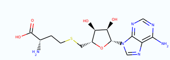
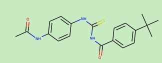
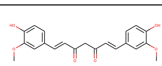
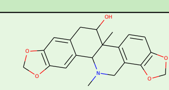
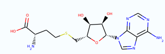
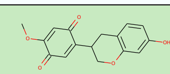
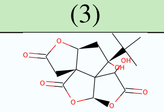
Pathway name	N.	KEGG ID
Ubiquinone and other terpenoid-quinone biosynthesis	1	C14156
Photosynthesis	1	C00005
Fc gamma R-mediated phagocytosis	1	C00219
Fc epsilon RI signaling pathway	1	C00219
Biosynthesis of alkaloids derived from shikimate pathway	1	C00074
Phenylalanine, tyrosine and tryptophan biosynthesis	1	C00074
Carbon metabolism	1	C00074
Diterpenoid biosynthesis	1	C09118
Biosynthesis of plant secondary metabolites	3	C00074 C01902 C01595
GnRH signaling pathway	1	C00219
PPAR signaling pathway	1	C02165
Rheumatoid arthritis	1	C00584
Carbon fixation pathways in prokaryotes	1	C00074
Retrograde endocannabinoid signaling	2	C00219 C00427
Bile secretion	2	C00584 C02165
Amoebiasis	2	C00219 C00584
Pathways in cancer	1	C00584
Prostaglandins	1	C00427
Biosynthesis of terpenoids and steroids	2	C00074 C01902
Citrate cycle (TCA cycle)	1	C00074

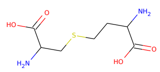
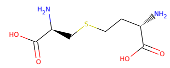
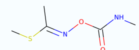
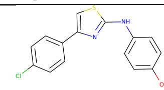
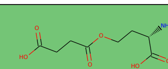
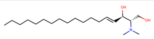
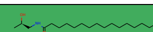
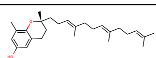
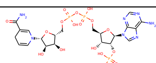
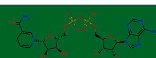
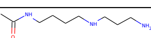
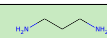
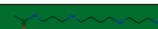
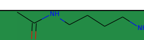
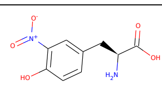
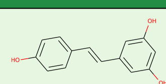
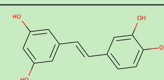
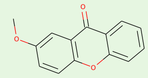
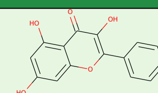
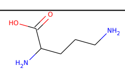
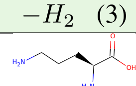
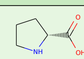
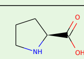
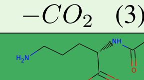
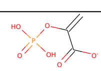
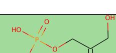
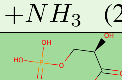
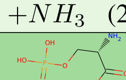
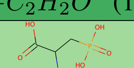
Pathway name	N.	HMDB ID
Pyrimidine Nucleosides and Analogues	1	273
Organic Oxoanionic Compounds	1	263
Prenol Lipids	6	30008 6218 3550 6220 36591 36811
Carboxylic Acids and Derivatives	1	672
Lineolic Acids and Derivatives	1	673
Fatty Acid Esters	1	736
Amino Acids and Derivatives	2	1904 739
Glycerophospholipids	2	10614 10604
Fatty Acids and Conjugates	5	31072 1043 6528 1999 638
Steroids and Steroid Derivatives	2	32 2759
Purine Nucleotides	1	221
Pyridine Nucleotides	1	1132
Purine Nucleosides and Analogues	1	2044
Benzopyrans	1	1518
Fatty Alcohols	1	6294
Eicosanoids	5	4240 1220 1085 12587 1381

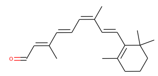
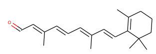
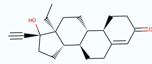
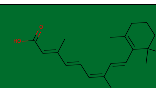
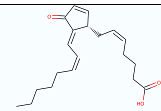
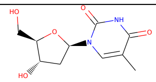
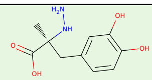
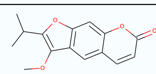
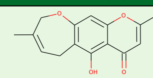
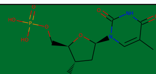
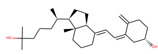
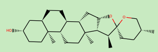
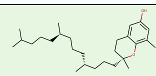
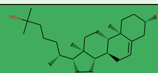
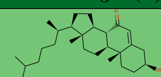
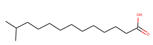
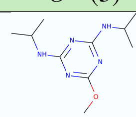
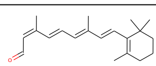
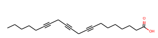
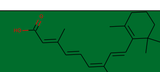
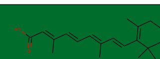
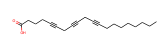
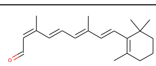
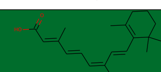
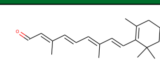
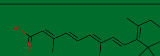
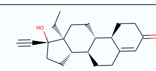
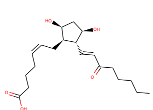
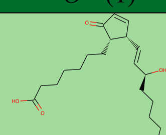
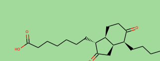
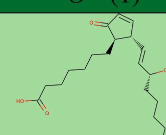
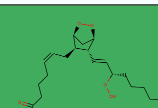
Table S7: Experimental validation

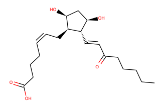
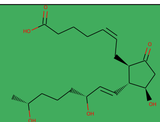
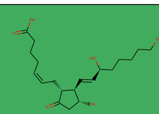
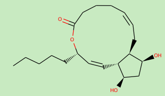
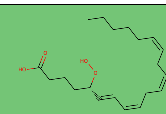
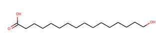
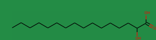
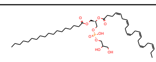
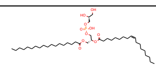
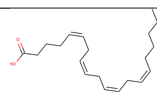
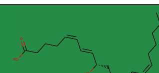
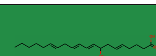
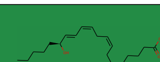
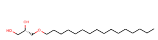
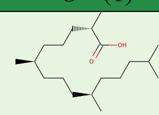
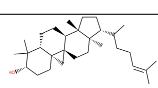
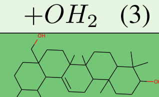
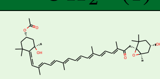
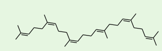
Structural similarity with target:

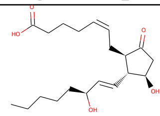
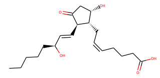
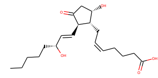
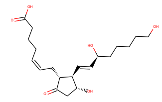
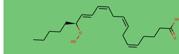
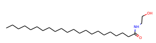
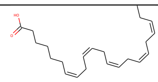
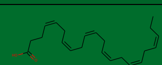
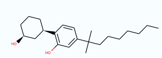
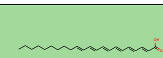
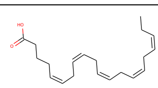
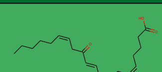
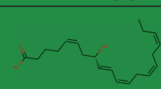
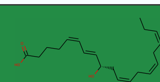
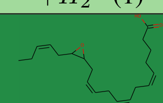
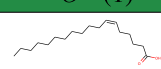
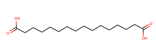
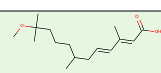
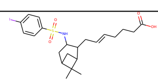
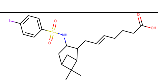
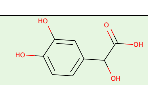
0.0 - 0.1	0.1 - 0.2	0.2 - 0.3	0.3 - 0.4	0.4 - 0.5	0.5 - 0.6	0.6 - 0.7	0.7 - 0.8	0.8 - 0.9	0.9 - 1.0

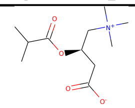
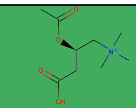
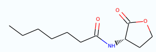
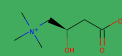
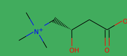
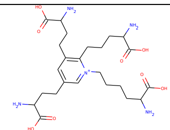
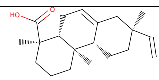
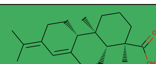
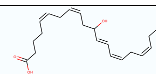
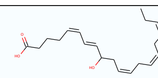
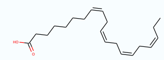
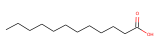
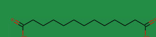
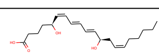
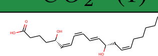
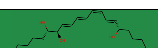
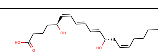
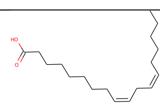
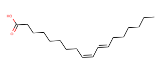
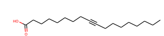
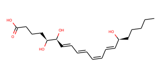
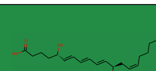
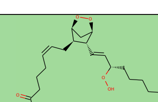
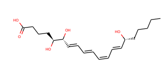
Target compound	**Candidate neighbors**			
 (1) 65131 (-),10,20,40	 6802 0.581 $+O$ (1)	 73318 0.581 $+O$ (1)	 65074 0.146 $+CO_2$ (3)	 70976 0.030 $+C_8H_6O$ (1)
 (2) 135313 (+),10,20,40	 121396 0.213 $+NH-O$ (1)	 33032 0.103 $+CH_2N_2$ (1)	 22880 0.075 $+CH_2N_2$ (2)	 92904 0.073 $-O$ (3)
 (3) 9943542 (+),10,20,40	 5287620 0.282 $-H_2$ (3)	 643732 0.144 $+CH_2$ (3)	 133098 0.135 $+CH_2$ (1)	 5375199 0.091 $+CH_2$ (3)
 (4) 5281767 (-),10,20,40	 5810067 0.312 $+H_2$ (3)	 969516 0.192 –	 439155 0.096 $-O$ (3)	 1013376 0.049 $+CH_3-N$ (3)
 (5) 5281767 (+),10,20,40	 185284 0.224 $+O-NH$ (3)	 439155 0.176 $-O$ (3)	 100072 0.100 $+C_5H_6O$ (3)	 25090658 0.069 $+C_6H_2-O_2$ (3)

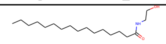
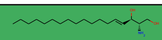
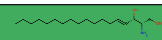
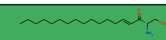
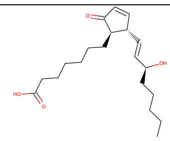
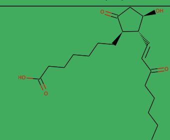
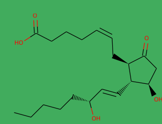
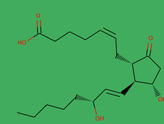
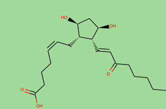
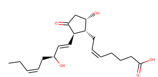
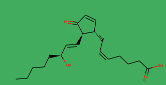
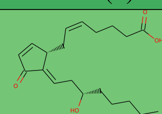
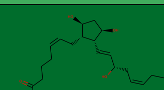
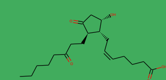
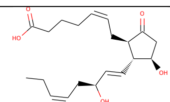
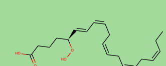
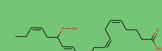
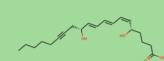
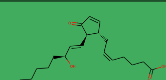
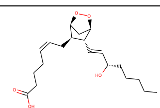
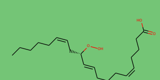
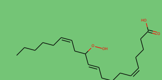
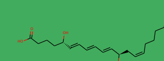
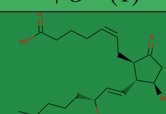
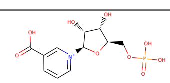
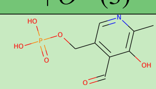
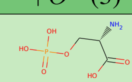
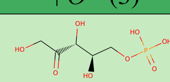
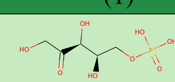
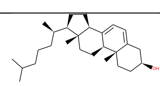
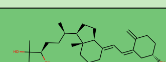
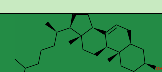
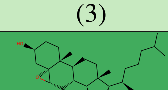
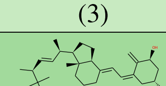
Target compound	**Candidate neighbors**			
 (6) 834 (+),10,20,40	 439258 0.411 –	 5353758 0.071 $+C_2H_4O_2$ (3)	 753704 0.068 –	 439406 0.066 $+HNS-CO_2$ (2)
 (7) 5282309 (+),20	 5283454 0.499 $+H_2$ (2)	 53394482 0.407 $+OH_2$ (2)	 18395309 0.316 $+CH_2$ (2)	 18431802 0.267 $+CH_2$ (2)
 (8) 5282350 (-),10,20,40	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –
 (9) 5884 (+),10,20,40	 439153 0.331 $+HPO_3$ (1)	- 0.000 –	- 0.000 –	- 0.000 –
 (10) 123689 (+),10,20,40	 1102 0.245 $+C_2H_2O$ (1)	 428 0.194 $+C_6H_{11}NO$ (1)	 916 0.165 $-C_3H_7N$ (1)	 122356 0.095 $+C_3H_7N$ (1)
 (11) 65124 (+),10,20,40	 445154 0.215 $-H_2$ (3)	 667639 0.176 $-OH_2$ (3)	 71034 0.139 – (3)	 5281616 0.108 $-CO_2$ (3)
 (12) 389 (+),10,20,40	 6262 0.645 –	 8988 0.323 $+NH_3$ (2)	 145742 0.323 $+NH_3$ (2)	 439232 0.313 $-C_2H_2O$ (1)
 (13) 1005 (-),10,20,40	 729 0.387 $-H_2$ (2)	 439183 0.369 $-OH_2$ (2)	 68841 0.336 $-NH_3$ (2)	 3857 0.258 $+NH_3-O$ (2)

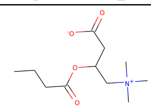
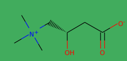
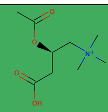
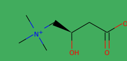
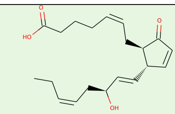
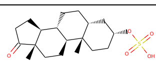
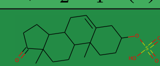
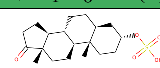
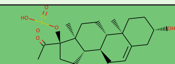
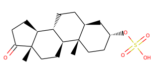
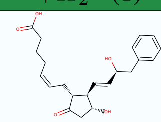
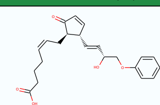
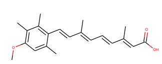
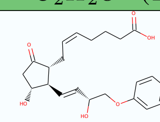
Target compound	Candidate neighbors			
 (14) 6436082 (+),10,20,40	 638015 0.518 –	 13109 0.446 $-CO$ (3)	 449171 0.442 $-O$ (1)	 5311211 0.405 $-O_2$ (3)
 (15) 5789 (+),10,20,40	 34359 0.259 $+O$ (3)	 8616 0.193 $-O$ (3)	 3646533 0.181 $-O$ (3)	 9700 0.150 $-HPO_3$ (1)
 (16) 5283731 (+),10,40	 91439 0.578 $-O$ (3)	 92094 0.573 $-H_2$ (3)	 65094 0.547 $-H_2$ (1)	 91474 0.518 – (2)
 (17) 520298 (-),10,20,40	 4928 0.195 $+NH_3-O$ (3)	- 0.000 –	- 0.000 –	- 0.000 –
 (18) 6436079 (-),10,20,40	 9543604 0.806 $-O$	 449171 0.741 $-O$ (1)	 444795 0.709 $-O$ (1)	 1781 0.532 $-O$
 (19) 6436079 (+),10,20,40	 449171 0.627 $-O$ (1)	 638015 0.569 –	 444795 0.513 $-O$ (1)	 13109 0.509 $-CO$ (3)
 (20) 5280887 (-),10,20,40	 6473771 0.949 $+O$ (1)	 35027042 0.873 $+O$ (2)	 5283050 0.873 $+O$ (1)	 5280883 0.800 $-O$ (2)

Target compound	**Candidate neighbors**			
 (21) 5280887 (+),10,20,40	 35022148 0.697 $-O$ (1)	 5283034 0.697 $-O$ (1)	 5283073 0.697 $+O$ (2)	 5280778 0.666 $+O$ (3)
 (22) 10466 (-),20	 92836 0.250 $-$ (1)	- 0.000 $-$	- 0.000 $-$	- 0.000 $-$
 (23) 46891853 (+),20	- 0.000 $-$	- 0.000 $-$	- 0.000 $-$	- 0.000 $-$
 (24) 24779551 (+),20	- 0.000 $-$	- 0.000 $-$	- 0.000 $-$	- 0.000 $-$
 (25) 444899 (-),10,20,40	 5283166 0.871 $-O$ (1)	 5283167 0.835 $-O$ (1)	 1898 0.835 $-O$ (1)	 5280724 0.799 $-O$ (1)
 (26) 72733 (+),20	 10936445 0.416 $-CH_2$ (1)	 9971931 0.366 $+OH_2$ (3)	 1380 0.351 $-CH_2$ (1)	 11825433 0.284 $-H_3PO$ (1)
 (27) 92110 (+),20	 72326 0.230 $-$ (3)	 3266408 0.144 $-$ (3)	 5281239 0.101 $-$ (3)	 5280370 0.097 $+O$ (3)

Target compound	Candidate neighbors			
 (28) 5280360 (-),10,20,40	 448457 0.933 –	 5283094 0.896 –	 5283034 0.896 $-O$	 5280893 0.896 $+O$ (3)
 (29) 3023585 (+),20	 3787294 0.178 $+C_2H_4$ (1)	- 0.000 –	- 0.000 –	- 0.000 –
 (30) 5497182 (-),10,20,40	 445580 0.544 $+H_2$ (1)	 53394526 0.544 $-H_2$ (1)	 12788231 0.544 $-H_2$ (3)	 175542 0.521 $+H_2$ (1)
 (31) 446284 (-),10,20,40	 5283162 0.780 $-O$ (1)	 5283186 0.780 $-O$ (1)	 5283187 0.746 $-O$ (1)	 16061088 0.712 $-O$ (1)
 (32) 637517 (-),10,20	 445639 1.000 –	 5281125 1.000 –	 6436624 0.740 – (1)	- 0.000 –
 (33) 10459 (-),20	 13849 0.510 $+CO_2$ (1)	 5353760 0.316 – (3)	- 0.000 –	- 0.000 –
 (34) 6439355 (-),10,20,40	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –
 (35) 6439355 (+),10,20,40	 85782 0.740 – (3)	- 0.000 –	- 0.000 –	- 0.000 –

Target compound	Candidate neighbors			
 (36) 168379 (+),20	 439756 0.147 $+C_2H_4$ (1)	 443437 0.144 $+OH_2$ (3)	 10917 0.108 $+C_4H_6O$ (1)	 2724480 0.101 $+C_4H_6O$ (1)
 (37) 13811 (+),20	- 0.000 —	- 0.000 —	- 0.000 —	- 0.000 —
 (38) 442048 (-),10,20,40	 221118 0.668 — (1)	 16061130 0.668 $-O$ (3)	 16061129 0.668 $-O$ (3)	 11722594 0.473 $-H_2$ (3)
 (39) 3893 (-),10,20,40	 10458 0.333 $-CO_2$ (1)	- 0.000 —	- 0.000 —	- 0.000 —
 (40) 5280492 (-),20	 35025115 0.916 —	 5280915 0.916 $-O$ (1)	 5283130 0.879 —	 1898 0.843 $+O$ (1)
 (41) 5280450 (-),10,20	 5280644 0.960 —	 1449 0.774 —	 5282796 0.740 —	 68167 0.740 —
 (42) 42607306 (-),20	 5283182 0.834 $+O$ (1)	 5283128 0.800 $+O$ (1)	 5280883 0.800 $-O$ (3)	 9841438 0.767 —

Target compound	Candidate neighbors			
 (43) 4671 (+),10,20,40	 5280335 0.640 − (1)	 11130338 0.640 − (1)	 14455158 0.510 $-CH_2$ (1)	 9839212 0.408 $+H_2$ (1)
 (44) 5281912 (-),10,20,40	 5280710 0.904 $-O$ (1)	 5283086 0.830 $-O$ (1)	 10247450 0.830 $-O$ (1)	 5283215 0.794 $-O$ (1)
 (45) 5282260 (-),20	 5280884 0.758 $+O$ (1)	 5280885 0.758 $+O$ (1)	 5283217 0.665 $-H_2$ (1)	 5283036 0.665 $-H_2$ (1)
 (46) 5280937 (-),20	 5283184 0.847 $+O$ (3)	 5283196 0.847 $+O$ (3)	 5283132 0.847 $+O$ (3)	 5280884 0.742 $+O$ (1)
 (47) 445049 (-),10,20,40	 5280892 0.939 $+O$ (3)	 5283175 0.939 $+O$ (3)	 5283128 0.901 $+O$ (3)	 5283086 0.864 − (1)
 (48) 121992 (-),10,20,40,IP	 1051 0.124 $+C_3H_4O_3$ (3)	 68841 0.100 $+C_8H_6O_3$ (3)	 439184 0.100 $+C_6H_3NO$ (3)	 439190 0.088 $+C_6H_3NO$ (3)
 (49) 439423 (+),20,40,IP	 5281059 0.453 $-O_2$ (2)	 65728 0.387 $-H_2$ (1)	 227037 0.299 $-O_H2$ (1)	 9547243 0.283 $-CO_2$ (1)

Target compound	Candidate neighbors			
 (50) 439829 (+),10,20,40,IP	 2724480 0.145 $+C_4H_6O$ (1)	 439756 0.133 $+C_2H_4$ (1)	 10917 0.099 $+C_4H_6O$ (1)	 5283112 0.096 $-$ (3)
 (51) 9907417 (-),10,20,40,IP	 12594 0.249 $+H_2$ (1)	 5244 0.226 $+H_2$ (1)	 159663 0.192 $-$	 10201764 0.122 $-C_2H_2O$ (1)
 (52) 159663 (+),10,20,40,IP	 5283067 0.251 $-H_2$ (3)	 12767333 0.199 $-$ (3)	 5284513 0.157 $+CO_2$	 13174090 0.085 $-OH_2$ (3)

Legend:

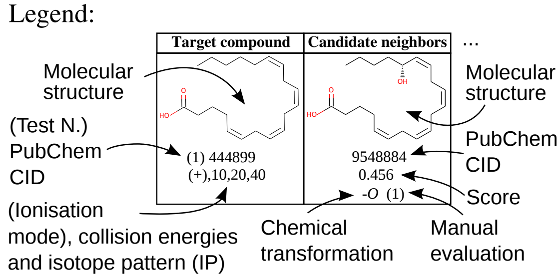


Table S8: Pathways represented in the leave-one-out cross-validation tests

Pathway name	N.	KEGG ID
Biosynthesis of alkaloids derived from histidine and purine	3	C00149 C00119 C00042
Degradation of aromatic compounds	6	C06577 C00156 C00108 C14088 C00042 C0116
Toxoplasmosis	1	C06314
Inflammatory mediator regulation of TRP channels	1	C14768
Inflammatory mediator regulation of TRP channels	1	C1476
Biosynthesis of plant hormones	5	C00149 C00119 C00108 C08491 C00042
Biosynthesis of unsaturated fatty acids	1	C01595
Acridone alkaloid biosynthesis	1	C00108
Propanoate metabolism	1	C00042
Aminoacyl-tRNA biosynthesis	1	C00183
Phenylpropanoid biosynthesis	2	C01752 C05608
Naphthalene degradation	1	C14088
Progesterone androgen and estrogen receptor agonists/antagonists	1	C00410
Inositol phosphate metabolism	1	C00191
Biosynthesis of alkaloids derived from terpenoid and polyketide	2	C00149 C00042
Synthesis and degradation of ketone bodies	1	C01089
Oocyte meiosis	1	C00410
Biosynthesis of type II polyketide products	1	C06570
Galactose metabolism	4	C05402 C01613 C00089 C00492
Starch and sucrose metabolism	3	C00089 C00167 C00191
Phosphonate and phosphinate metabolism	1	C06454
Arginine and proline metabolism	1	C00763
AMPK signaling pathway	1	C00085
2-Oxocarboxylic acid metabolism	5	C01179 C00233 C00183 C02504 C00141
Dioxin degradation	1	C00108
Benzoate degradation	3	C00196 C01144 C00156
Ubiquinone and other terpenoid-quinone biosynthesis	2	C01179 C00156
Ascorbate and aldarate metabolism	3	C00167 C00191 C05422
Histidine metabolism	1	C00119
Prolactin signaling pathway	1	C00410
Carbon metabolism	8	C00168 C00149 C00136 C00085 C00119 C01144 C00117 C00042
Choline metabolism in cancer	1	C00588

Pathway name	N.	KEGG ID
D-Alanine metabolism	1	C00993
Bisphenol degradation	2	C00156 C13636
Vitamin B6 metabolism	1	C00250
Alanine aspartate and glutamate metabolism	2	C00402 C00042
Glycine serine and threonine metabolism	3	C00168 C00430 C06231
Toluene degradation	1	C00156
Vitamin digestion and absorption	1	C00250
Oxidative phosphorylation	1	C00042
Proximal tubule bicarbonate reclamation	1	C00149
Purine metabolism	3	C00119 C00117 C04051
Fatty acid metabolism	2	C00136 C01144
Glutathione metabolism	3	C01419 C01672 C05422
Caffeine metabolism	1	C16364
Ovarian steroidogenesis	2	C14768 C00410
Sulfur metabolism	1	C00042
GABAergic synapse	1	C00042
Fc epsilon RI signaling pathway	1	C05952
Phosphotransferase system (PTS)	3	C01094 C00089 C03752
Methane metabolism	4	C01179 C00168 C00149 C00085
Amphetamine addiction	1	C01161
Selenocompound metabolism	1	C05335
Central carbon metabolism in cancer	4	C00149 C00085 C00183 C00042
Vascular smooth muscle contraction	1	C14768
Tryptophan metabolism	6	C01144 C00108 C01598 C00463 C03722 C00643
Carotenoid biosynthesis	1	C08601
Tuberculosis	1	C01673
Alcoholism	1	C01161
Dopaminergic synapse	1	C01161
Valine leucine and isoleucine degradation	4	C00233 C02939 C00183 C00141
Porphyrin and chlorophyll metabolism	2	C00931 C00430
Glucocorticoid and meneralocorticoid receptor agonists/antagonists	2	C00762 C00410
Lysine degradation	3	C00136 C01144 C01672
Microbial metabolism in diverse environments	22	C01179 C00196 C00168 C00149 C08737 C06577 C00085 C01094 C00119 C01144 C00156 C00108 C13636 C14088 C00250 C00117 C03752 C01380 C06231 C16364 C00042 C01161
Thyroid hormone synthesis	1	C00117
Fructose and mannose metabolism	1	C01094
Penicillin and cephalosporin biosynthesis	1	C00183

Pathway name	N.	KEGG ID
Biosynthesis of secondary metabolites	29	C01179 C00196 C00168 C01674 C01724 C00149 C11700 C00931 C00729 C00233 C00119 C06180 C00430 C00156 C00108 C08601 C06570 C00183 C09189 C00117 C00463 C02504 C01672 C17742 C11864 C00167 C00141 C00042 C05608
Aminobenzoate degradation	4	C00196 C01144 C00156 C00108
Renal cell carcinoma	1	C00149
Butanoate metabolism	4	C00136 C01144 C01089 C00042
Glyoxylate and dicarboxylate metabolism	5	C00168 C00149 C00136 C01380 C00042
Carbon fixation pathways in prokaryotes	3	C00149 C01144 C00042
Progesterone-mediated oocyte maturation	1	C00410
Citrate cycle (TCA cycle)	2	C00149 C00042
Benzoic acid family	2	C00156 C00108
Pentose and glucuronate interconversions	3	C00476 C00167 C00191
Polyketide sugar unit biosynthesis	1	C11918
Carbon fixation in photosynthetic organisms	3	C00149 C00085 C00117
Endocrine and other factor-regulated calcium reabsorption	1	C01673
beta-Alanine metabolism	2	C02642 C03722
Neuroactive ligand-receptor interaction	3	C05952 C01598 C06314
Carbohydrate digestion and absorption	1	C00089
Biosynthesis of siderophore group nonribosomal peptides	1	C00196
Valine leucine and isoleucine biosynthesis	4	C00233 C00183 C02504 C00141
Biosynthesis of alkaloids derived from shikimate pathway	6	C01179 C00149 C00108 C00183 C00042 C00643
Diterpenoid biosynthesis	2	C11700 C11864
Novobiocin biosynthesis	1	C01179
Aldosterone-regulated sodium reabsorption	1	C00762
Chlorocyclohexane and chlorobenzene degradation	1	C00042
Arachidonic acid metabolism	5	C05952 C14773 C14768 C04577 C06314
Sphingolipid metabolism	1	C00319
Amino sugar and nucleotide sugar metabolism	3	C01674 C00167 C00191
Steroid biosynthesis	2	C01724 C01673
Bile secretion	1	C06570
D-Glutamine and D-glutamate metabolism	1	C02237
Serotonergic synapse	3	C14773 C14768 C00643
Circadian entrainment	1	C01598

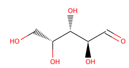
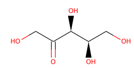
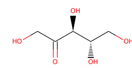
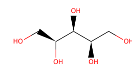
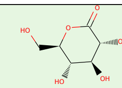
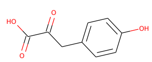
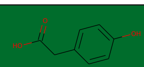
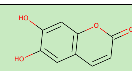
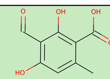
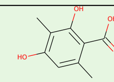
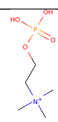
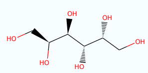
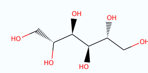
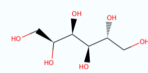
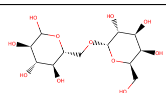
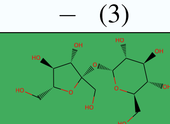
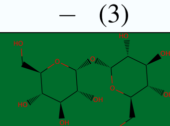
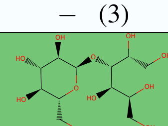
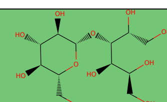
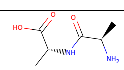
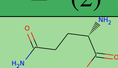
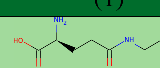
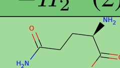
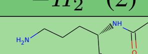
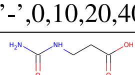
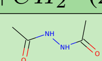
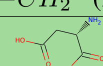
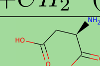
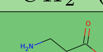
Pathway name	N.	KEGG ID
Steroid hormone biosynthesis	3	C02140 C00762 C00410
alpha-Linolenic acid metabolism	2	C16300 C08491
Pathways in cancer	3	C00149 C00762 C00410
Glycerophospholipid metabolism	1	C00588
Prostate cancer	2	C00762 C00410
Cocaine addiction	1	C01161
Vancomycin resistance	1	C00993
Biosynthesis of plant secondary metabolites	11	C01179 C00149 C01595 C00119 C00430 C00108 C00183 C08491 C01672 C00042 C03722
Peptidoglycan biosynthesis	1	C00993
Chloroalkane and chloroalkene degradation	1	C01380
Two-component system	2	C00149 C00042
Pyruvate metabolism	3	C00149 C02504 C00042
Isoquinoline alkaloid biosynthesis	1	C01179
Atrazine degradation	1	C08737
Xylene degradation	1	C06577
Tropane piperidine and pyridine alkaloid biosynthesis	4	C00729 C06180 C10772 C01672
Cyanoamino acid metabolism	1	C00183
Plant hormone signal transduction	2	C08491 C11864
Phenylalanine metabolism	3	C05853 C00156 C00042
Benzoxazinoid biosynthesis	1	C00463
Pyrimidine metabolism	5	C02642 C01103 C00119 C00239 C00295
cAMP signaling pathway	2	C01089 C00042
Drug metabolism - other enzymes	1	C16614
Biosynthesis of amino acids	8	C01179 C00233 C00119 C00108 C00183 C00117 C02504 C00141
Biosynthesis of phenylpropanoids	7	C00196 C00149 C00156 C00108 C01752 C00042 C05608
Tyrosine metabolism	4	C01179 C05588 C00042 C01161
Pantothenate and CoA biosynthesis	3	C02642 C00183 C00141
Prion diseases	1	C02140
ABC transporters	6	C05402 C01674 C00089 C00430 C00183 C00492
Nicotinate and nicotinamide metabolism	1	C03722
Linoleic acid metabolism	1	C01595
Fatty acid degradation	2	C00136 C01144
Glucosinolate biosynthesis	3	C00233 C00183 C00141
Biosynthesis of vancomycin group antibiotics	1	C01179
Stilbenoid diarylheptanoid and gingerol biosynthesis	1	C17742

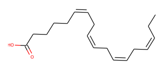
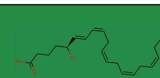
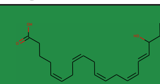
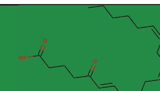
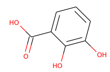
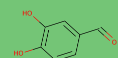
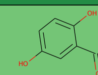
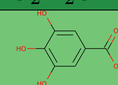
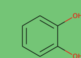
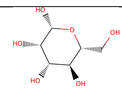
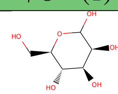
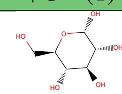
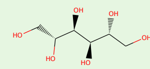
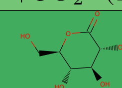
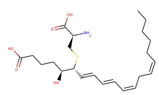
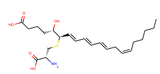
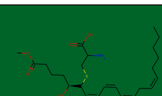
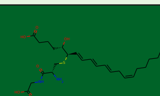
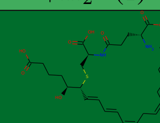
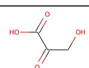
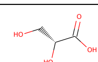
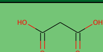
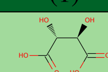
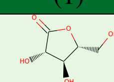
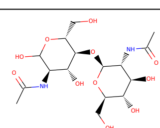
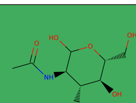
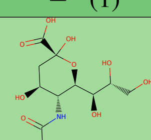
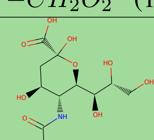
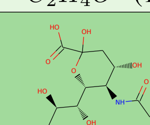
Pathway name	N.	KEGG ID
Phenylalanine tyrosine and tryptophan biosynthesis	5	C01179 C01094 C00119 C00108 C00463
Protein digestion and absorption	3	C00183 C00463 C01672
Tetracycline biosynthesis	1	C06570
Pentose phosphate pathway	4	C00119 C01801 C00117 C03752
Taste transduction	1	C00089
Biosynthesis of alkaloids derived from ornithine, lysine and nicotinic acid	7	C00149 C00729 C06180 C10772 C01672 C00042 C03722
Biosynthesis of type II polyketide backbone	1	C00136
Fatty acid elongation	1	C00136
Mineral absorption	2	C01673 C00183
Biosynthesis of terpenoids and steroids	4	C00149 C08601 C09189 C00042

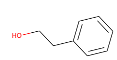
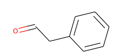
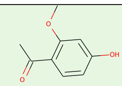
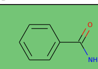
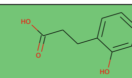
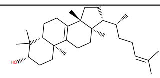
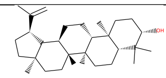
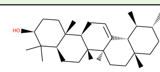
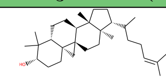
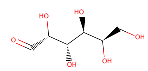
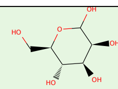
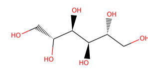
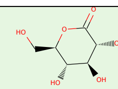
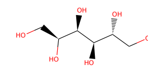
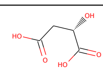
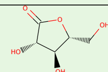
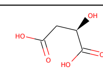
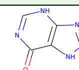
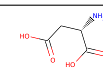
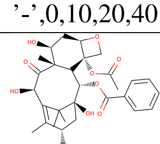
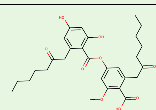
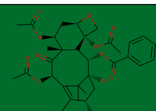
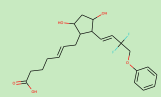
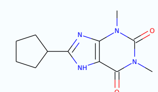
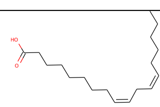
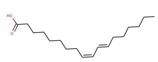
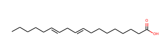
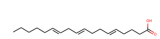
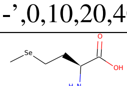
Table S9: Leave-one-out cross-validation

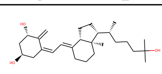
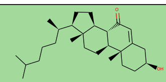
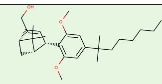
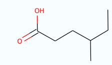
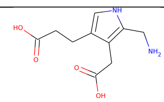
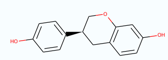
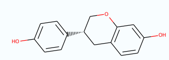
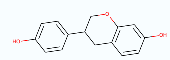
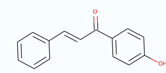
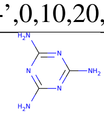
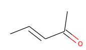
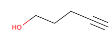
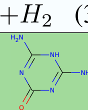
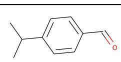
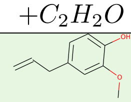
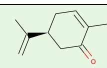
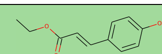
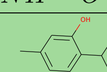
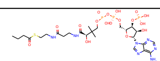
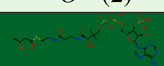
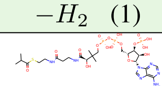
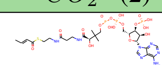
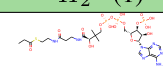
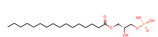
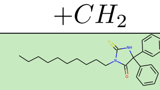
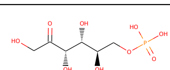
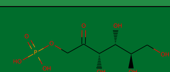
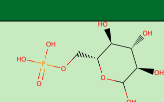
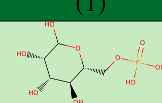
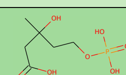
Structural similarity with target:

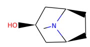
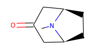
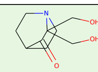
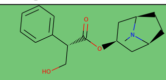
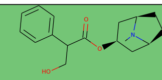
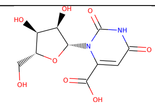
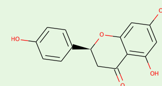
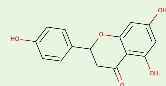
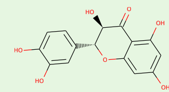
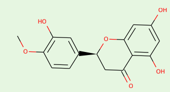
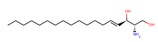
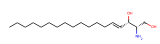
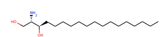
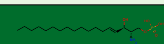
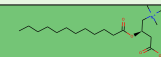
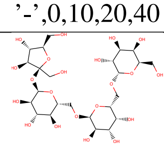
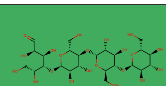
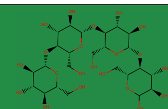
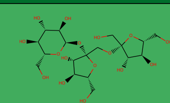
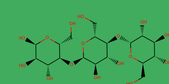
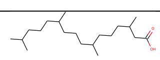
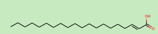
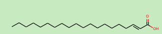
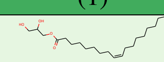
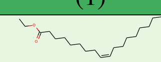
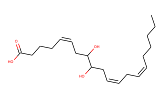
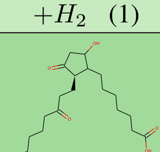
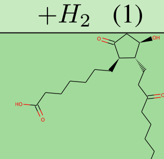
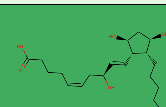
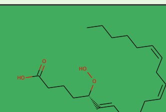
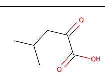
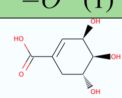
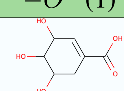
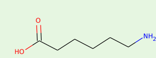
0.0 - 0.1	0.1 - 0.2	0.2 - 0.3	0.3 - 0.4	0.4 - 0.5	0.5 - 0.6	0.6 - 0.7	0.7 - 0.8	0.8 - 0.9	0.9 - 1.0

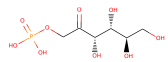
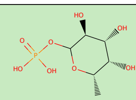
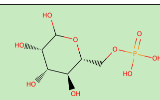
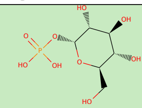
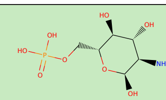
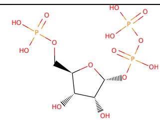
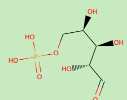
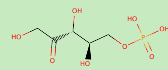
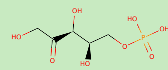
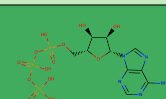
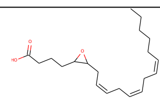
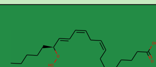
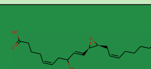
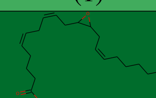
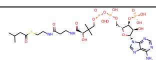
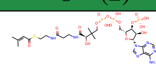
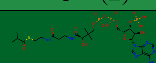
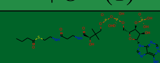
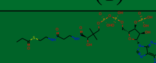
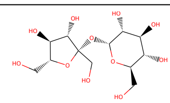
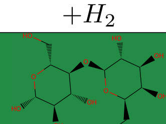
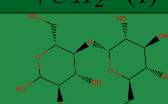
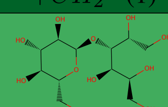
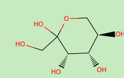
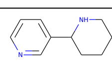
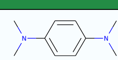
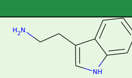
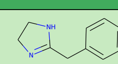
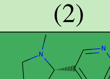
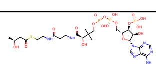
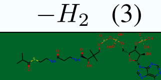
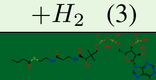
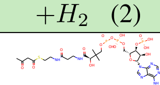
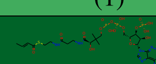
Target compound	**Candidate neighbors**			
 (1) 439240 '+',0,10,20,40 '-',0,10,20,40	 439204 0.663 $-$	 439245 0.555 $-$	 6912 0.392 $-H_2$	 7027 0.374 $-CO$ (1)
 (2) 979 '-',0,10,20,40	 127 0.472 $+CO$ (1)	 5281416 0.429 $+H_2$ (2)	 6710680 0.418 $-O$ (3)	 736228 0.409 $-H_2$ (3)
 (3) 1014 '+',0,10,20,40 '-',0	 11850 0.518 $-$ (3)	 6251 0.436 $-$ (3)	 5780 0.436 $-$ (3)	- 0.000 $-$
 (4) 440658 '-',0,10,20,40	 5988 0.365 $-$ (2)	 7427 0.346 $-$ (1)	 493591 0.330 $-H_2$ (2)	 160514 0.330 $-H_2$ (2)
 (5) 20112020 '+',0,10,20,40 '-',0,10,20,40	 5961 0.229 $+CH_2$ (2)	 439378 0.229 $-CH_2$ (2)	 145815 0.206 $+CH_2$ (2)	 439232 0.206 $-CH_2$ (2)
 (6) 111 '+',0,10,20,40	 72884 0.469 $+O$ (2)	 5960 0.451 $+NH-O$ (2)	 83887 0.397 $+NH-O$ (2)	 239 0.344 $+CHNO$ (2)

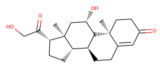
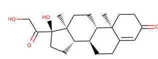
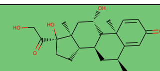
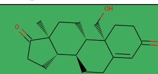
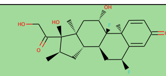
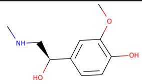
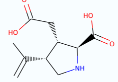
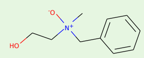
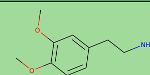
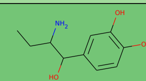
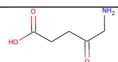
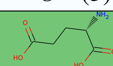
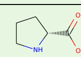
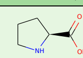
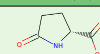
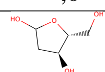
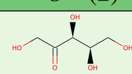
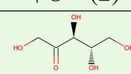
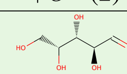
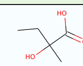
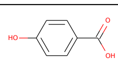
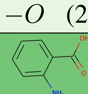
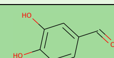
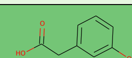
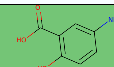
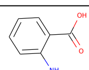
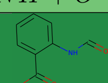
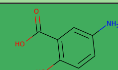
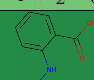
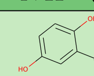
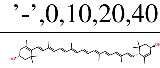
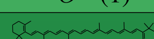
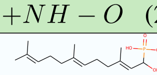
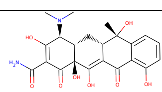
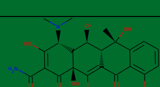
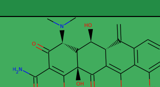
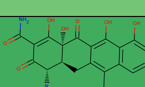
Target compound	**Candidate neighbors**			
 (7) 5312508 '+',0,10,20,40 '-',0,10,20,40	 128849 0.483 $-H_2$ (1)	 5283193 0.416 $-C_2H_2O$ (1)	 16061132 0.383 $-C_2H_2O$ (1)	 5283159 0.367 $-C_2H_2O$ (1)
 (8) 19 '-',0,10,20,40	 8768 0.769 $+O$ (1)	 70949 0.466 $+O$ (1)	 370 0.466 $-O$ (1)	 289 0.464 $+CO_2$ (1)
 (9) 439680 '-',0,10,20,40	 18950 0.726 –	 79025 0.693 –	 6251 0.611 $-H_2$ (1)	 7027 0.586 $+H_2$ (1)
 (10) 5280879 '+',0,10,20,40 '-',0,10,20,40	 5283136 0.403 –	 71684545 0.380 $-CH_2$ (1)	 5283135 0.271 $-C_2H_3NO$ (1)	 5280938 0.179 $-C_5H_7NO_3$ (1)
 (11) 964 '-',0,10,20,40	 439194 0.380 $-H_2$	 867 0.348 – (1)	 444305 0.241 $-CH_2O_2$ (1)	 17723 0.148 $-C_2H_4O$ (1)
 (12) 439544 '+',0,10,20,40 '-',0,10,20,40	 439174 0.099 $+C_8H_{13}NO_5$ (1)	 445063 0.061 $+C_5H_9NO_2$ (2)	 444885 0.061 $+C_5H_9NO_2$ (2)	 439197 0.061 $+C_5H_9NO_2$ (2)

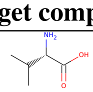
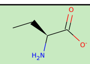
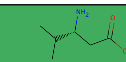
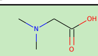
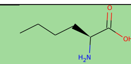
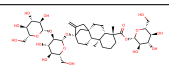
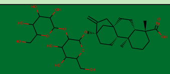
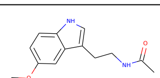
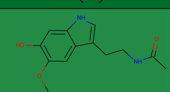
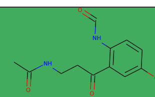
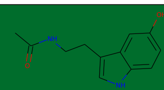
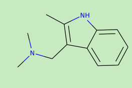
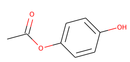
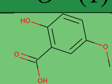
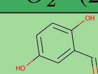
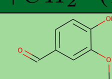
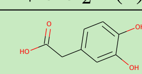
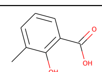
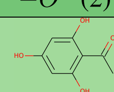
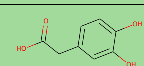
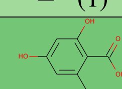
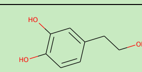
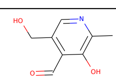
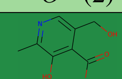
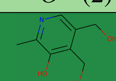
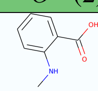
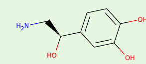
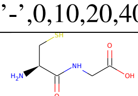
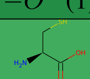
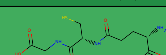
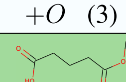
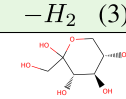
Target compound	Candidate neighbors			
 (13) 6054 '+',0,10,20,40	 998 0.404 $+H_2$	 529402 0.346 $-CO_2$ (2)	 2331 0.322 $CH_3 - N$ (1)	 873 0.314 $-CO_2$ (1)
 (14) 246983 '-',0,10,20,40	 259846 0.706 $-$	 73170 0.672 $-$	 92110 0.548 $-$	- 0.000 $-$
 (15) 3037556 '-',0,10,20,40	 18950 0.751 $-$ (1)	 6251 0.632 $-H_2$	 7027 0.632 $+H_2$ (1)	 5780 0.580 $-H_2$
 (16) 222656 '+',0,10,20,40 '-',0,10,20,40	 17723 0.298 $-CH_2$ (2)	 92824 0.240 $-$	 790 0.226 $-H_2$	 5960 0.121 $+NH - O$
 (17) 154272 '+',0,10,20,40 '-',0,10,20,40	 6710684 0.231 $+O$ (3)	 10908329 0.057 $-C_4H_4O_2$ (1)	 53394479 0.020 $+C_8H_6O_2$ (3)	 1917 0.012 $+C_{14}H_{12}N_6O_2$ (3)
 (18) 5280450 '+',0,10,20,40 '-',0,10,20,40	 5280644 0.774 $-$	 5282457 0.518 $-$	 128849 0.499 $+H_2$	 5282796 0.462 $-$
 (19) 105024 '+',0,10,20,40 '-',0,10,20,40	- 0.000 $-$	- 0.000 $-$	- 0.000 $-$	- 0.000 $-$

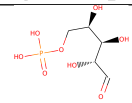
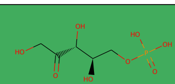
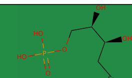
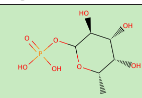
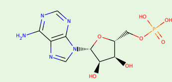
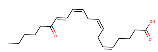
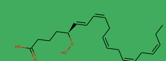
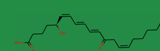
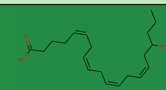
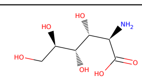
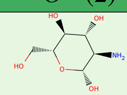
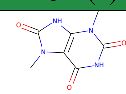
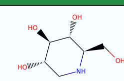
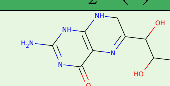
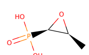
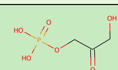
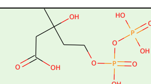
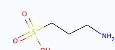
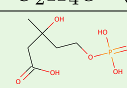
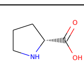
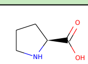
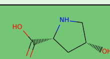
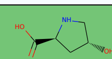
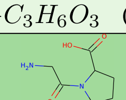
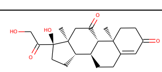
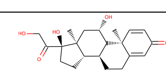
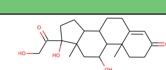
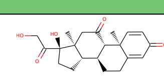
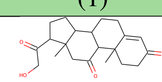
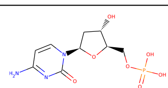
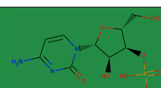
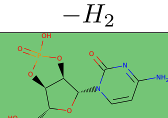
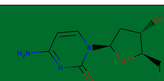
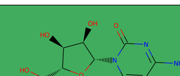
Target compound	Candidate neighbors			
 (20) 5280453 '+',0,10,20,40	 91474 0.548 $+O$ (2)	 11553430 0.222 $+H_2$ (3)	 15271 0.054 $+C_{20}H_{30}O$ (3)	- 0.000 $-$
 (21) 1021 '+',0,10,20,40 '-',0,10,20,40	 91469 0.374 $-O$ (3)	 6950272 0.356 $-O$ (3)	 382975 0.338 $-O$ (3)	 5282362 0.277 $+H_2$ (3)
 (22) 7955 '+',0,10,20,40	 12248 0.146 $+C_2H_2O$	 79346 0.129 $+C_2H_2O$	 5283349 0.117 $-CH_2$	 12583 0.083 $+NH-O$ (1)
 (23) 326 '+',0,10,20,40	 3314 0.493 $-O$ (2)	 16724 0.272 $-H_2$ (1)	 349172 0.234 $-CO_2$ (2)	 6989 0.231 $-H_2$ (1)
 (24) 122283 '-',0,10,20,40	 9543037 0.723 $-O$ (1)	 3036931 0.605 $-$	 5497143 0.579 $+H_2$	 439164 0.472 $+CH_2$
 (25) 89566 '+',0,10,20,40 '-',0,10,20,40	 78401 0.089 $+C_5H_8O_4$ (1)	 14900 0.073 $+HO_3P$ (1)	 44474572 0.052 $-C_5H_9NO_2$ (1)	 11502403 0.041 $+H_2$ (3)
 (26) 69507 '-',0,10,20,40	 439394 0.648 $-$ (1)	 5958 0.648 $-$ (1)	 65127 0.562 $-$ (1)	 476 0.545 $+O_2$ (1)

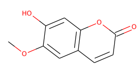
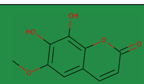
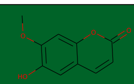
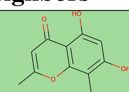
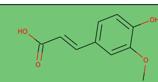
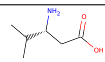
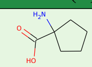
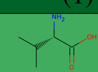
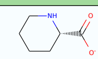
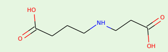
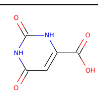
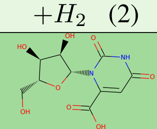
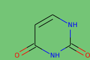
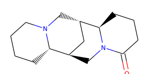
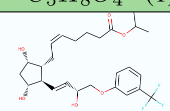
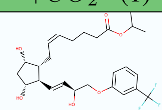
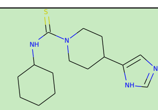
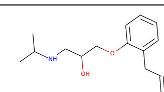
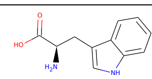
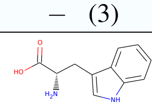
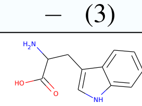
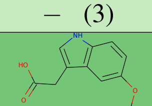
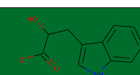
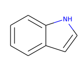
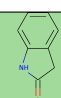
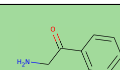
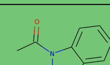
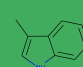
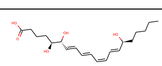
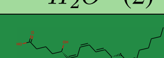
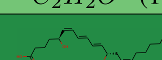
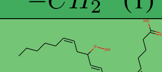
Target compound	**Candidate neighbors**			
 (27) 8424 '+',0,10,20,40	 79038 0.543 $+H_2$	 322968 0.342 $-CO_2$ (2)	 154417 0.117 $-C_9H_8O_2$ (1)	 174174 0.106 $-C_9H_8O_2$ (1)
 (28) 92751 '+',0,10,20,40 '-',0,10,20,40	 439246 0.309 $+O$ (3)	 932 0.309 $+O$ (3)	 439533 0.232 $-O$ (3)	 72281 0.229 $-CH_2$ (3)
 (29) 5280335 '+',0,10,20,40 '-',0,10,20,40	 11130338 0.548 $-$	 91486 0.499 $-H_2$	 5283560 0.301 $-HO_3P$ (1)	 168381 0.251 $-CO_2$ (2)
 (30) 439531 '+',0,10,20,40 '-',0,10,20,40	 439639 0.518 $-$ (1)	 439626 0.250 $-$ (1)	 440080 0.161 $+C_6H_{10}O_5$ (1)	 439586 0.107 $+C_6H_{10}O_5$ (1)
 (31) 26840 '+',0,10,20,40 '-',0,10,20,40	 3035221 0.573 $+H_2$ (1)	 6438157 0.548 $+H_2$ (1)	 5283468 0.208 $-CO_2$ (1)	 5363269 0.206 $+H_2$ (1)
 (32) 5283144 '+',0,10,20,40 '-',0,10,20,40	 71684618 0.569 $-O$ (1)	 165284 0.539 $-O$ (1)	 11245181 0.510 $-O$ (1)	 5280778 0.491 $+H_2$ (1)
 (33) 70 '-',0,10	 8742 0.489 $-CO_2$ (2)	 1094 0.388 $-CO_2$ (2)	 564 0.183 $-NH_3+O$ (1)	- 0.000 $-$

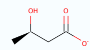
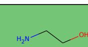
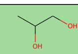
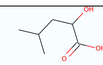
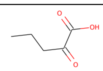
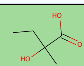
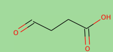
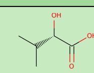
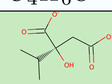
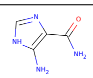
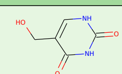
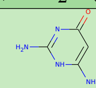
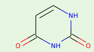
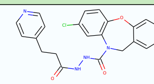
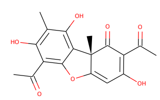
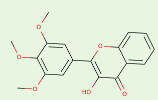
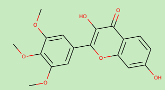
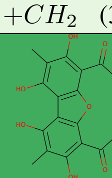
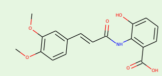
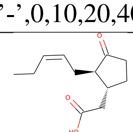
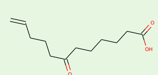
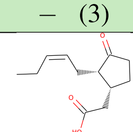
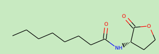
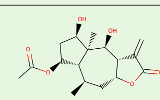
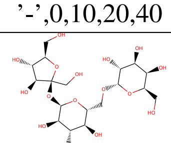
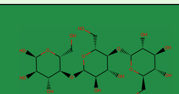
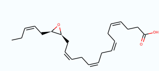
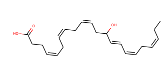
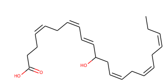
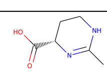
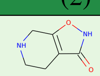
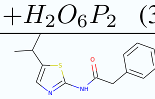
Target compound	**Candidate neighbors**			
 (34) 439394 '+',0,10,20,40 '-',0,10,20,40	 65369 0.403 $+O$ (1)	 65127 0.231 $-$ (1)	 65533 0.202 $-$ (1)	 440997 0.190 $+NH+O$ (1)
 (35) 7339 '-',0,10,20,40	 439167 0.178 $+H_2O_6P_2$ (1)	 439184 0.171 $+H_2O_6P_2$ (1)	 439190 0.171 $+H_2O_6P_2$ (1)	 5957 0.073 $+C_5H_3N_5-O$ (1)
 (36) 5283202 '+',0,10,20,40	 5280893 0.578 $-O$ (2)	 5280743 0.548 $-O$ (2)	 444899 0.462 $+O$ (1)	 5283204 0.410 $-$ (1)
 (37) 165435 '-',0,10,20,40	 9549326 0.530 $+H_2$	 3036931 0.416 $+CH_2$ (1)	 122283 0.416 $+CH_2$ (1)	 439164 0.206 $+C_2H_4$ (1)
 (38) 5988 '+',0,10,20,40 '-',0,10,20,40	 6134 0.647 $-$ (1)	 439186 0.502 $-$ (1)	 493591 0.355 $-H_2$ (1)	 439312 0.192 $+C_6H_{10}O_5$ (2)
 (39) 2181 '+',0,10,20,40	 7490 0.539 $-H_2$ (3)	 1150 0.493 $+H_2$ (3)	 5504 0.471 $+H_2$ (2)	 89594 0.441 $-$ (1)
 (40) 9543037 '-',0,10,20,40	 3036931 0.884 $+O$ (1)	 122283 0.884 $+O$ (1)	 92153 0.625 $+H_2$	 5497143 0.577 $+H_2O$ (1)

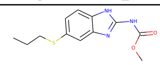
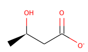
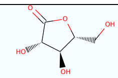
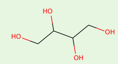
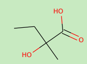
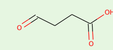
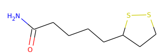
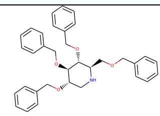
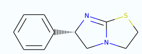
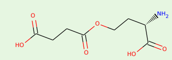
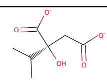
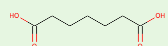
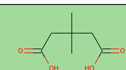
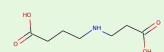
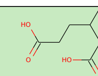
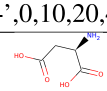
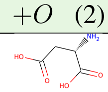
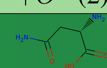
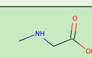
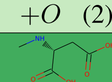
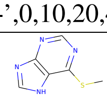
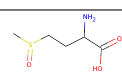
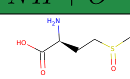
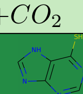
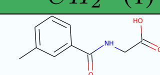
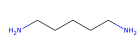
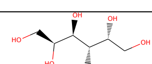
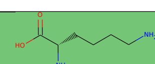
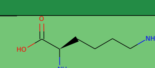
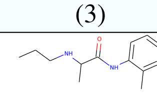
Target compound	Candidate neighbors			
 (41) 5753 '-',0,10,20,40	 440707 0.288 –	 6741 0.155 $-CO$ (1)	 252379 0.077 $+C_2H_4O$ (1)	 16490 0.031 $-O_4$ (2)
 (42) 688084 '+',0,10,20,40	 10255 0.708 $-O$ (3)	 264996 0.431 $+O$ (3)	 8421 0.431 $+O$ (2)	 18538 0.388 – (2)
 (43) 137 '+',0,10,20,40 '-',0	 33032 0.659 $-O$ (2)	 8988 0.597 $+O$ (2)	 145742 0.539 $+O$ (2)	 439685 0.402 $+H_2$ (1)
 (44) 9828112 '+',0,10,20,40	 439204 0.634 $-O$ (2)	 439245 0.605 $-O$ (2)	 439240 0.498 $-O$ (2)	 95433 0.498 $+O$ (3)
 (45) 135 '+',0,'-',0,10 20,40	 227 0.351 $-NH+O$ (1)	 8768 0.314 – (1)	 12122 0.236 $-CH_2$ (1)	 4075 0.206 $-NH$ (2)
 (46) 227 '+',0,10,20,40 '-',0,10,20,40	 101399 0.579 $-CO$ (1)	 4075 0.411 $-O$ (1)	 67069 0.362 $-CH_2$ (1)	 70949 0.251 $+NH-O$ (2)
 (47) 5281243 '+',0,10,20,40	 5280899 0.288 – (1)	 5280489 0.204 $+O_2$ (1)	 446925 0.191 $+O_2$ (2)	 5353399 0.012 $+C_{18}H_34O$ (3)
 (48) 54675776 '+',0,10,20,40 '-',0,10,20,40	 54675779 0.608 $-O$ (1)	 54675785 0.408 $+H_2$ (1)	 54675758 0.351 $+H_2O$ (1)	- 0.000 –

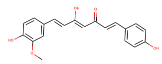
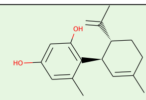
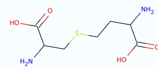
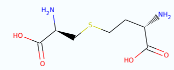
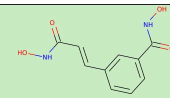
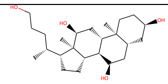
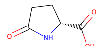
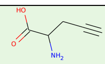
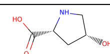
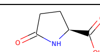
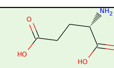
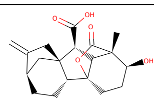
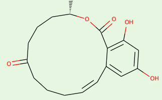
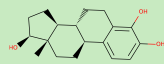
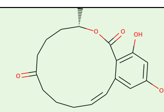
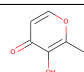
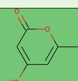
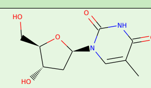
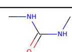
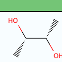
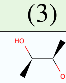
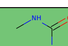
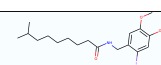
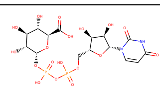
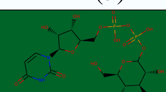
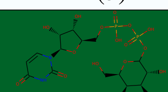
Target compound	Candidate neighbors			
 (49) 6287 '+',0,10,20,40 '-',0,10	 80283 0.386 $+CH_2$ (1)	 2761558 0.353 $-CH_2$ (1)	 673 0.353 $+CH_2$ (2)	 21236 0.337 $-CH_2$ (2)
 (50) 442089 '+',0,10,20,40 '-',0,10,20,40	 44386142 0.107 $+C_6H_{10}O_5$ (1)	- 0.000 $-$	- 0.000 $-$	- 0.000 $-$
 (51) 896 '+',0,10,20,40 '-',0,10,20,40	 1864 0.512 $-O$ (1)	 171161 0.298 $-O_2$ (2)	 903 0.256 $+CH_2$ (1)	 181513 0.164 $+CO_2$ (2)
 (52) 96009 '-',0,10,20,40	 75787 0.516 $-O$ (2)	 70949 0.370 $+CH_2$ (2)	 1183 0.365 $-$ (1)	 547 0.327 $-O$ (1)
 (53) 6738 '-',0,10,20,40	 68073 0.649 $-O$ (2)	 547 0.560 $-O$ (2)	 68072 0.560 $-O$ (2)	 82755 0.437 $-H_2$ (2)
 (54) 1050 '+',0,10,20,40 '-',0,10,20,40	 6723 0.605 $-O$ (1)	 1054 0.391 $-H_2$ (1)	 67069 0.350 $+O$ (3)	 439260 0.307 $-H_2$ (3)
 (55) 439498 '+',0,10,20,40 '-',0,10,20,40	 5862 0.128 $+C_2H_3NO$ (1)	 124886 0.088 $-C_5H_7NO_3$ (1)	 73917 0.076 $+O_2$ (3)	 439192 0.071 $-H_2$

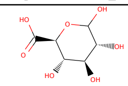
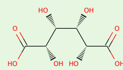
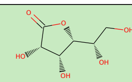
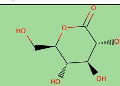
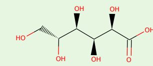
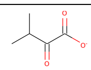
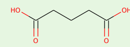
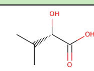
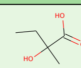
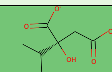
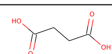
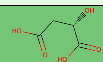
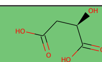
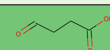
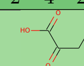
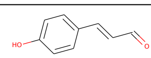
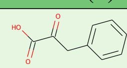
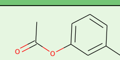
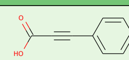
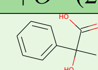
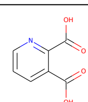
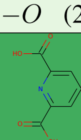
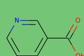
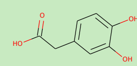
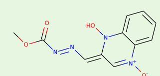
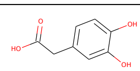
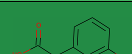
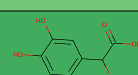
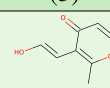
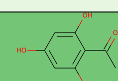
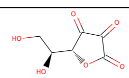
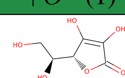
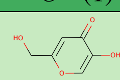
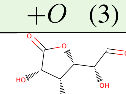
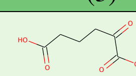
Target compound	Candidate neighbors			
 (56) 439167 '+',0,10,20,40 '-',0,10,20,40	 439184 0.516 $-$ (1)	 439288 0.427 $+O$ (1)	 65369 0.234 $-CH_2$ (2)	 6083 0.157 $-C_5H_3N_5+O$ (2)
 (57) 5280701 '+',0,10,20,40	 5283184 0.639 $-O$ (2)	 5280876 0.639 $-O$ (2)	 6442740 0.594 $-H_2$ (1)	 5283157 0.594 $-H_2$ (1)
 (58) 73563 '+',0,10,20,40	 441477 0.168 $+O$ (1)	 83126 0.101 $-NH+O$	 29435 0.086 $+O_2$ (3)	 252 0.084 $-C_2H_4O$ (3)
 (59) 446987 '-',0,10,20,40	 729 0.217 $-O_2$ (2)	 516 0.091 $-$ (3)	 1646 0.091 CH_3-N (3)	 476 0.030 $-C_3H_6O_3$ (3)
 (60) 8988 '+',0,10,20,40 '-',0,10,20	 145742 0.846 $-$	 440014 0.518 $-O$ (1)	 5810 0.462 $-O$ (1)	 79101 0.320 $-C_2H_3NO$ (1)
 (61) 222786 '-',0,10,20,40	 5755 0.490 $-$	 3640 0.475 $-H_2$	 5865 0.453 $+H_2$	 6290 0.336 $+O$
 (62) 13945 '+',0,10,20,40 '-',0,10,20,40	 66535 0.469 $-O$ (2)	 68934 0.408 $+H_2$ (1)	 13711 0.310 $+HO_3P$ (1)	 6175 0.144 $+HO_2P$ (1)

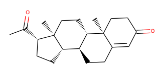
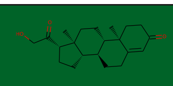
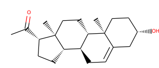
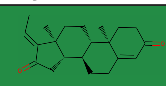
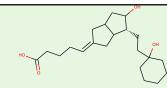
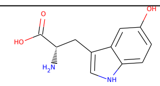
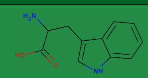
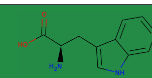
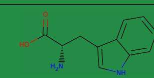
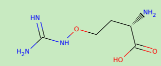
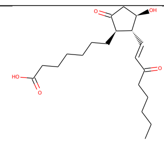
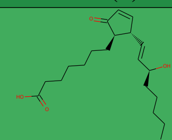
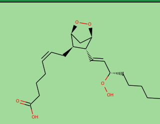
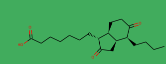
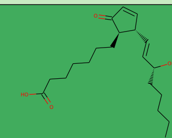
Target compound	Candidate neighbors			
 (63) 5280460 '+',0,10,20,40 '-',0,10,20,40	 5273569 0.666 $-O$ (1)	 69894 0.548 $-$ (1)	 5318562 0.440 $-CH_2$ (3)	 445858 0.428 $-H_2$ (1)
 (64) 2761558 '+',0,10,20,40 '-',0,10	 2901 0.445 $+H_2$ (2)	 6287 0.380 $+CH_2$ (1)	 439227 0.347 $+H_2$ (2)	 2572 0.346 $-CO_2$ (2)
 (65) 967 '+',0,'-',0,10 20,40	 92751 0.163 $-C_5H_8O_4$ (1)	 1174 0.156 $+CO_2$ (1)	- 0.000 $-$	- 0.000 $-$
 (66) 442956 '+',0,10,20,40	 5282226 0.259 $-$ (3)	 35023509 0.146 $-$ (3)	 3035905 0.062 $-$ (3)	 2119 0.055 $-NH+O$
 (67) 9060 '+',0,10,20,40 '-',0,10,20,40	 6305 0.296 $-$	 1148 0.260 $-$	 18986 0.254 $+NH-O$ (1)	 92904 0.241 $+NH-O$ (1)
 (68) 798 '+',0,10,20,40	 321710 0.616 $-O$ (1)	 11952 0.339 $-H_2O$ (2)	 68470 0.332 $-C_2H_2O$ (1)	 6736 0.322 $-CH_2$ (1)
 (69) 5280914 '+',0,10,20,40 '-',0,10,20,40	 5280892 0.903 $+O$ (2)	 35025115 0.866 $+O$ (2)	 5283130 0.793 $+O$ (2)	 5283175 0.793 $+O$ (2)

Target compound	Candidate neighbors			
 (70) 174 '+',0,10,20,40	 92135 0.303 $-C_2H_2O$ (2)	 700 0.253 $+NH+O$ (1)	 1030 0.189 $-CH_2$ (1)	 92779 0.087 $-C_4H_6O$ (1)
 (71) 74563 '+',0,10,20,40 '-',0,10,20,40	 95433 0.308 $-H_2$ (1)	 1112 0.296 $+CH_2$ (2)	 853180 0.274 $-H_2$ (1)	 5280523 0.243 $-C_2H_4O_2$ (2)
 (72) 9679 '+',0,10,20,40 '-',0,10,20,40	 78168 0.144 $-O$ (3)	 2944 0.137 $-$ (3)	 1174 0.112 $+CH_2$ (3)	 132749 0.101 $-$ (3)
 (73) 442614 '+',0,10,20,40 '-',0,10,20,40	 688827 0.490 $+O$ (3)	 6250403 0.292 $-$ (3)	 4317971 0.292 $-$ (1)	 11617114 0.267 $+NH+O$ (3)
 (74) 5281166 '+',0,10,20,40 '-',0,10,20,40	 4374065 0.327 $-H_2$ (1)	 7251183 0.250 $-$	 6914579 0.135 $-H_3N$ (3)	 10019140 0.103 $-C_5H_6O_3$ (3)
 (75) 439242 '-',0,10,20,40	 439586 0.267 $-$ (2)	 35020781 0.099 $+H_2O_6P_2$ (3)	 11559259 0.099 $+H_2O_6P_2$	 11631564 0.099 $+H_2O_6P_2$
 (76) 126041 '+',0,10,20,40 '-',0,10,20,40	 3448 0.226 $+H_2$ (3)	 16760027 0.135 $-$ (3)	- 0.000 $-$	- 0.000 $-$

Target compound	Candidate neighbors			
 (77) 2082 '+',0,10,20,40	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –
 (78) 92135 '+',0,10,20,40 '-',0,10,20,40	 17723 0.341 $-CO_2$ (2)	 8998 0.308 $-H_2O$ (2)	 95433 0.308 $-CH_2$ (2)	 1112 0.267 $+H_2$ (1)
 (79) 863 '+',0,10,20,40	 11027759 0.137 –	 26879 0.079 $+NH+O$ (3)	 439406 0.048 $-CH_2$ (3)	- 0.000 –
 (80) 5280523 '+',0,10,20,40 '-',0,10,20,40	 385 0.465 $+O$ (2)	 20984 0.398 $+O$ (2)	 2572 0.325 $-NH+O$ (2)	 12292 0.298 $+O$ (2)
 (81) 83887 '+',0,10,20,40 '-',0,10,20,40	 5960 0.846 –	 6267 0.489 $+NH+O$ (1)	 1088 0.333 $+CO_2$ (2)	 22880 0.181 $-CH_2$ (1)
 (82) 5778 '+',0,10,20,40 '-',0,10,20,40	 847 0.068 $-NH+O$	 158980 0.068 $-NH+O$	 667490 0.067 $+CH_2$ (1)	 99223 0.051 $+C_2H_5N-O$ (3)
 (83) 273 '+',0,10,20,40	 11850 0.183 –	 5962 0.156 $-CO_2$ (1)	 57449 0.156 $-CO_2$ (1)	 4906 0.106 $-C_8H_6O$

Target compound	Candidate neighbors			
 (84) 5324476 '+',0,10,20,40 '-',0,10,20,40	 45073499 0.028 $+O_3S$ (3)	 834 0.023 $+C_5H_8O_3$ (3)	 439258 0.023 $+C_5H_8O_3$ (3)	 2602 0.016 $+C_5H_8O_3$ (3)
 (85) 5284054 '+',0,10,20,40	- 0.000 —	- 0.000 —	- 0.000 —	- 0.000 —
 (86) 439685 '+',0,10,20,40 '-',0,10,20	 95575 0.564 $+O$ (2)	 440014 0.362 $-H_2$	 7405 0.348 —	 33032 0.296 $-H_2O$ (2)
 (87) 443457 '+',0,10,20,40 '-',0,10,20,40	 5281576 0.246 $+CH_2$ (3)	 5282360 0.121 $+CO_2$ (3)	 5281576 0.095 $+H_2N_2-O$ (3)	- 0.000 —
 (88) 8369 '+',0,10,20,40	 54675757 0.410 — (1)	 5789 0.160 $-C_4H_8N_2O_2$ (3)	- 0.000 —	- 0.000 —
 (89) 7293 '+',0,10,20,40	 439888 0.165 — (3)	 225936 0.156 — (3)	 11719 0.134 $+CH_2$ (1)	 35026498 0.106 — (3)
 (90) 17473 '-',0,10,20,40	 53477679 0.370 $+H_2-O$ (1)	 18068 0.333 $+H_2-O$ (1)	- 0.000 —	- 0.000 —

Target compound	Candidate neighbors			
 (91) 94715 '-',0,10,20,40	 3037582 0.729 $-O$ (2)	 439373 0.608 $+O$ (2)	 7027 0.579 $+O$ (1)	 10690 0.450 $-H_2$ (1)
 (92) 49 '-',0,10,20,40	 743 0.410 $-O$ (2)	 853180 0.346 $-H_2$	 95433 0.308 $-H_2$ (1)	 5280523 0.249 $-C_2H_4O_2$ (1)
 (93) 1110 '-',0,10,20,40	 222656 0.655 $-O$ (1)	 92824 0.594 $-O$ (1)	 1112 0.594 $+O$ (1)	 58 0.536 $+O$ (2)
 (94) 641301 '-',0,10,20,40	 997 0.461 $-O$ (2)	 67406 0.443 $-H_2$ (2)	 69475 0.402 $+H_2$ (1)	 1303 0.296 $-H_2O$ (2)
 (95) 1066 '-',0,10,20,40	 10367 0.336 $-$ (1)	 938 0.156 $+CO_2$ (1)	 547 0.112 $-CH_3+N$ (3)	 2553 0.072 $-C_4H_5N_3$ (3)
 (96) 547 '-',0,10,20,40	 12122 0.851 $+O$ (1)	 85782 0.625 $-O$ (1)	 5986986 0.625 $+O$ (3)	 68073 0.567 $-$ (3)
 (97) 440667 '-',0,10,20,40	 54670067 0.589 $-H_2$	 3840 0.456 $+O_2$ (2)	 92283 0.257 $-H_2$	 71 0.185 $+OH_2$ (2)

Target compound	Candidate neighbors			
 (98) 5994 '+',0,10,20,40	 6166 0.645 $-O$ (1)	 8955 0.477 $-H_2$	 6439929 0.377 $+H_2$ (1)	 71433980 0.359 $-O_2$ (3)
 (99) 439280 '+',0,10,20,40 '-',0	 1148 0.387 $+O$ (1)	 9060 0.387 $+O$ (1)	 6305 0.304 $+O$ (1)	 439202 0.074 $+CO_2$ (3)
 (100) 5280710 '+',0,10,20,40 '-',0,10,20,40	 5281912 0.901 $+O$ (1)	 5280883 0.864 $-O$ (2)	 35027042 0.828 $+O$ (2)	 5283050 0.792 $+O$ (1)

Legend:

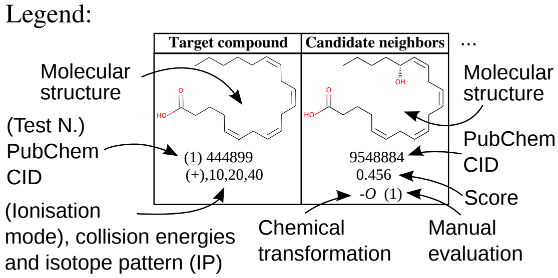
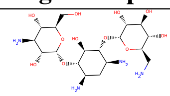
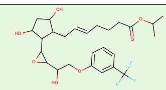
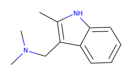
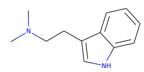
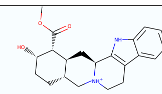
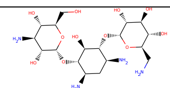
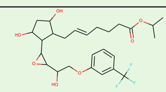
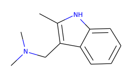
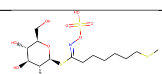
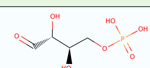
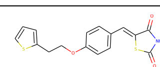
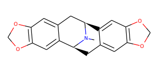
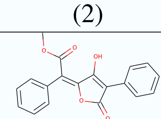
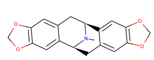
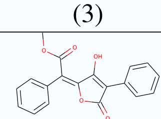
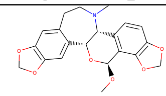
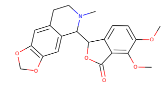
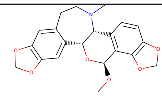
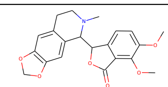
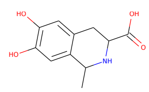
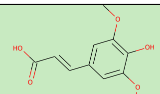
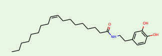
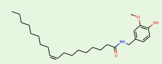
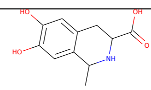
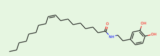
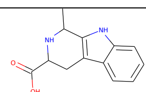
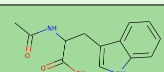
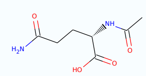
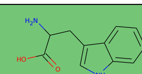
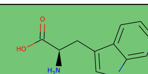
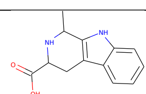
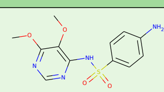
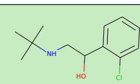
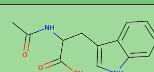
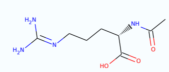


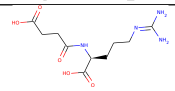
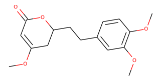
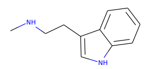
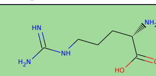
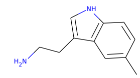
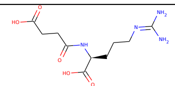
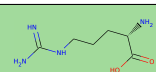
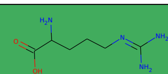
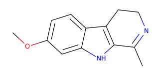
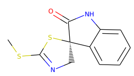
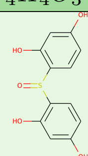
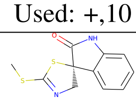
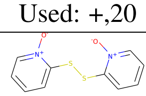
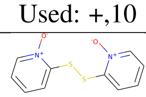
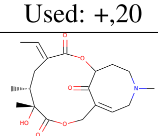
Table S10: CASMI challenges 2012-2016 approximate CE

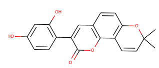
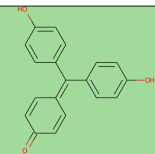
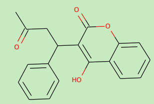
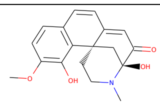
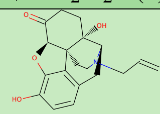
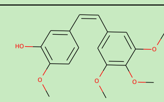
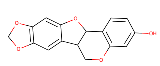
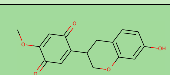
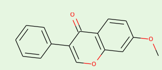
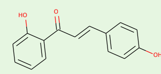
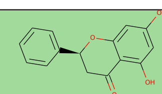
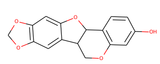
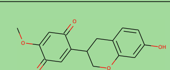
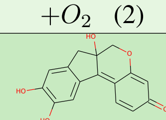
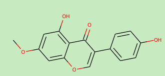
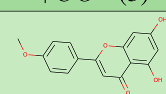
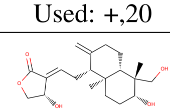
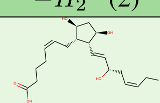
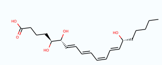
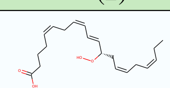
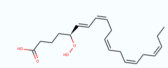
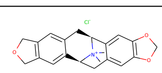
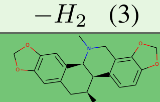
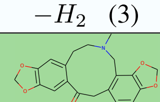
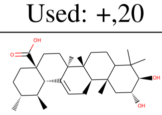
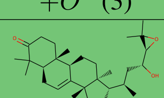
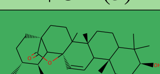
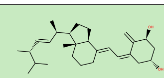
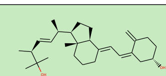
Structural similarity with target:

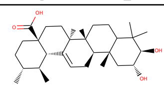
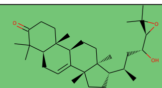
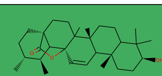
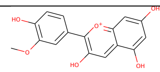
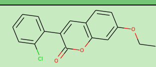
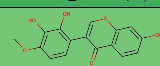
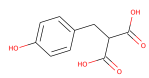
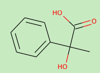
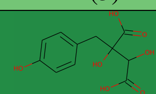
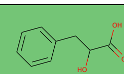
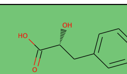
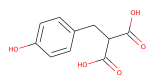
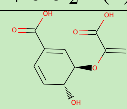
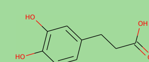
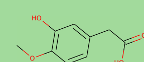
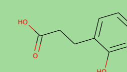
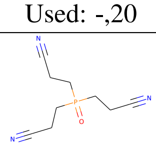
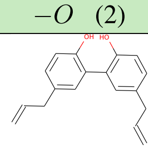
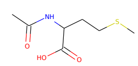
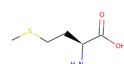
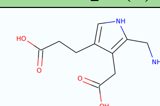
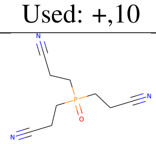
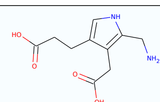
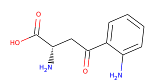
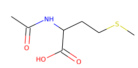
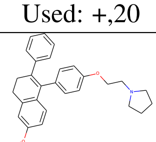
0.0 - 0.1	0.1 - 0.2	0.2 - 0.3	0.3 - 0.4	0.4 - 0.5	0.5 - 0.6	0.6 - 0.7	0.7 - 0.8	0.8 - 0.9	0.9 - 1.0

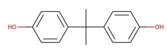
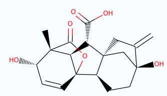
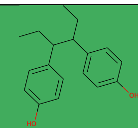
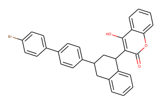
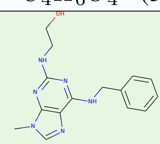
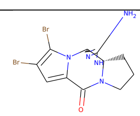
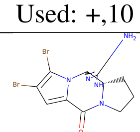
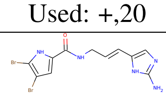
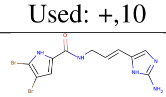
Target compound	Candidate neighbors			
 (1) 6032 Real: +,10,20,30 Used: +,10,20,40	 53394147 0.072 $-O_2$ (3)	 181513 0.045 $+C_{12}H_{16}N_4O_5$	 6089 0.030 $+C_{12}H_{16}N_4O_5$	 11869220 0.022 $+C_5H_7NO_3$ (3)
 (2) 6032 Real: +,10,20,30 Used: +,10,20	 53394147 0.096 $-O_2$ (3)	 181513 0.026 $+C_{12}H_{16}N_4O_5$	- 0.000 –	- 0.000 –
 (3) 46173875 Real: -,10,20,30,40 Used: -,10,20,40	 122357 0.066 $+C_9H_{15}NO_5S$ (2)	 53394148 0.043 $+C_7H_6N_2$	- 0.000 –	- 0.000 –
 (4) 12304178 Real: +,10,20,30 Used: +,10,20,40	 54684634 0.245 $+NH_3-O$ (3)	- 0.000 –	- 0.000 –	- 0.000 –
 (5) 12304178 Real: +,10,20,30 Used: +,10,20	 54684634 0.258 $+NH_3-O$ (3)	- 0.000 –	- 0.000 –	- 0.000 –

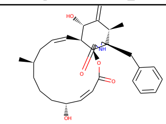
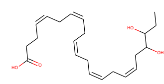
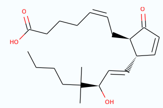
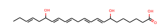
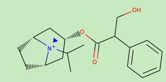
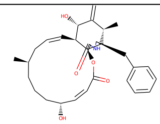
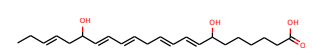
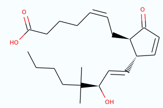
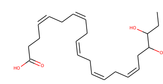
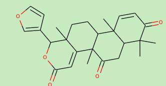
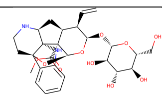
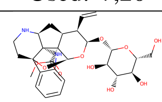
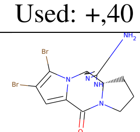
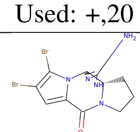
Target compound	Candidate neighbors			
 (6) 197775 Real: +,10,20,30 Used: +,10,20,40	 1309 0.270 –	- 0.000 –	- 0.000 –	- 0.000 –
 (7) 197775 Real: +,10,20,30 Used: +,10,20	 1309 0.436 –	- 0.000 –	- 0.000 –	- 0.000 –
 (8) 162692 Real: +,15 Used: +,10	 637775 0.236 $+NH-O$ (3)	 5282106 0.114 $+O-C_{15}H_{30}$ (3)	 5311093 0.085 $+O-C_{15}H_{30}$ (3)	- 0.000 –
 (9) 162692 Real: +,15 Used: +,20	 5282106 0.053 $+O-C_{15}H_{30}$ (3)	- 0.000 –	- 0.000 –	- 0.000 –
 (10) 73530 Real: +,15 Used: +,10	 2002 0.261 $-O$ (2)	 182230 0.128 $+C_6H_2-O_2$ (3)	 1148 0.108 $+C_2H_2$ (2)	 9060 0.108 $+C_2H_2$ (2)
 (11) 73530 Real: +,15 Used: +,20	 17134 0.168 – (3)	 5606 0.078 – (3)	 2002 0.078 $-O$ (2)	 67427 0.052 $+O-H_2$ (3)

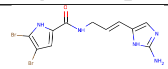
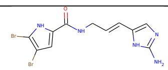
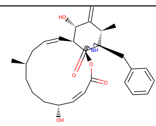
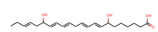
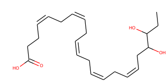
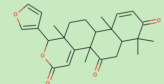
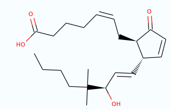
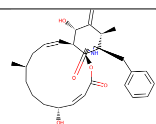
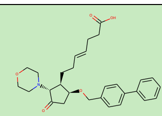
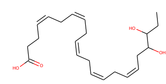
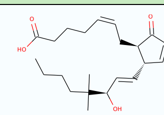
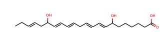
Target compound	Candidate neighbors			
 (12) 439968 Real: +,15 Used: +,10	 592220 0.067 $-H_2O$	 6088 0.066 $+C_4H_4O_3$	 6322 0.059 $+C_4H_4O_3$ (1)	 15760 0.035 $+C_4H_4O_3$
 (13) 439968 Real: +,15 Used: +,20	 6322 0.077 $+C_4H_4O_3$ (1)	 232 0.077 $+C_4H_4O_3$ (1)	 5280951 0.037 $+C_2H_4O_2$	- 0.000 –
 (14) 188830 Real: +,15 Used: +,10	 5044578 0.740 $-O$ (3)	- 0.000 –	- 0.000 –	- 0.000 –
 (15) 188830 Real: +,15 Used: +,20	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –
 (16) 3109 Real: +,15 Used: +,10	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –
 (17) 3109 Real: +,15 Used: +,20	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –
 (18) 6433332 Real: +,25 Used: +,20	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –

Target compound	Candidate neighbors			
 (19) 15380912 Real: -,25 Used: -,20	 5100 0.219 $+CH_2O_2$ (3)	 54678486 0.177 $+CO$ (3)	- 0.000 –	- 0.000 –
 (20) 51041521 Real: +,25 Used: +,20	 5284596 0.244 $-CH_2$ (2)	 5351344 0.180 – (3)	- 0.000 –	- 0.000 –
 (21) 363863 Real: +,15 Used: +,10	 100072 0.653 $-H_2$ (2)	 638006 0.483 $+O_2$ (2)	 5357218 0.324 $+CO_2$ (3)	 68071 0.251 $+CO$ (3)
 (22) 363863 Real: +,15 Used: +,20	 100072 0.627 $-H_2$ (2)	 6453902 0.533 – (2)	 5281804 0.479 – (2)	 5280442 0.479 – (3)
 (23) 5318517 Real: +,6 Used: +,10	 5283217 0.602 $-H_2$ (3)	 9841438 0.577 $-H_2$ (3)	 5283195 0.479 $+O$ (3)	 5283184 0.378 $+O$ (3)
 (24) 45266443 Real: +,25 Used: +,20	 197810 0.846 $+O$ (3)	 4970 0.846 $+O$ (3)	- 0.000 –	- 0.000 –
 (25) 21363056 Real: +,15 Used: +,10	 44559946 0.360 $+O$ (3)	 6708656 0.168 $+H_2O$ (2)	 24779618 0.103 $+C_2H_4O_2$ (3)	 5710148 0.103 $+C_2H_4O_2$ (3)

Target compound	Candidate neighbors			
 (26) 21363056 Real: +,15 Used: +,20	 44559946 0.462 $+O$ (3)	 6708656 0.181 $+H_2O$ (2)	- 0.000 –	- 0.000 –
 (27) 441773 Real: +,25 Used: +,20	 689007 0.259 – (3)	 5318834 0.231 – (3)	- 0.000 –	- 0.000 –
 (28) 17840182 Real: -,15 Used: -,10	 1303 0.238 $+CO_2$ (2)	 120693 0.204 $-CH_2O_2$ (2)	 3848 0.202 $+CO_2$ (1)	 643327 0.191 $+CO_2$ (2)
 (29) 17840182 Real: -,15 Used: -,20	 12039 0.305 $-O$ (2)	 348154 0.216 $+CO$ (2)	 160562 0.178 $+CO$ (2)	 873 0.176 $+CO_2$ (2)
 (30) 74039 Real: +,15 Used: +,10	 72300 0.134 –	 6180 0.083 $+H_2O$	 6137 0.054 $+C_2H_4O_2$	 1021 0.036 $-NH_3$ (3)
 (31) 74039 Real: +,15 Used: +,20	 1021 0.097 $-NH_3$ (3)	 161166 0.096 $+O-HN$	 6180 0.089 $+H_2O$	- 0.000 –
 (32) 4416 Real: +,25 Used: +,20	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –

Target compound	Candidate neighbors			
 (33) 6623 Real: -,25 Used: -,20	 9819600 0.146 $-C_4H_6O_4$ (3)	 3606 0.097 $-C_3H_6$ (2)	- 0.000 –	- 0.000 –
 (34) 54680676 Real: +,25 Used: +,20	 4592 0.081 $+C_3H_2N_2O_6P_2$ (3)	- 0.000 –	- 0.000 –	- 0.000 –
 (35) 42636938 Real: +15 Used: +,10	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –
 (36) 42636938 Real: +15 Used: +,20	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –
 (37) 6312649 Real: +15 Used: +,10	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –
 (38) 6312649 Real: +15 Used: +,20	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –

Target compound	Candidate neighbors			
 (39) 5311281 Real: +15 Used: +,10	 16061148 0.095 $+C_7H_3NO$	 5283083 0.078 $+C_7H_3NO$ (3)	 53394047 0.072 $+C_7H_3NO$	 657309 0.054 $+C_9H_8O_2$ (3)
 (40) 5311281 Real: +15 Used: +,20	 53394047 0.140 $+C_7H_3NO$	 5283083 0.127 $+C_7H_3NO$ (3)	 16061148 0.099 $+C_7H_3NO$	 3550024 0.082 $+C_3H_7N$ (3)
 (41) 118984484 Real: +30 Used: +,20	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –
 (42) 118984484 Real: +30 Used: +,40	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –
 (43) 42636938 Real: +30 Used: +,20	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –
 (44) 42636938 Real: +30 Used: +,40	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –

Target compound	Candidate neighbors			
 (45) 6312649 Real: +30 Used: +,20	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –
 (46) 6312649 Real: +30 Used: +,40	- 0.000 –	- 0.000 –	- 0.000 –	- 0.000 –
 (47) 5311281 Real: +30 Used: +,20	 53394047 0.149 $+C_7H_3NO$	 16061148 0.111 $+C_7H_3NO$	 3550024 0.062 $+C_3H_7N$ (3)	 5283083 0.059 $+C_7H_3NO$ (3)
 (48) 5311281 Real: +30 Used: +,40	 12786332 0.417 $+H_2$ (3)	 16061148 0.111 $+C_7H_3NO$	 5283083 0.111 $+C_7H_3NO$ (3)	 53394047 0.100 $+C_7H_3NO$

Legend: