Chemical-Damage MINE: a database of curated and predicted spontaneous metabolic reactions

James G. Jeffryes[1], Claudia Lerma-Ortiz[2], Filipe Liu[1,2], Alexey Golubev[3], Thomas D. Niehaus[4], Mona Elbadawi-Sidhu[5], Oliver Fiehn[5], Andrew D. Hanson[4], Keith EJ Tyo[6], Christopher S. Henry[1,2]*

1. Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, U.S.A., 2. Consortium for Advanced Science and Engineering, University of Chicago, Chicago, IL 60637, U.S.A., 3. Department of Biochemistry, Saint Petersburg State University, Saint Petersburg, Russia 4. Horticultural Sciences Department, University of Florida, Gainesville, FL 32611, U.S.A., 5. West Coast Metabolomics Center, University of California, Davis, California, U.S.A., 6. Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, U.S.A.

*Correspondence may be addressed to this author

Christopher S. Henry
Argonne National Laboratory
Full postal address missing

Tel: +1-630-252-6501
Email: chenry@ci.uchicago.edu

Abstract

Spontaneous reactions between metabolites are often neglected in favor of emphasizing enzymatically catalyzed chemistry because it is assumed that spontaneous reaction rates are insignificant under physiological conditions. However, synthetic biology and metabolic engineering efforts can greatly raise the levels of natural metabolites or introduce unnatural ones, so that previously innocuous or nonexistent spontaneous reactions can become problematic. This is because in SynBio systems spontaneous reactions rates can exceed the capacity of the platform organism to dispose of toxic or chemically active reaction products. While a synthetic biologist can consult various sources of high quality to discover competing or toxic enzymatic pathways' side reactions, no corresponding extensive compilation of spontaneous side reactions exists. We address this deficiency by constructing a large database of spontaneous metabolite reactions that occur in physiological conditions. We assembled this database by extensive curation of literature data. We then leveraged this resource to construct 148 reaction rules describing the known spontaneous chemistry in a substrate-generic way. We applied these rules to all compounds in the ModelSEED database, predicting 180891 putative spontaneous reactions. The resulting dataset is accessible via a web application at minedatabase.mcs.anl.gov and through developer tools. Here we demonstrate how damage-prone intermediates and end products are amply distributed in diverse metabolic pathways. We also show how predictions of spontaneous chemical damage can help rationalize toxicity and carbon loss with examples from published pathways to commercially relevant products. We explain how the analysis of damage prone areas in metabolism can help in the design of effective strategies in SynBio and engineering. Finally, we disclose the detection and confirmation of N-carbamoyl proline, one of the damage products predicted by the Chemical Damage (CD)-MINE by using our rule-based reaction prediction toolset and that had not been previously detected in biological samples.

Keywords

Computational biochemistry, metabolite damage, metabolic engineering, side-product, spontaneous reaction, synthetic biology.

1. Introduction

In synthetic biology, metabolic pathways are often viewed as pipework systems that can be replumbed by adding or removing enzymatic catalysts (Colijn et al., 2009; Dueber et al., 2009; Boyle and Silver, 2012). While intuitive for those with a background in engineering, an unfortunate effect of this analogy is that it ignores the intrinsic chemical nature of both intermediates and enzymes. In fact, metabolites are not passive objects of manipulation by enzymes, and many of them can react spontaneously (Linster et al., 2013; Hanson et al., 2016; Lerma-Ortiz et al., 2016; Golubev et al., 2017).

There is a paucity of information regarding these reactions, which is understandable. Stable metabolites which accumulate in the cell, can be detected and identified (DeBerardinis and Thompson, 2012; Dromms and Styczynski, 2012; Sévin et al., 2015). However, derivatives may also be removed or restored to their original state by damage-control mechanisms, masking their presence (Linster et al 2013; Hanson et al 2016). In other instances, metabolites become reactive when exposed to light or oxygen (Frelin et al., 2015; Lerma-Ortiz et al., 2016). High reactivity has two consequences: a) degradation in *ex vivo* during extraction, derivatization, and ionization (Halket et al., 2005; Koek et al., 2011; Park et al., 2016); b) participation in chain reactions (Phaniendra et al., 2015; Lerma-Ortiz et al., 2016), in which case the formation of degraded metabolites is inferred for the damage they cause, rather than for their accumulation.

Moreover, metabolic modeling is based solely in comparative genomics, so it is unable to consider spontaneous reactions, because it uses sequence homology of the enzymes to identify reaction catalysts (Desai et al., 2011). Due to these facts, the discovery of both damaged intermediates and repair mechanisms has been a largely anecdotal enterprise (Lerma-Ortiz et al., 2016), in contrast to the rich and systematized array of enzymatic pathway databases (Kanehisa et al., 2014; Seaver et al., 2014; Caspi et al., 2016).

Synthetic biology's methodology upsets normal cellular homeostasis, as it introduces heterologous pathways and intermediates, it modulates enzyme concentrations and removes natural cell barriers (Li J et al., 2018). By disposing of pre-emption and repair mechanisms in its systems design, it can leave unhindered and even increase spontaneous and/or enzymatic promiscuous reactions. The resulting toxic and inhibitory effects often uncover these damaged products as well as dynamic pre-emption and repair mechanisms which would otherwise remain hidden in the cell (Withers et al., 2007; Kildegaard et al., 2014; Poust et al., 2015). Once such effects are exposed, they can be ameliorated using various strategies. These include spatially separating reactions into micro-compartments (Kerfeld and Erbilgin, 2015; Koppel and Balskus, 2016), exploring scaffolding enzymes (Dueber et al., 2009; Danchin, and Sekowska, 2014) or controlling enzyme expression (Andre et al., 2013; Kildegaard et al., 2014; Jones et al., 2015).

Although it would be reasonable to generalize that most of the spontaneous reactions are deleterious, whereas all the enzymatic reactions are tuned to optimal metabolic function, this is not the case. Some non-enzymatic reactions have high specificity and are part of important pathways (Keller et al., 2015), and quite a few enzymes are promiscuous and catalyze side reactions (see supplementary data). Moreover, enzymatic reactions share reaction operators with their parallel spontaneous reactions (Jeffryes et al., 2019), so an indirect way to identify a promiscuous enzymatic reaction, would be to identify a spontaneous reaction and infer the type of enzyme that could catalyze it. Deleterious reactions, being spontaneous and/or catalyzed by promiscuous enzymes, may affect metabolism's

efficiency (due to loss of building materials and energy sources) at best, and produce toxic compounds at worst (Linster et al., 2013; Bommer et al., 2019; Jeffryes et al., 2019).

Considering all the information above, it is evident that both spontaneous and promiscuous enzymes reactions, and cellular damage control play a very important role in metabolic modeling and synthetic biology. However, there remains a pressing need for a resource that allows researchers to identify and predict deleterious reactions that could affect the performance of their metabolic designs. We addressed this issue by doing an extensive curation of published work and building a database of spontaneous reactions that occur under physiological conditions. We used this data set to infer reactive motifs ('rules') from these reactions and then used the rules to predict novel spontaneous reactions in metabolites from ModelSEED (Henry et al., 2010) metabolic models. Here we demonstrate how this resulting database, called CD-MINE, enables metabolic engineers and synthetic biologists to detect novel compounds and identify potential sources of carbon loss and toxicity in metabolic pathways. These capabilities make the CD-MINE a valuable source of testable hypothesis for troubleshooting altered strains and cell-free SynBio systems.

2. Material and methods

2.1. Compounds library

A set of all compounds in the ModelSEED database was obtained from https://github.com/ModelSEED/ModelSEEDDatabase (commit 2d8a6af) and tautomerized using ChemAxon's Standardizer program (v.16.1.11.0 www.chemaxon.com).

2.2. CD-MINE Construction; starter reaction rules system

We show a flowchart of our database's construction in Fig. 1. Previous work identified 30 basic metabolite families prone to spontaneous reactions under physiological conditions (Lerma-Ortiz et al., 2016). We used this information and extended our literature search to cover a wider variety of metabolite chemical structures. Other reactions of this type, which are part of metabolic pathways and have been annotated as such in MetaCyc 20.5 (released Dec 17th 2016), were obtained and associated with compound structures according to their MetaCyc Frame ID. Filtering out transport reactions and those missing fully defined chemical structures resulted in a second collection, which we combined with that originally curated from the literature (Table S1).

Reaction rules were written to capture the chemical transformation motifs (Fig. 1 and Table S2) identified as described below. We defined each reaction rule guided by chemical reasoning and empirical testing to balance accuracy and exploratory power. For bimolecular reactions, one reactive partner was defined by the rule while the other was allowed to vary. For example, there are three Michael Addition rules, one for each cysteine, glutathione, and fumarate, as reactive partners. This excludes some possible reactions but reduces the combinatorial explosion and the combination of two rare compounds.

2.3. CD-MINE identifiers for compounds and reaction rules

All spontaneous reaction chemistry rules are represented as using SMART's strings (https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html) that are compatible with the

open/source RDKit cheminformatics software. RDKit was applied with these rules and the compounds in the ModelSEED database (ModelSEED.org) (Seaver et al., 2019) to generate the compounds and reactions in the CD-MINE. Compound entries contain the chemical formula, exact mass, InChIKey, canonical SMILES (Simplified Molecular Input Line Entry System) (O'Boyle, 2012), the fingerprints called MACCS keys (http://www.dalkescientific.com/writings/diary/archive/2014/10/17/maccs_key_44.html) and RDKit Fingerprints (http://rdkit.blogspot.com/2018/10/using-new-fingerprint-bit-rendering-code.html), and are uniquely identified by a SHA1 hash of the canonical SMILES. A compound may also be annotated with additional information such as common names or database links propagated by InChIKey matching from the ModelSEED, KEGG (Kanehisa et al., 2014) and PubChem (https://pubchem.ncbi.nlm.nih.gov/) databases. Reactions are uniquely identified by the SHA256 hash of the sorted chemical reaction as SMILES. Reaction entries contain stoichiometric and compound identifiers for reactants and products as well as a list of the reaction rules that predicted the reaction.

2.4. Searching for products of spontaneous reactions of pathway intermediates.

2.5 Mapping damage-prone metabolites in biosynthetic pathways networks.

2.6. Extent of chemical reactivity in synthetic biology systems

The reactions in pathways described by Zhang et al (2016) were obtained from the paper's supplemental information. To eliminate the confounding factor of reactive cofactors, only the compounds in the reaction described as "carbon transfer" compounds by the authors were considered. For the same reason, the reactivity of acetyl-CoA, S-adenosyl-L-methionine and S-adenosyl-L-homocysteine which appear in carbon transfer reaction as carbon donors were not considered in the determining reactive steps in the pathway.

2.7. Synthesis and purification of N-carbamoyl proline

N-Carbamoyl proline was synthesized by combining 0.2 mmol L-proline and 0.2 mmol sodium cyanate in 1 mL of water and incubating at 22°C for one hour. N-carbamoyl proline was purified by HPLC on a Polaris C18-A column (150 x 4.6 mm, 3 µm, Agilent) with 0.1% formic acid as mobile phase (flow rate 1 mL/min). Fractions containing N-carbamoyl proline (detected by absorbance at 200 nm) were collected and successive runs were pooled and lyophilized.

2.8. LC-MS analysis of N-carbamoyl proline

Escherichia coli strain MG1655 cells and medium were extracted and analyzed by LC-MS as previously described (Jeffryes et al., 2015). Briefly, replicate cell pellets were extracted with 1 mL cold acetonitrile: isopropanol: water (3:3:2) and analyzed by hydrophilic interaction liquid chromatography (HILIC) on an Agilent 1290 Infinity LC system coupled to quadrupole time-of-flight mass spectrometer (Agilent Technologies, Santa Clara, CA). Full scan and tandem mass spectrometry data were collected in positive and negative polarity with a 4 spectra/s scan rate; mass calibration was maintained by constant infusion of reference ions. Data processing was performed using the open-source software MZmine 2 (http://mzmine.github.io/) and MS-FLO post-processor.

Compound identification was based on accurate-mass and MS/MS matching in the NIST MSPepSearch GUI (http://peptide.nist.gov/) with precursor ion and fragment peak tolerance of 0.05 m/z units using the NIST, Metlin, MassBank and LipidBlast libraries; additional MS/MS and retention time

matching to in-house libraries was also performed. Unknown features were investigated using the EcoCyc MINE database and N-carbamoyl proline was a returned hit for the unknown feature with m/z 159.0764. The experimental MS/MS spectrum was interrogated against the predicted MS/MS spectrum generated by MassFrontier and found to be a potential positive hit. Authentic standards of N-carbamoyl proline were analyzed by LC-MS as above and shown to have the same retention time and MS/MS spectrum as the unknown feature with m/z 159.0764. Further, when *E. coli* extract and N-carbamoyl proline were co-injected, a single peak eluted with an identical MS/MS spectrum as authentic standard.

3. Results and Discussion

3.1. The CD-MINE

In order to create a suitable library of chemical compounds containing regular metabolites as well as damaged products, we structured the CD-MINE starting compounds set on the complete ModelSEED's metabolites data (Henry CS et al., 2010) as described in methods. Only 24087 compounds in the database had a defined chemical structure, to which reaction operators could be applied, and the other 3134 compounds contained a general or undefined structure and were identified by names like "protein" or "residue".

Using the data reported in the literature and MetaCyc for over 700 spontaneous reactions (Table S1) (http://minedatabase.mcs.anl.gov/cdmine/#/datatables/litrxns), we created 148 reaction rules (Table S2) (http://minedatabase.mcs.anl.gov/cdmine/#/datatables/rxnrules), which once applied to the chemically defined compounds pool predicted 180891 reactions. Merging together the products of these reactions and the original metabolites in the database, the ModelSEED now has 189807 putative metabolites of defined chemical structure (Fig, 1).This increment on metabolite content is much smaller than the 18 to 363-fold increment observed when constructing the Enzymatic MINE (Jeffryes et al.,2015). Most CD-MINE compounds are less than 2000 Daltons in mass although the database does contain some very large starch and lipid structures from ModelSEED. The increase in average mass in the CD-MINE is not just a result of adduct-forming reactions but also due to the generation of high-molecular weight molecules with multiple reactive sites. The shift towards more hydrophilic compounds in the CD-MINE is consistent with the effect of oxidation and hydration damage reactions.

3.2 Distribution of reactive intermediates in metabolic pathways and identification of their spontaneous reaction products

We first determined which metabolites had chemical identifiers in the ModelSEED database, so we could assess their spontaneous reactions as determined by our rules (Table S2). We describe this process in methods and we show the corresponding reactive intermediates and the number of possible products derived from their spontaneous reactions in Table S3. We also mapped these reactive intermediates into the metabolic pathways as designed in the KEGG network (Table S4 and Figs. S1 A and S1 B). The intermediates of the metabolic pathways of porphyrin and chlorophyll, and arginine and proline showed the highest number of possible derivatives. The compounds participating in purine metabolism, fatty acid elongation and degradation, biosynthesis of unsaturated fatty acids, peptidoglycan and folate biosynthesis, and amino and nucleotide sugar metabolism also showed a high

incidence of spontaneous reactivity. These results are consistent with our previous observation that pathway-end products like porphirins, B vitamins, nucleotides, sugars and some amino acids are damage-prone (Lerma-Ortiz et al., 2016), but also show that a high proportion of their precursors are subject to spontaneous reactions. Our results stress the high reactivity of fatty acids, which is due to their propensity to undergo autoxidation (Morrow et al., 1990; Bernoud-Hubac et al., 2001). It is interesting to note that for fatty acids synthesis, organisms have evolved a mechanism using acyl-carrier proteins (ACP) to transport intermediates from one enzyme to the next one (Chan and Vogel, 2010), whereas in the degradation process organisms use coenzyme A for this transportation (Houten et al., 2016). That said, the stability of the thioester bond with the ACP depends on the length of the fatty acid chain, making fatty acids longer than 15 carbons more labile to hydrolysis (Zornetzer, et al., 2010), and consequently, oxidation. Also, eukaryotes have a fused fatty acid synthetase complex that increases production efficiency and decreases exposure of nascent fatty acids to oxidation.

There is a high incidence of reactive metabolites in the degradation pathways of aromatic and branched-chain amino acids and lysine. In contrast, this number is low for the intermediates in the biosynthetic pathways of the corresponding amino acids (Figs. S1 A and Figs. S2-S6). We need to consider the ubiquitous presence of acyl-coenzyme A esters in these degradation metabolic pathways. This is because the coenzyme A moiety itself can give rise to damaged products (Lerma-Ortiz et al., 2016), which can inflate the number of derivatives and eclipse the damage input of other reactive compounds. For example, in phenylalanine metabolism, the CD-MINE predicts 33 derivatives from 3-Oxo-5,6-dehydrosuberyl-CoA semialdehyde (Table S3). Hence, the overall chemical structure of the individual intermediates should be considered, when analyzing the distribution of reactive intermediates in the metabolic network. Also, it is important to identify and map the reactive metabolites (hot spots) in the pathways themselves. It is possible to find whole areas of high spontaneous reactivity (Figs. S2-S6) and explore the nature of chemical and structural phenomena which cells have developed around them. In other words, the identification and analysis of the coping mechanisms of pre-emption and repair which some organisms have developed (Linster CL, et al., 2013), as well as the discovery of structural strategies-like fusions (Henry et al 2016), formation of enzymatic complexes and compartmentations. If metabolic engineers are to implement novel designs based on biosynthetic pathway networks, they need to be aware of these spontaneous reactions' hot spots. Here we give some examples to illustrate our point.

3.3. Commonality of reactive compounds in pathways of commercial interest

To explore the scope of metabolic engineering efforts that may be impacted by spontaneous reactions, we examined 262 chemicals with commercial value that could potentially be synthesized by engineered *Escherichia coli* grown on minimal glucose medium (Zhang X et al., 2016). The heterologous pathways introduced to synthesize these products are drawn from the KEGG database and therefore pathway intermediates and products were matched to corresponding CD-MINE compounds by their KEGG ID numbers. Almost half (49.6%) endpoint products are predicted to undergo spontaneous reactions. Fig. 2 shows that a further 92 pathways (35.1%) contain reactive intermediates which may be a source of carbon loss or cellular stress. In particular, pathways originating outside of central carbon metabolism seem to be more prone to reactive intermediates although this effect is partially accounted for by greater pathway length (median pathway length of 5 vs. 3 for central carbon origination) with correspondingly more opportunities for reactive intermediates (See Table S3 for more details). These results should not be interpreted to mean that most engineered pathways will be stymied by chemical

reactivity but rather that the impact of spontaneous reactions is likely to be widespread and is currently underappreciated. The CD-MINE provides a unique collection of these potential activities for systematic evaluation.

Anthranilate synthesis

Anthranilate (a petroleum derivative) is an important chemical product, but the current manufacturing process is costly. Being an intermediate in aromatic amino acids biosynthesis, it could alternatively be produced in a sustainable manner by a synthetic biology method. This has been tried using an *Escherichia coli* W3110 trpD9923 mutant which has a fused trpD gene where the first domain is intact with a very small fraction of the second domain attached to it (Balderas-Hernandez VE et al., 2009). Metabolic engineers have improved this system by inactivating the PTS operon, transforming the mutant to increase glucokinase and galactose permease activities, and mutating its 3-deoxy-7-phosphoheptulonate synthase to be inhibition resistant (no feedback control). When a fed-batch process to grow the engineered mutant was implemented, a 19-fold surge of anthranilate production was observed, but the final biomass output decreased (Balderas-Hernandez VE et al., 2009).

As we show in Fig. 3, and Fig. S2 anthranilate is a hot spot of spontaneous reactivity. In the case of the aromatic amino acids biosynthetic pathway, organisms dispatch anthranilate, N-(5-Phospho-D-ribosyl) anthranilate, and prephenate efficiently. They have evolved several fused enzymes, which channel these compounds from one reactive center to the next one, so preventing long exposure to oxygen and their interaction with other metabolites (Henry et al., 2016) (Fig. 3). In the *E. coli* W3110 trpD9923 mutant, the removal of the phosphoribosyltransferase domain causes anthranilate accumulation, but it necessarily increases its exposure to spontaneous chemical reactions. CD-MINE predicts 13 derivatives from spontaneous reactions of anthranilate, among them N-acetylanthranilic acid, 2-[(3-carboxypropanoyl)amino]benzoic acid and 2-(ethylideneamino)benzoic acid. Moreover, it is thought that amino aromatics like anthranilate form polymers in the presence of oxygen (Braun, 1984). In general, aromatic compounds like benzoate, form radicals in the presence of hydroxyl radicals (Wu C, 2017) and experimental evidence has shown that methyl anthranilate (a similar molecule) is degraded by UV and H_2O_2, and can be decarboxylated (Lanzafame GM, 2017). An alternative to overcome possible side reactions due to oxidation of the accumulated anthranilate, would be to engineer an *E. coli* W3110 trpD9923 mutant with catalase overexpression, or bioengineer anthranilate production in an anaerobic organism (Braun K, 1984). Also, the addition of N-acetylanthranilate amidase cloned from *Arthrobacter Sp* (Overhage et al., 2005) could be tried, to recover anthranilate from its acetylated derivative.

Alkaloids synthesis

The synthesis of plant alkaloids is another example of a pathway of commercial significance where spontaneous reactions may also play an underappreciated role. Some of these alkaloids have medicinal properties like berberine (Koshla et al., 1992), corytuberine (Zetler, 1988), and magnoflorine (Li C and Wang, 2014). Since the plant production of alkaloids has not a very high yield and their extraction is a complicated process, researchers are developing new synthesis methodologies.

A synthetic biology system to make reticuline (a precursor of these alkaloids), has been engineered in *Escherichia coli* (Minami et al., 2008; Nakagawa et al., 2011). Some melanin-like pigments were

observed in the production system, which indicated the presence of oxidation products. The incorporation of a tyrosinase from *Ralstonia solanacearum*, which displays low o-diphenolase activity, increased reticuline production 7-fold. Fig. 4 shows L-dopa (Fig. 4 (1, 5)) and dopamine (Fig. 4 (2, 5)), which are some of the first intermediates in the reticuline synthesis pathway, and the spontaneous reactions (oxidation, deamination, cyclization or addition) involving these unstable compounds, and their immediate derivatives. Although some of these spontaneous reactions have been incorporated into the tyrosine metabolism (Fig. S4) and melanin biosynthetic pathway, some of these compounds are also good substrates for deleterious side reactions of promiscuous enzymes. One example is the deamination of dopamine (Fig. 4 (2, 5)) to produce 3,4-dihydroxyphenylacetaldehyde (dopal) (Fig. 4 (3)), which is a toxic compound, as it reacts with proteins and some amino acids (Anderson et al., 2016). Deamination of dopamine may be catalyzed by monoamine oxidase in mammals (Arai et al., 1996; Marchitti et al., 2007), an enzyme of low substrate specificity. It is also a side reaction of pig kidney dopamine decarboxylase (Bertoldi and Voltattorni, 2009). These enzymatic side reactions might also be present in plants and bacteria. Moreover, under certain conditions, spontaneous deamination can occur (Vianello et al., 2016), obviously in all organisms. Another example is the promiscuous reaction of tyrosinase, which as explained above, has o-diphenolase activity (Ortiz-Ruiz et al., 2015), oxidizing L-dopa and dopamine. This reaction can also happen spontaneously, in cells in high oxidation state. The oxidation products (Fig. 4 (9)) undergo either cyclization, turning out into the corresponding dopaminechrome and dopachrome (Fig 4 (8)), or they can react in a Michael addition with cysteine (Agrup et al., 1982), glutathione (Fig. 4 (7)), alanine (Fig.4 (6)) and other amino acids. Considering the above, in any synthetic system where L-dopa and dopamine are involved, improving production will require the optimization of the enzymatic components. For example, some bacterial azoreductases of low substrate specificity can reduce many quinones (Ryan et al., 2014), so engineering this type of enzyme into the system might increase reticuline yield. However, optimizing the enzymatic components will be just part of the solution because researchers also need an understanding of the spontaneous reactions that may occur as intermediates concentrations increase.

From these observations we can conclude that when designing a new system, it is important for engineers to know which intermediates in their pathway system are reactive, the possibility of damage-control and pre-emption that could be implemented in their designs (Sun et al., 2017) as well as have an idea of the architecture of their biological material, from protein structure to cell compartmentation. These spontaneous products may even be able to serve as markers for undesirable compounds accumulation, allowing metabolic engineers to prioritize bottleneck enzymes without quantifying the concentration of all pathway intermediates.

3.4. Annotation of new metabolites

Experimental confirmation of N-carbamoyl proline

To show the utility of the CD-MINE for the annotation of novel metabolites, we detected and confirmed the presence of N-carbamoyl proline in a biological sample. This modified amino acid has been chemically synthesized (Delgado et al., 2012) and is commercially available. However, other than one paper describing its presence in the N-terminus of a crystallized protein (Olajuyigbe et al., 2010), there was little evidence for it being a naturally occurring metabolite; it is not present in metabolic databases such as MetaCyc (Caspi et al., 2016) and KEGG, nor is it found in metabolomics databases

such as METLIN (Smith et al., 2005) or HMDB (Wishart et al., 2013). CD-MINE however, predicts that this compound may arise in the presence of either carbamoyl phosphate or urea.

Using CD-MINE's accurate mass/charge ratio search tools, we found carbamoyl proline to be a potential annotation for an unknown feature in *E. coli* at 159.0764 m/z. This annotation was confirmed (Fig. 5) by comparing the candidate's retention time and mass spectrum with a chemically synthesized standard (see methods for details).

Web access

The CD-MINE is accessible through a dedicated website that offers many features to query and display compound information (https://minedatabase.mcs.anl.gov/cdmine/#/home). Users may search the database by textual information like chemical formula or metabolic pathways, or by chemical structure to uncover novel metabolites. For a given chemical structure, they can explore molecular properties, and reactions that are predicted to generate or consume the compound.

4. Conclusion

"I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail." This observation by Abraham Maslow rings true for both fields metabolic engineering as well as synthetic biology. It is tempting to reduce metabolism to its enzymes and their catalyzed reactions only, when methodology allows experimentalists to manipulate enzyme concentrations expecting to improve a system's productivity. However, organisms' metabolic functions follow chemical rules, and both enzymes and compounds are subject to them. This is a fundamental axiom to keep when designing metabolic systems *in vitro* or *in silico*. We have shown how important it is to consider metabolites' chemical reactions when designing systems in the mentioned research fields.

Figure Legends

Fig. 1. CD-MINE construction workflow. Inputs(top), process (middle) and outputs available to the public (bottom), of CD-MINE database's development. Details are given in the results and methods sections.

Fig. 2. Incidence of spontaneous reactions as predicted by CD-MINE for the *Escherichia coli* production system of commercially valuable chemicals. The systems analyzed are the heterologous pathways introduced in *Escherichia coli* to synthesize 262 products by Zhang et al., 2016. Each final product has its own pathway: 123 central carbon metabolism pathways and 139 other metabolic pathways. The intermediates in each pathway were identified and matched to identifiers in the database, so its structure could be analyzed by the CD-MINE tools. The corresponding reaction rules were applied computationally to determine the possibility of a spontaneous reaction for each intermediate or final product. The digits inside each bar section represent the number of pathways where: a) only their intermediates may undergo spontaneous reactions (red); b) there could be spontaneous reactions at their synthesis endpoint (blue); c) there are no expected spontaneous reactions along them (green).

Fig. 3. Mapping damage-prone intermediates in the aromatic amino acids biosynthesis pathway. Number after the bar in each cell, represents quantity of damage products of such compound as predicted by our model (See Table S3). Color hue proportional to this number as shown on gradient bar.

Compounds lacking a colored bar are those whose derivatives were not reported, for reasons explained in methods. Only main intermediates are shown (for detailed pathway see Supplementary Fig.S2). Fusion events are represented by dotted lines and reactions catalyzed by such enzymes by contiguous arrows of the same color.

Fig. 4. Spontaneous and enzymatic reactions involving L-dopa and dopamine. Red arrows represent spontaneous reactions and blue arrows, enzymatic reactions. DDCX, Dopa decarboxylase [EC 4.1.1.28]; MAO, monoamine oxidase [EC 1.4.3.4], an enzyme with low substrate specificity. The reactive compounds are: **1**. L-Dopa; **2**. Dopamine; **3**. Dopal; **4**. For R = H, leucoaminochrome; for R = COOH, leucodopachrome or 5,6-dihydroxyindoline-2-carboxylic acid; **5**. For R = H, dopamine; for R = COOH, L-dopa. **6**. For R= H, 6-N-alanyl dopamine; for R= COOH, 6-N-alanyl dopa**; 7**. For R= H, and R^1= $C_3H_6NO_2$, 5-S-cysteinyl dopamine; for R= COOH, and R^1 = $C_3H_6NO_2$, 5-S-cysteinyl dopa. Also, for R= H, and R^1= $C_{10}H_{16}N_3O_6$, 5-S-glutathionyl dopamine; for R= COOH, and R^1 = $C_{10}H_{16}N_3O_6$, 5-S-glutathionyl dopa. **8**. For R = H, dopaminechrome; for R = COOH, dopachrome; **9**. For R = H, dopamine quinone; for R = COOH, L-dopa quinone; **10**. Examples only of cysteinyl adducts cyclization: for R = H, 7-(2-aminoethyl)-5-hydroxy-1,4-benzothiazine-3-carboxylate; for R = COOH, 1,4-benzothiazine.

Fig. 5. Prediction of N-carbamoyl proline as a damaged cellular product. (**A**) Spontaneous reaction of urea with carbomyl-proline to generate N-carbamoyl proline as revealed by reactions rules in CD-MINE. (**B**) Comparison of MS/MS spectra of the *Escherichia coli* product (*in vivo* synthesis) and the authenticated N-carbamoyl proline standard prepared as described in the methods section (*in vitro* synthesis); m/z, mass per charge ratio.

In folder:

Figure 1. flowchart of ModelSEED database's construction

Fig. 2. Incidence of spontaneous reactions as predicted by CD-MINE for the Escherichia coli production system of commercially valuable chemicals

Fig. 3. Mapping damage-prone intermediates in the aromatic amino acids biosynthesis pathway

Fig. 4. Spontaneous and enzymatic reactions involving L-dopa and dopamine

Fig. 5. Prediction of N-carbamoyl proline as a damaged cellular product

Supplementary Tables and Figures

On line:

Table S1: Spontaneous reactions from literature and MetaCyc:

http://minedatabase.mcs.anl.gov/cdmine/#/datatables/litrxns

Table S2: List of CD-MINE operators:

http://minedatabase.mcs.anl.gov/cdmine/#/datatables/rxnrules

Table S3. List of reactive metabolites and end products in the ModelSEED database

Table S4. Distribution of reactive compounds in the metabolic pathway network

Table S5. Examples of side reactions of enzymes with low substrate specificity or catalytic promiscuity

Fig. S1A and S1B. Comparison of occurrence of reactive intermediates in different metabolic pathways

Fig. S2: Damage-prone metabolites in aromatic amino acids biosynthesis

Fig. S3: Damage-prone metabolites in phenylalanine metabolism

Fig. S4. Damage-prone metabolites in tyrosine metabolism

Fig. S5. Damage-prone metabolites in lysine biosynthesis

Fig. S6. Damage prone metabolites in lysine degradation

Figure legends

Fig. S1A and S1B. Comparison of occurrence of reactive intermediates in different metabolic pathways. Values represent the number of products of spontaneous reactions undergone by the metabolites participating in each pathway, and the names of the latter are based on the KEGG database classification (Table S4). Values have been corrected by filtering out products of cofactors or metabolites which participate in more than ten pathways (methods). Abbreviations and symbols: glycosylphosphatidylinositol (GPI); glycosaminoglycan (GAG); glycosphingolipid (GSP); cytochrome (cyt); dichlorodiphenyltrichloroethane (DDT). Pathways are grouped in relation to their metabolic areas and their corresponding bars colored accordingly.

Fig. S2 to S6: Damage prone hot spots in some amino acid metabolic pathways. Arteries based on KEGG database's network, mapped with the Escher program (Rowe et al., 2018) and integrated into the ModelSEED hub (https://modelseed.github.io/). Reactions identified by rxn numbers given in ModelSEED (http://modelseed.org/); Standard abbreviations used for cofactors names. Non-standard abbreviations: PACT: phenylacetic acid; SALC: salicylic acid; DHI: 5,6-dihydroxyindole or dopamine lutine; DHICA: 5,6-dihydroxyindole-2-carboxylate; T4: thyroxine or 2-amino-3-[4-(4-hydroxy-3,5-diiodophenoxy)-3,5-diiodophenyl]propanoic acid. DAHP: dihydroxyacetone phosphate; PQQH2; reduced pyrroloquinoline-quinone; Aminoadip_S, alpha-Aminoadipoyl-S-acyl enzyme. Numbers after metabolites names show the amounts of derivatives that could be produced by spontaneous reactions of such compounds according to the CD-MINE operator rules and calculated as explained in methods (Table S3). Hot spots circles' size and color hue proportional to corresponding metabolite reactivity. Notice that L-2-aminoadipate 6-semialdehyde is also called allysine.

Missing: Table legends for Tables S3 and S4.

Table S5. Examples of side reactions of enzymes with low substrate specificity (a) or catalytic promiscuity (b). Cases of enzymes belonging to most groups (oxidireductases, transferases, hydrolases, lyases, isomerases and ligases) and organized by EC numbers are shown with their prevalent activity and side reactions. Damage reactions are followed by repair or prevention enzymatic reactions and examples are taken from bacteria, fungi, plants and animals. Literature references for each reaction are given in the corresponding row as PMID numbers (if applicable).

Figures

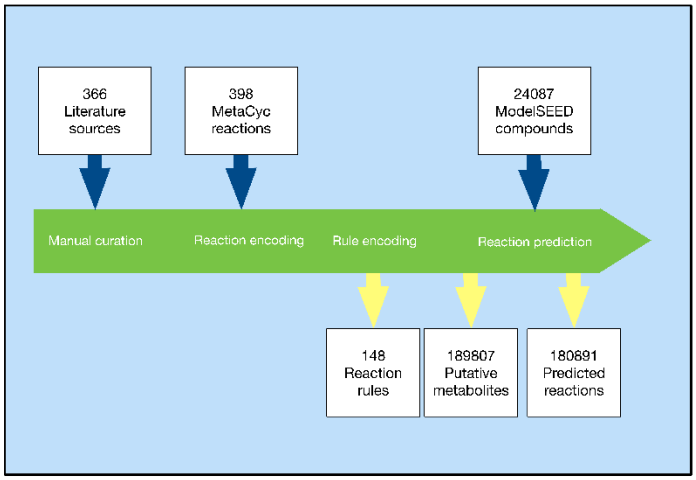


Fig. 1. **CD-MINE construction workflow**

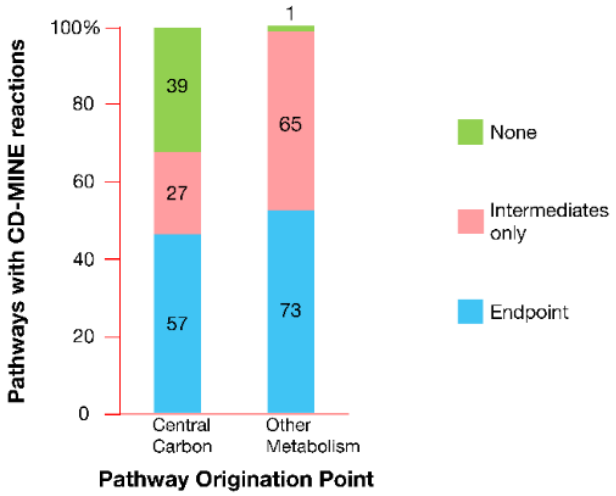


Fig. 2 . **Incidence of spontaneous reactions as predicted by CD-MINE for the *Escherichia coli* production system of commercially valuable chemicals**

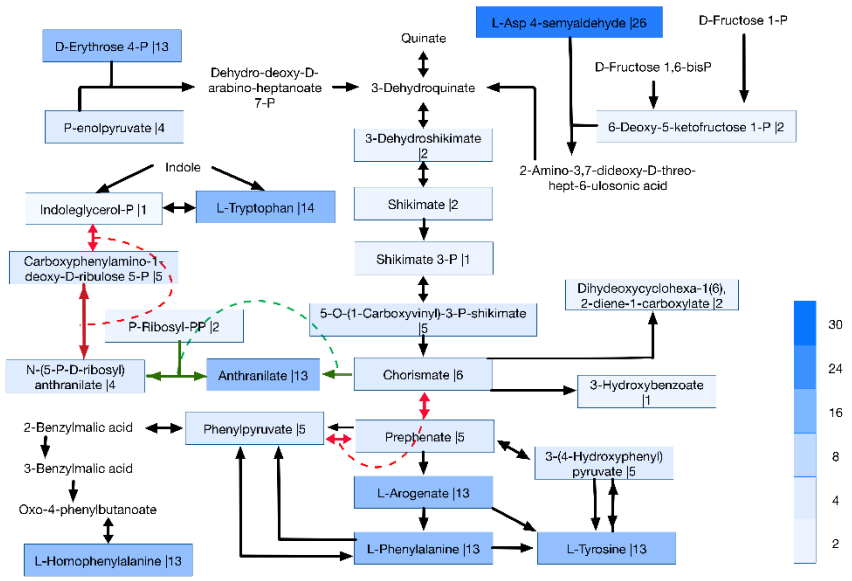


Fig. 3. **Mapping damage-prone intermediates in the aromatic amino acids biosynthesis pathway**

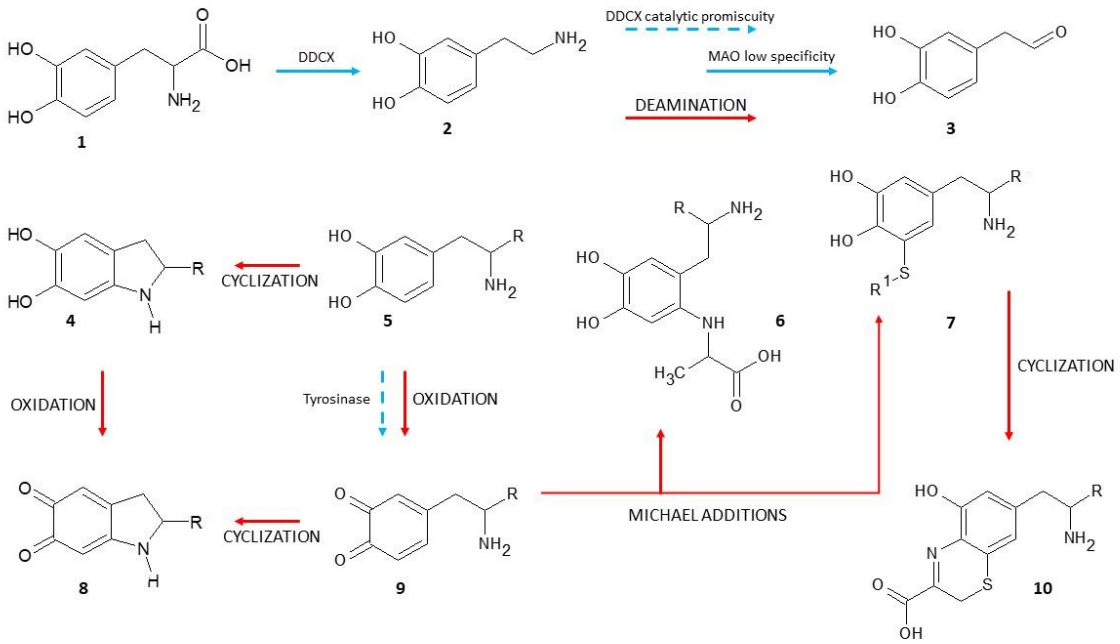


Fig. 4. **Spontaneous and enzymatic reactions involving L-dopa and dopamine**

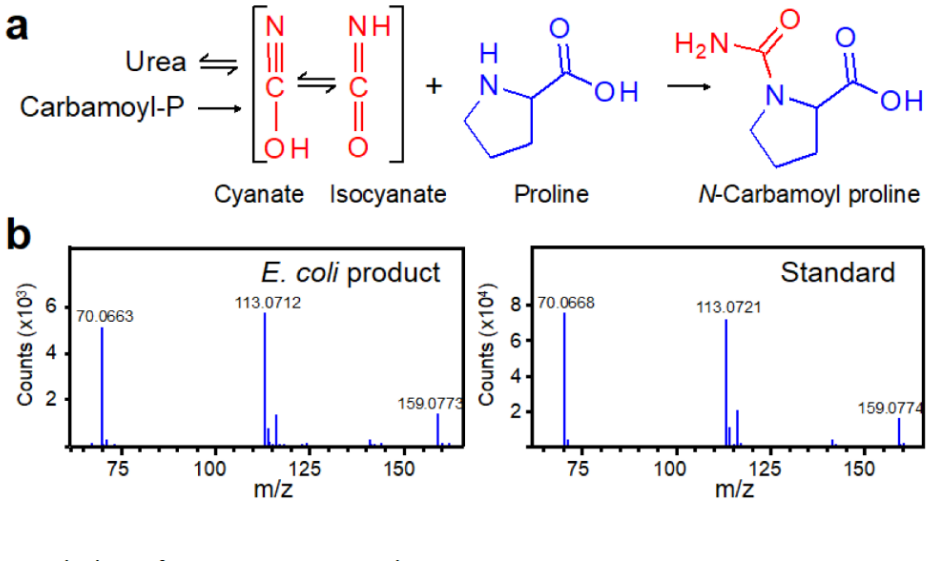


Fig. 5. Prediction of N-carbamoyl proline as a damaged cellular product

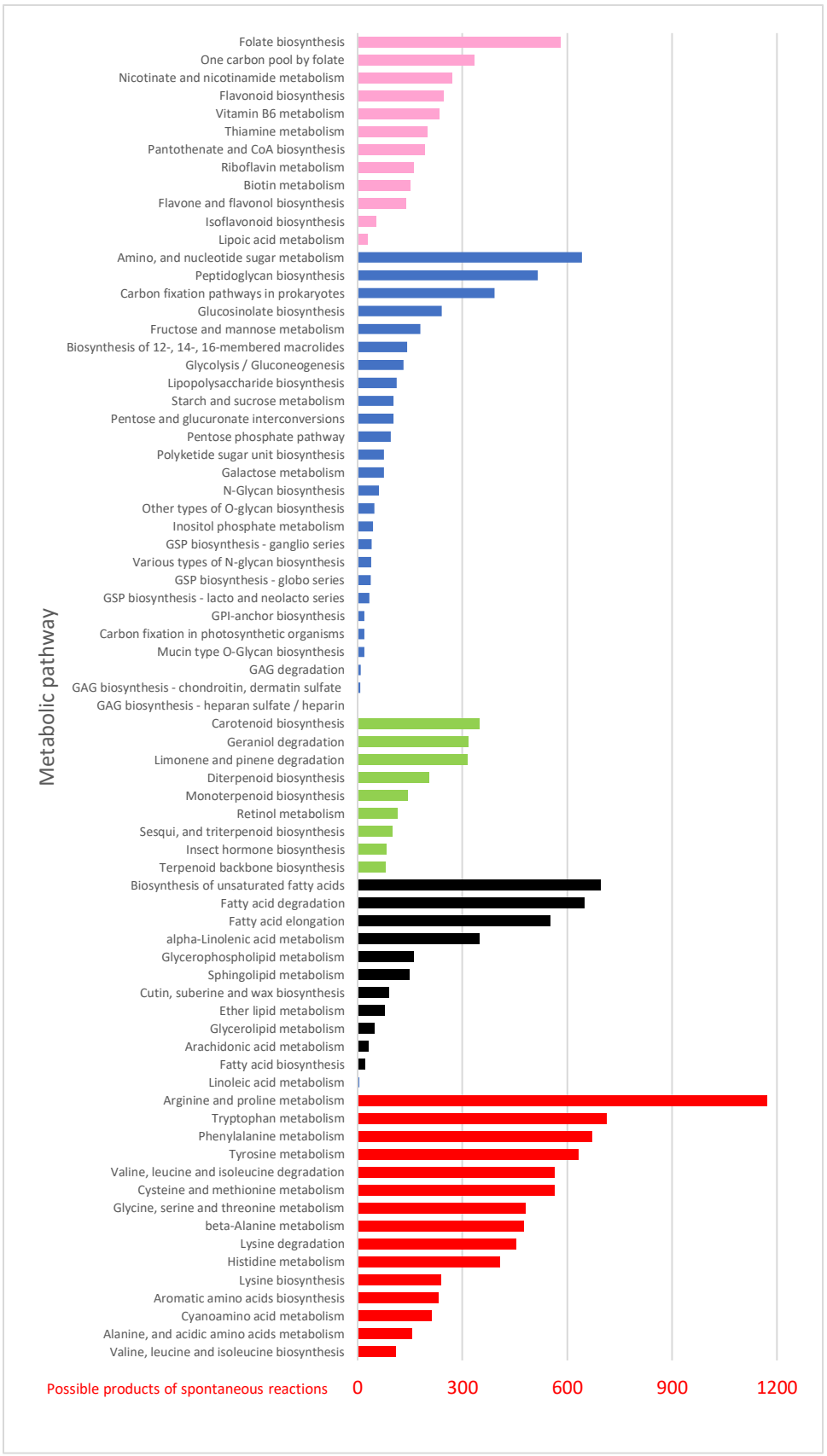


Fig. S1A. Comparison of occurrence of reactive intermediates in different metabolic pathways

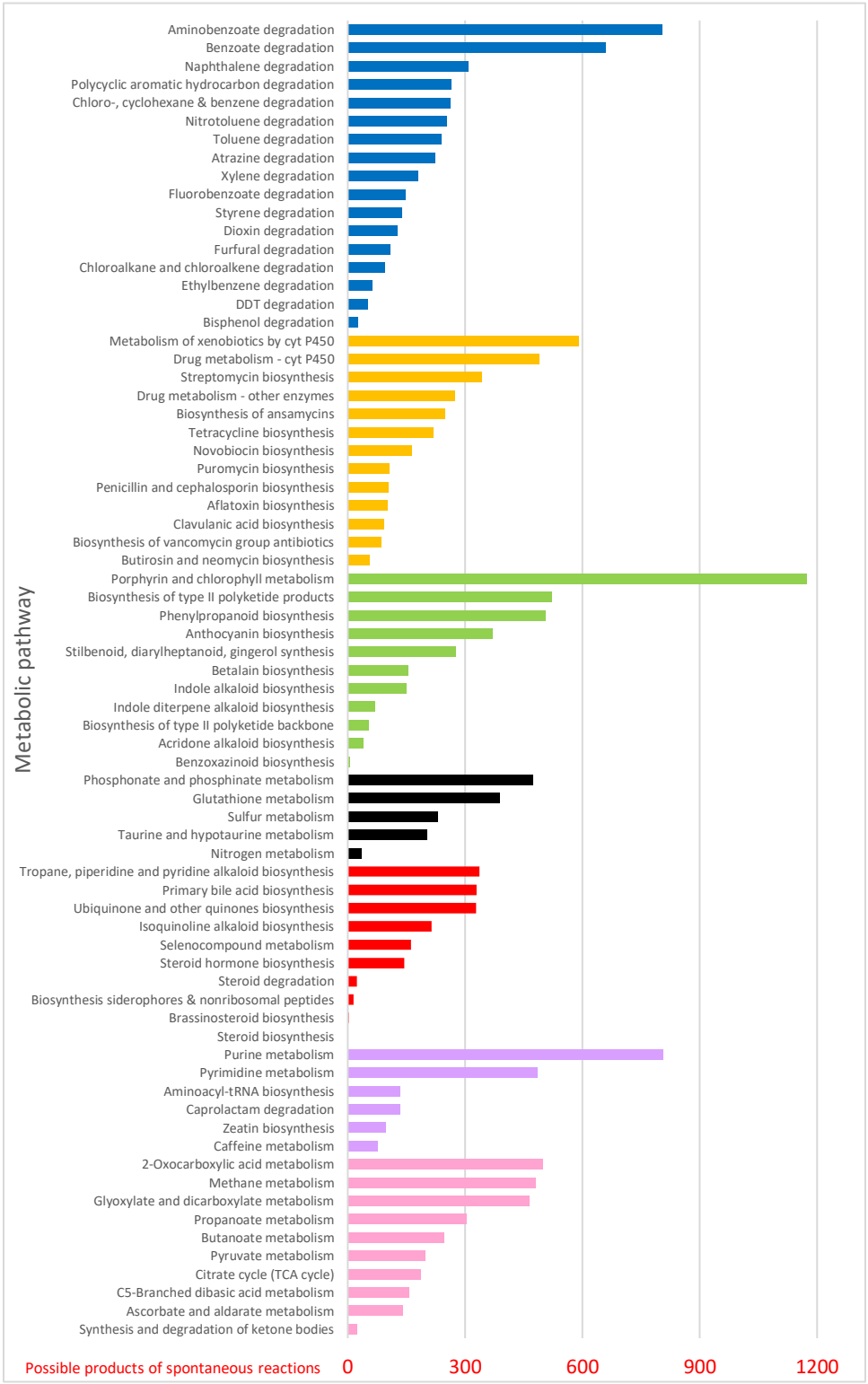
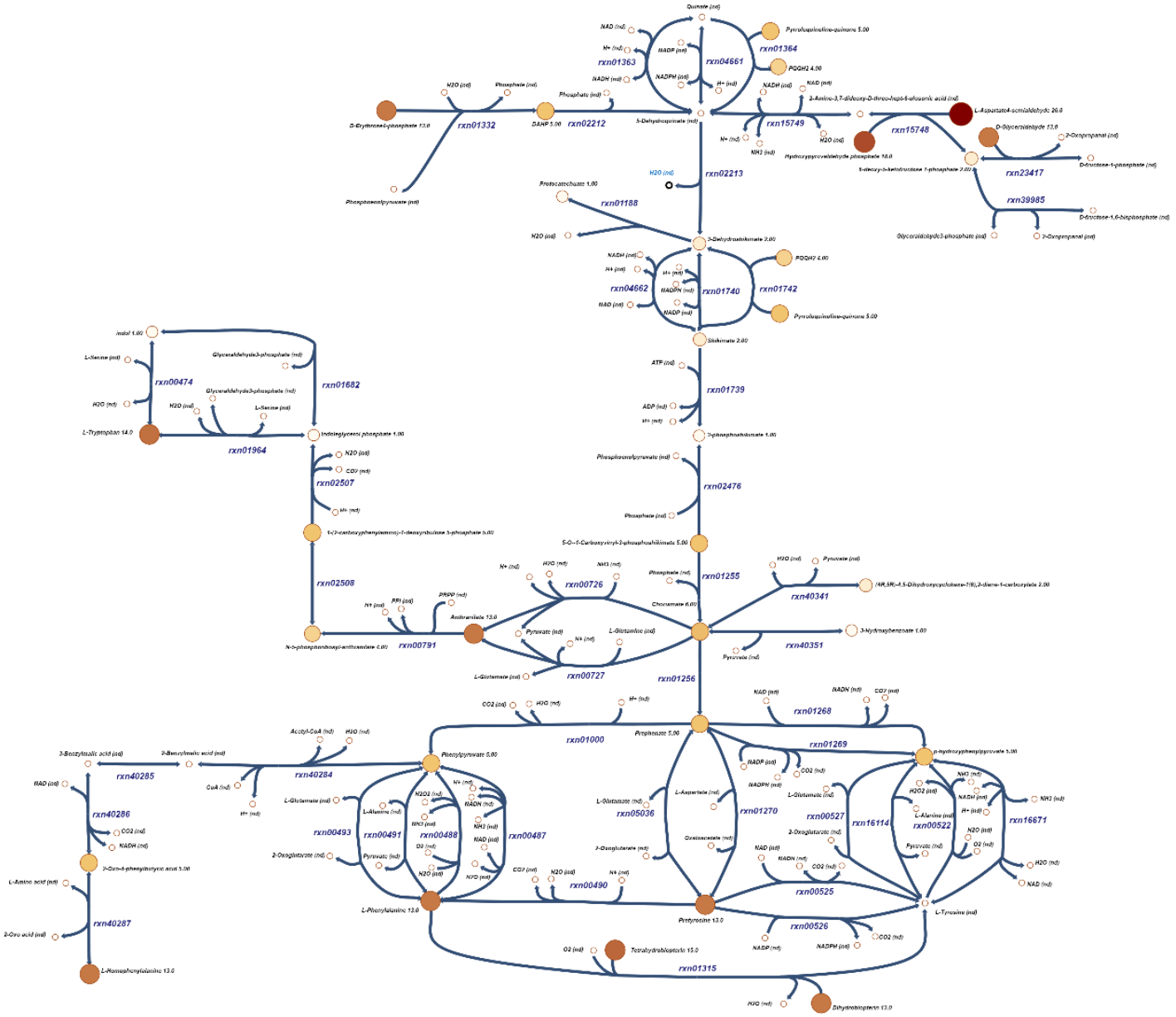


Fig. S1B. Comparison of occurrence of reactive intermediates in different metabolic pathways



Aromatic Amino Acids Biosynthesis

Fig. S2: Damage-prone metabolites in aromatic amino acids biosynthesis

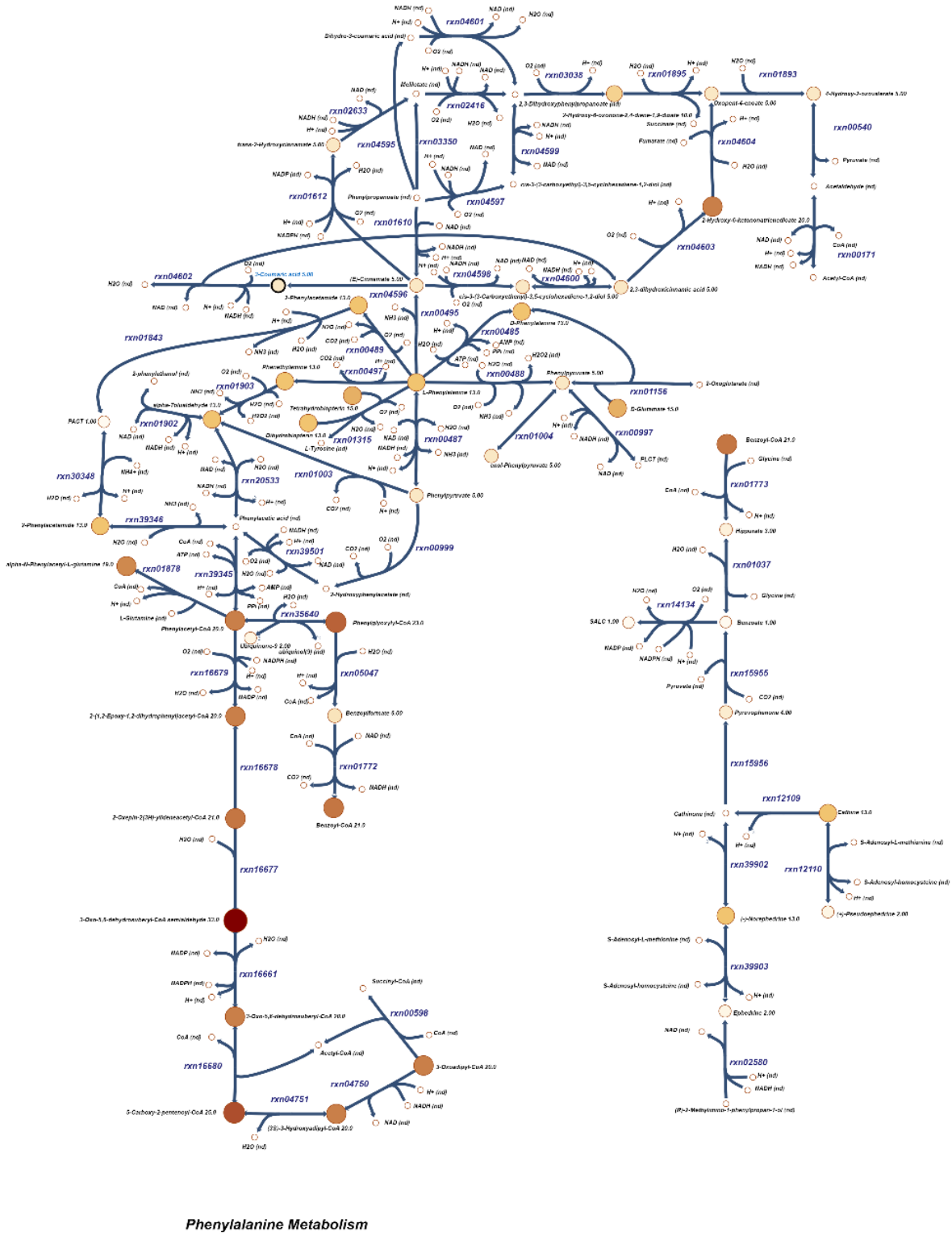
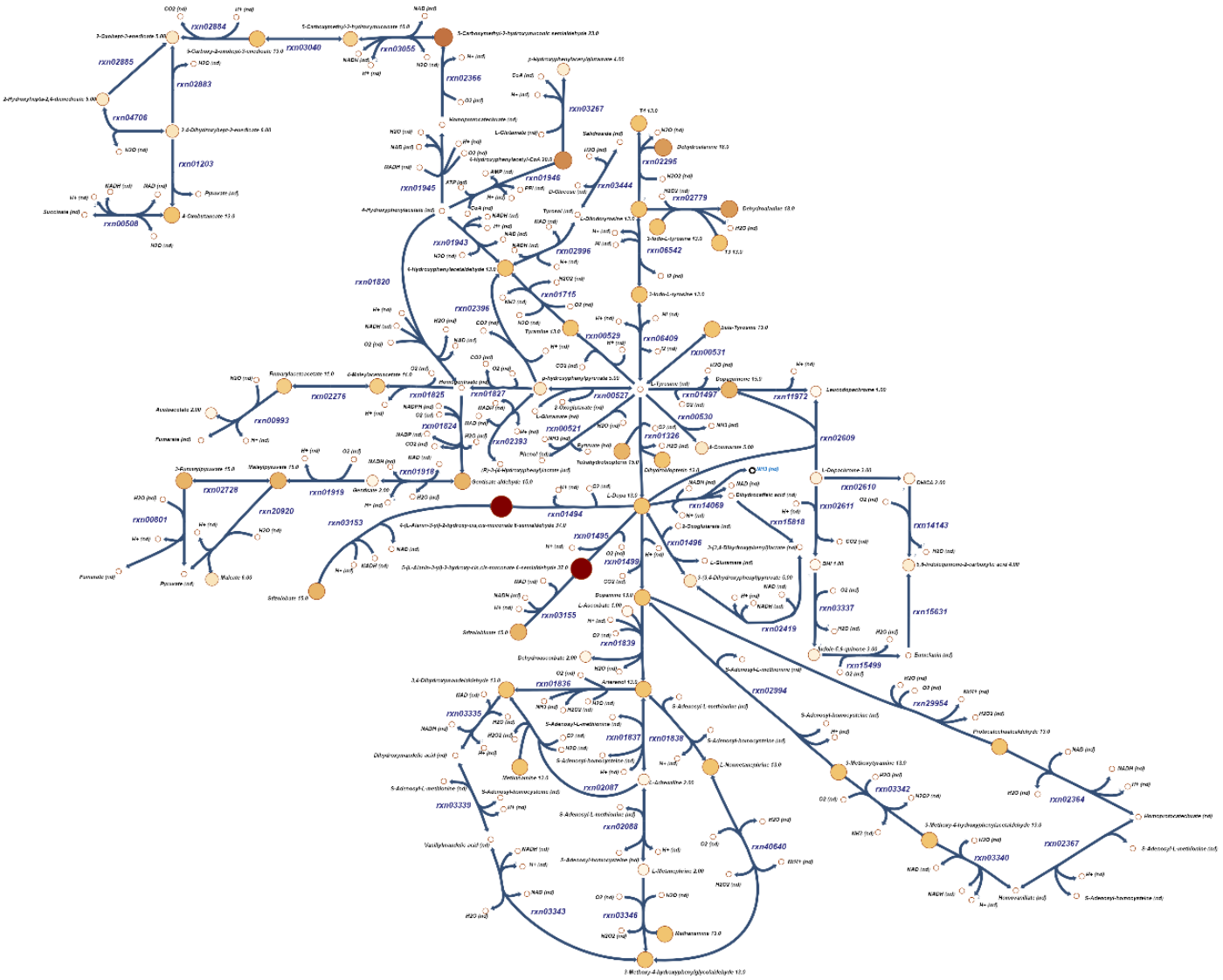
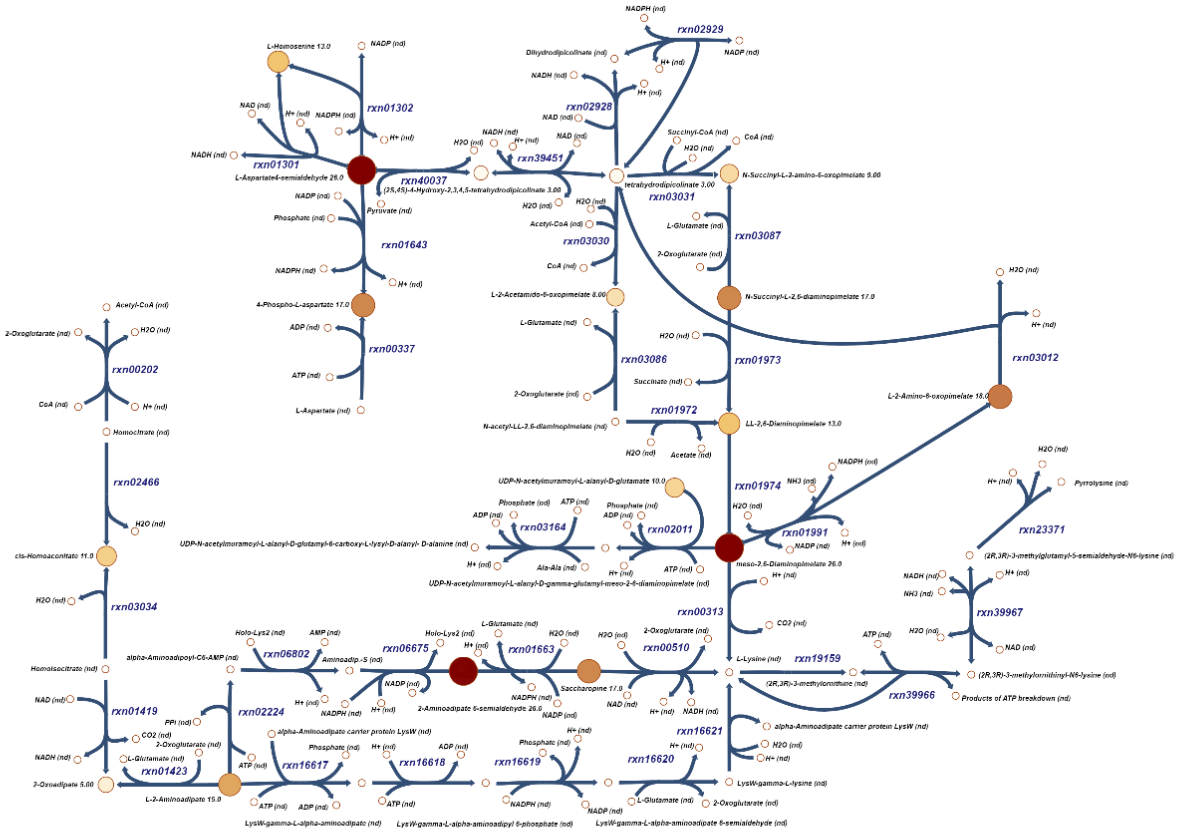


Fig. S3. **Damage-prone metabolites in phenylalanine metabolism**



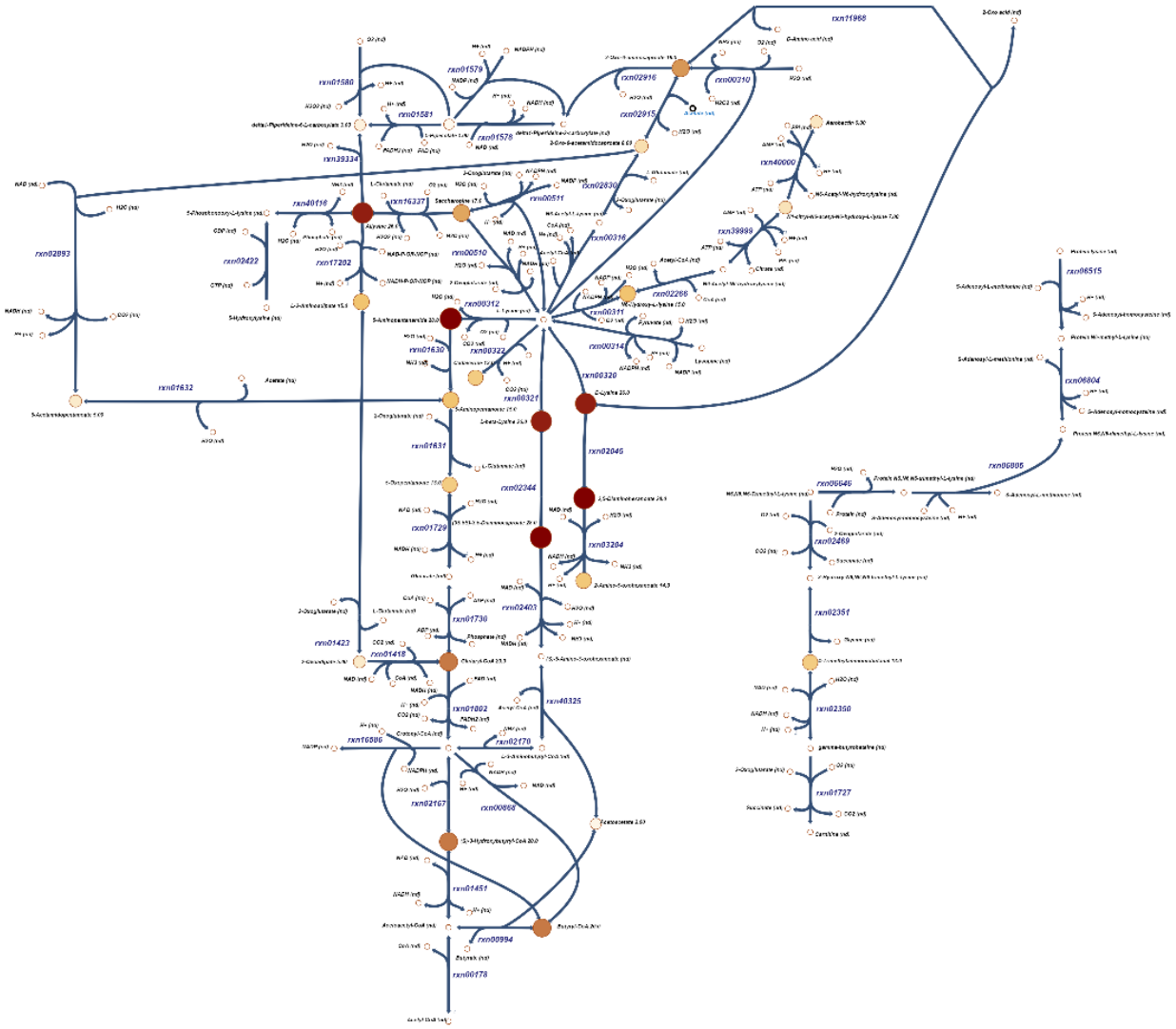
Tyrosine Metabolism

Fig. S4. Damage-prone metabolites in tyrosine metabolism



Lysine Biosynthesis

Fig. S5. Damage-prone metabolites in lysine biosynthesis



Lysine Degradation

Fig. S6. Damage prone metabolites in lysine degradation

References

Agrup, G., Hansson, C., Rorsman, H. Rosengren, E., 1982. The effect of cysteine on oxidation of tyrosine, dopa, and cysteinyldopas. Arch Dermatol Res. 272,103-115. https://doi.org/10.1007/bf00510400

Anderson, D.G., Florang, V.R., Schamp, J.H., Buettner, G.R., Doorn, J.A., 2016. Antioxidant-Mediated Modulation of Protein Reactivity for 3,4-Dihydroxyphenylacetaldehyde, a Toxic Dopamine Metabolite. Chem Res Toxicol. 29, 1098-1107. https://doi.org/10.1021/acs.chemrestox.5b00528

Andre, C., Kim, S.W., Yu, X.H., Shanklin, J., 2013. Fusing catalase to an alkane-producing enzyme maintains enzymatic activity by converting the inhibitory byproduct H2O2 to the cosubstrate O2. Proc. Natl. Acad. Sci. U S A 110, 3191-3196. https://doi.org/10.1073/pnas.1218769110

Arai, R., Karasawa, N., Nagatsu, I., 1996. Dopamine produced from L-DOPA is degraded by endogenous monoamine oxidase in neurons of the dorsal raphe nucleus of the rat: an immunohistochemical study. Brain Res. 722, 181-184. https://doi.org/10.1016/0006-8993(96)00252-1

Balderas-Hernández, V.E., Sabido-Ramos, A., Silva, P., Cabrera-Valladares, N., Hernández-Chávez, G., Báez-Viveros, J.L., Martínez, A., Bolívar, F., Gosset, G., 2009. Metabolic engineering for improving anthranilate synthesis from glucose in Escherichia coli. Microb Cell Fact. 8, 19. doi: 10.1186/1475-2859-8-19

Bernoud-Hubac, N., Davies, S.S., Boutaud, O., Montine, T.J., Roberts, L.J.,2001. Formation of highly reactive gamma-ketoaldehydes (neuroketals) as products of the neuroprostane pathway. J Biol Chem. 276, 30964-30970. https://doi.org/10.1074/jbc.M103768200

Bertoldi, M., Voltattorni, C.B., 2009. Multiple roles of the active site lysine of Dopa decarboxylase. Arch Biochem Biophys. 488, 130-139. https://doi.org/10.1016/j.abb.2009.06.019.

Bettey, M., Ireland, R.J., Smith, A.M., 1992. Purification and characterization of acetyl CoA carboxylase from developing pea embryos. J. Plant Physiol. 140, 513-520.

Boyle, P.M., Silver, P.A., 2012. Parts plus pipes: synthetic biology approaches to metabolic engineering. Metab. Eng. 14, 223-232. https://doi.org/10.1016/j.ymben.2011.10.003

Braun, K., Gibson, D.T., 1984. Anaerobic degradation of 2-aminobenzoate (anthranilic acid) by denitrifying bacteria. Appl Environ Microbiol. 48, 102-107.

Bommer, G.T., Van Schaftingen, E., Veiga-da-Cunha, M., 2019. Metabolite Repair Enzymes Control Metabolic Damage in Glycolysis. Trends Biochem Sci. pii: S0968-0004, 30176-30178. https://doi.org/10.1016/j.tibs.2019.07.004

Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D.S., Karp, P.D., 2016. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 44, D471-D480. https://doi.org/10.1093/nar/gkv1164

Chan, D., Vogel, H.J., 2010. Current understanding of fatty acid biosynthesis and the acyl carrier protein. Biochem J. 430:1-19. https://doi.org/10.1042/BJ20100462.

Colijn, C., Brandes, A., Zucker, J., Lun, D.S., Weiner, B., Farhat, M.R., Cheng, T.Y., Moody, D.B., Murray, M., Galagan, J.E., 2009. Interpreting expression data with metabolic flux models: predicting Mycobacterium tuberculosis mycolic acid production. PLoS Comput. Biol. 5, e1000489. https://doi.org/10.1371/journal.pcbi.1000489.

Danchin, A., Sekowska, A., 2014. The logic of metabolism and its fuzzy consequences. Environ. Microbiol. 16, 19-28. https://doi.org/10.1111/1462-2920.12270

DeBerardinis, R. J.,Thompson, C.B., 2012. Cellular metabolism and disease: what do metabolic outliers teach us? Cell 148, 1132-1144. https://doi.org/10.1016/j.cell.2012.02.032

Delgado, G.E., Seijas, L.E., Mora, A.J., González, T., Briceño, A., 2012. Synthesis, Crystal Structure and Hydrogen-Bonding Patterns in (RS)-1-Carbamoyl Pyrrolidine-2-Carboxylic Acid. J. Chem. Crystallogr. 42, 388–393. https://doi.org/10.1107/S0108270111054977

Desai, D.K., Nandi, S., Srivastava, P.K., Lynn, A.M., 2011. ModEnzA: Accurate Identification of Metabolic Enzymes Using Function Specific Profile HMMs with Optimised Discrimination Threshold and Modified Emission Probabilities. Adv. Bioinformatics 2011, 743782. https://doi.org/10.1155/2011/743782.

Dromms, R.A., Styczynski, M.P., 2012. Systematic applications of metabolomics in metabolic engineering. M etabolites 2, 1090-1122. https://doi.org/10.3390/metabo2041090

Dueber, J.E., Wu, G.C., Malmirchegini, G.R., Moon, T.S., Petzold, C.J., Ullal, A.V., Prather, K.L.J., Keasling, J.D., 2009. Synthetic protein scaffolds provide modular control over metabolic flux. Nat. Biotechnol. 27, 753-759. https://doi.org/10.1038/nbt.1557

Edmondson, D.L., , Kane, H.J., Andrews, T.J., 1990. Substrate isomerization inhibits ribulosebisphospate carboxylaseoxygenase during catalysis. FEBS letters 260, 62-66.

El-Sayed, A., Shindia, A., Zaher, Y., 2013. Purification and characterization of L-amino acid oxidase from the solid-state grown cultures of Aspergillus oryzae ASH. Microbiology 82, 762-771 (2013)

Felsted, R.L., Chu, A.E., Chaykin, S., 1973. Purification and properties of the aldehyde oxidases from hog and rabbit livers. J Biol Chem. 248, 2580-2587.

Frelin, O., Huang, L., Hasnain, G., Jeffryes, J.G., Ziemak, M.J., Rocca, J.R., Wang, B., Rice, J., Roje, S., Yurgel, S.N., Gregory, J.F., Edison, A.S., Henry, C.S., de Crécy-Lagard, V., Hanson, A.D., 2015. A directed-overflow and damage-control N-glycosidase in riboflavin biosynthesis. Biochem. J. 466, 137-145. https://doi.org/10.1042/BJ20141237

Golubev, A., Hanson, A.D., Gladyshev, V.N., 2017. Non-enzymatic molecular damage as a prototypic driver of aging. J. Biol. Chem. 292, 6029-6038. https://doi.org/10.1074/jbc.R116.751164

Hanson, A.D., Henry, C.S., Fiehn, O., de Crécy-Lagard, V., 2016. Metabolite Damage and Metabolite Damage Control in Plants. Annu. Rev. Plant Biol. 67, 131-152. https://doi.org/10.1146/annurev-arplant-043015-111648.

Halket, J.M., Waterman, D., Przyborowska, A.M., Patel, R.K.P., Fraser, P.D., Bramley, P.M., 2005. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. J. Exp. Bot. 56, 219-243. https://doi.org/10.1093/jxb/eri069

Henry, C.S., DeJongh, M., Best, A.A., Frybarger, P.M., Linsay, B., Stevens, R.L., 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat. Biotechnol. 28, 977-982. https://doi.org/10.1038/nbt.1672

Henry, C.S., Lerma-Ortiz, C., Gerdes, S.Y., Mullen, J.D., Colasanti, R., Zhukov, A., Frelin, O., Thiaville, J.J., Zallot, R., Niehaus, T.D., Hasnain, G., Conrad, N., Hanson, A.D., de Crécy-Lagard, V., 2016. Systematic identification and analysis of frequent gene fusion events in metabolic pathways. BMC Genomics 17, 473. https://doi.org/ 10.1186/s12864-016-2782-3

Hill, J.P., Dickinson, F.M., 1988. Pre-steady-state kinetics of aldehyde oxidation by pig liver cytosolic aldehyde dehydrogenase. Biochem. Soc. Trans. 16, 856-857.

Houten, S.M., Violante, S., Ventura, F.V., Wanders, R.J., 2016. The Biochemistry and Physiology of Mitochondrial Fatty Acid β-Oxidation and Its Genetic Disorders. Annu Rev Physiol. 78, 23-44. https://doi.org/10.1146/annurev-physiol-021115-105045.
PMID: 26474213
Jayaseelan, KV., Moreno, P., Truszkowski, A., Ertl, P., Steinbeck, C., 2012. Natural product-likeness score revisited: an open-source, open-data implementation. BMC Bioinformatics. May 20, 2012. https://doi.org/ 10.1186/1471-2105-13-106.

Jeffryes, J.G., Colastani, R.L., Elbadawi-Sidhu, M., Kind, T., Niehaus, T.D., Broadbelt, L.J., Hanson, A.D., Fiehn, O., Tyo, K.E.J., Henry, C.S., 2015. MINEs: open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics. J Cheminform. 7, 44. https://doi.org/10.1186/s13321-015-0087-1.

Jeffryes, J., Strutz, J., Henry, C.,Tyo, K.E.J., 2019. Metabolic In Silico Network Expansions to Predict and Exploit Enzyme Promiscuity. Methods Mol. Biol. 1927, 11-21. https://doi.org/10.1007/978-1-4939-9142-6_2

Jones, J.A., Toparlak, T.D., Koffas, M.A., 2015. Metabolic pathway balancing and its role in the production of biofuels and chemicals. Curr. Opin. Biotechnol. 33 ,52-59.

https://doi.org/10.1016/j.copbio.2014.11.013

Kahn, A., Marie, J., 1982. Pyruvate kinases from human erythrocytes and liver. Methods Enzymol. 90 Pt E:131-140.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M.,Tanabe, M., 2014. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 42, D199-205. https://doi.org/10.1093/nar/gkt1076

Keller, M.A., Piedrafita, G., Ralser, M., 2015. The widespread role of non-enzymatic reactions in cellular metabolism. Curr. Opin. Biotechnol. 34, 153-161. https://doi.org/10.1016/j.copbio.2014.12.020

Kerfeld, C.A., Erbilgin O., 2015. Bacterial microcompartments and the modular construction of microbial metabolism. Trends Microbiol. 23, 22-34. https://doi.org/10.1016/j.tim.2014.10.003

Khosla, P.K., Neeraj, V.I., Gupta, S.K., Satpathy, G., 1992. Berberine, a potential drug for trachoma. Rev Int Trach Pathol Ocul Trop Subtrop Sante Publique. 69: 147-165.

Kildegaard, K.R., Hallström, B.M., Blicher, T.H., Sonnenschein, N., Jensen, N.B., Sherstyk, S., Harrison, S.J., Maury, J., Herrgård, M.J., Juncker, A.S., Forster, J., Nielsen, J., Borodina, I., 2014. Evolution reveals a glutathione-dependent mechanism of 3-hydroxypropionic acid tolerance. Metab. Eng. 26:57-66. https://doi.org/10.1016/j.ymben.2014.09.004

Koek, M.M., Jellema, R.H., van der Greef, J., Tas, A.C., Hankemeier, T., 2011. Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. Metabolomics 7, 307-328. https://doi.org/10.1007/s11306-010-0254-3

Koppel, N., Balskus, E.P., 2016. Exploring and Understanding the Biochemical Diversity of the Human Microbiota. Cell Chem. Biol. 23, 18-30. https://doi.org/10.1016/j.chembiol.2015.12.008

Lanzafame GM, Sarakha M, Fabbri D, Vione D (2017) Degradation of Methyl 2-Aminobenzoate (Methyl Anthranilate) by H_2O_2/UV: Effect of Inorganic Anions and Derived Radicals. Molecules 22(4). pii: E619. https://doi.org/10.3390/molecules22040619.

Lerma-Ortiz, C., Jeffryes, J.G., Cooper, A.J.L., Niehaus, T.D., Thamm, A.M.K., Frelin, O., Aunins, T., Fiehn, O., de Crécy-Lagard, V., Henry, C.S., Hanson, A.-D., 2016. 'Nothing of chemistry disappears in biology': the Top 30 damage-prone endogenous metabolites. Biochem. Soc. Trans. 44, 961-971. https://doi.org/10.1042/BST20160073

Li, C., Wang, M.H., 2014. Potential Biological Activities of Magnoflorine: A Compound from Aristolochia debilis. Korean J. Plant Res. 27 :223-228. http://dx.doi.org/10.7732/kjpr.2014.27.3.223

Li, J., Zhang, L., Liu, W., 2018. Cell-free synthetic biology for in vitro biosynthesis of pharmaceutical natural products. Synth. Syst. Biotechnol. 3, 83-89. https://doi.org/10.1016/j.synbio.2018.02.002

Linster, C.L., Van Schaftingen, E., Hanson, A.D., 2013. Metabolite damage and its repair or pre-emption. Nat. Chem. Biol. 9, 72-80. https://doi.org/10.1038/nchembio.1141.

Marchitti, S.A., Deitrich, R.A., Vasiliou, V., 2007. Neurotoxicity and metabolism of the catecholamine-derived 3,4-dihydroxyphenylacetaldehyde and 3,4-dihydroxyphenylglycolaldehyde: the role of aldehyde dehydrogenase. Pharmacol Rev. 59 :125-150. https://doi.org/10.1124/pr.59.2.1

Minami, H., Kim, J.S., Ikezawa, N., Takemura, T., Katayama, T., Kumagai, H., Sato, F., 2008. Microbial production of plant benzylisoquinoline alkaloids. Proc Natl Acad Sci U S A. 105: 7393-7398. https://doi.org/10.1073/pnas.0802981105.

Morrow, J.D., Hill, K.E., Burk, R.F., Nammour, T.M., Badr, K.F., Roberts, L.J., 2nd., 1990. A series of prostaglandin F2-like compounds are produced in vivo in humans by a non-cyclooxygenase, free radical-catalyzed mechanism. Proc Natl Acad Sci U S A. 87, 9383-9387. https://doi.org/10.1073/pnas.87.23.9383

Nakagawa, A., Minami, H., Kim, J.S., Koyanagi, T., Katayama, T., Sato, F., Kumagai, H., 2011. A bacterial platform for fermentative production of plant alkaloids. Nat Commun. 2:326. https://doi.org/10.1038/ncomms1327. Erratum in: Nat Commun. 2012;3. https://doi.org/10.1038/ncomms1866.

O'Boyle, N.M., 2012. Towards a Universal SMILES representation - A standard method to generate canonical SMILES based on the InChI. J Cheminform. 4, 22. https://doi.org/10.1186/1758-2946-4-22

Olajuyigbe, F.M., Demitri, N., Ajele, J.O., Maurizio, E., Randaccio, L., Geremia, S., 2010. Carbamylation of N-terminal proline. ACS Med. Chem. Lett. 1, 254-257. https://doi.org/10.1021/ml100046d

Ortiz-Ruiz, C.V., Berna, J., Garcia-Molina, Mdel M., Tudela, J., Tomas, V., Garcia-Canovas, F., 2015. Identification of p-hydroxybenzyl alcohol, tyrosol, phloretin and its derivate phloridzin as tyrosinase substrates. Bioorg Med Chem. 23: 3738-3746. https://doi.org/10.1016/j.bmc.2015.04.016.

Overhage, J., Sielker, S., Homburg, S., Parschat, K., Fetzner,S., 2005. Identification of large linear plasmids in Arthrobacter spp. encoding the degradation of quinaldine to anthranilate. Microbiology. 151, 491-500. https://doi.org/10.1099/mic.0.27521-0

Park, J.O., Rubin, S.A., Xu, Y., Amador-Noguez, D., Fan, J., Shlomi, T., Rabinowitz, J.D., 2016. Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. Nat. Chem. Biol. 12, 482-489. https://doi.org/10.1038/nchembio.2077

Phaniendra, A., Jestadi, D.B., Periyasamy, L., 2015. Free radicals: properties, sources, targets, and their implication in various diseases. Indian J. Clin. Biochem. 30, 11-26. https://doi.org/10.1007/s12291-014-0446-0

Plaut, G.W.E., 1963. Isocitrate dehydrogenases. In: Boyer, P.D., Lardy, H., Myrbäck, K. (Eds.), The Enzymes, 2nd ed., vol. 7, Academic Press, New York, 1963, pp. 105-126.

Poust, S., Piety, J., Bar-Even, A., Louw, C., Baker, D., Keasling, J.D., Siegel, J.B., 2015. Mechanistic Analysis of an Engineered Enzyme that Catalyzes the Formose Reaction. Chembiochem. 16, 1950-1954. https://doi.org/10.1002/cbic.201500228

Rowe E., Palsson, B.O., King, Z.A., 2018. Escher-FBA: a web application for interactive flux balance analysis. BMC Syst Biol. 12, 84. https://doi.org/10.1186/s12918-018-0607-5

Ryan, A., Kaplan, E., Nebel, J.C., Polycarpou, E., Crescente, V., Lowe, E., Preston, G.M., Sim, E., 2014. Identification of NAD(P)H quinone oxidoreductase activity in azoreductases from P. aeruginosa: azoreductases and NAD(P)H quinone oxidoreductases belong to the same FMN-dependent superfamily of enzymes. PLoS One 9, e98551. https://doi.org/10.1371/journal.pone.0098551

Seaver, S.M.D., Gerdes, S., Frelin, O., Lerma-Ortiz, C., Bradbury, L.M.T., Zallot, R., Hasnain, G., Niehaus, T.D., El Yacoubi, B., Pasternak, S., Olson, R., Pusch, G., Overbeek, R., Stevens, R., de Crécy-Lagard, V., Ware, D., Hanson, A.D., Henry, C.S., 2014. High-throughput comparison, functional annotation, and metabolic modeling of plant genomes using the PlantSEED resource. Proc. Natl. Acad. Sci. U S A 111, 9645-9650. https://doi.org/10.1073/pnas.1401329111

Sévin, D.C., Kuehne, A., Zamboni, N., Sauer, U., 2015. Biological insights through nontargeted metabolomics. Curr. Opin. Biotechnol. 34, 1-8. https://doi.org/10.1016/j.copbio.2014.10.001

Smith, C.A., O'Maille, G., Want, E.J., Qin, C., Trauger, S.A., Brandon, T.R., Custodio, D.E., Abagyan, R., Siuzdak, G., 2005. METLIN: a metabolite mass spectral database. Ther Drug Monit. 27, 747-751. https://doi.org/10.1097/01.ftd.0000179845.53213.39

Sun, J., Jeffryes, J.G., Henry, C.S., Bruner, S.D., Hanson, A.D., 2017. Metabolite damage and repair in metabolic engineering design. Metab Eng. 44, 150-159. https://doi.org/10.1016/j.ymben.2017.10.006

Van Schaftingen, E., Rzem, R., Veiga-da-Cunha, M.,2009. L: -2-Hydroxyglutaric aciduria, a disorder of metabolite repair. J Inherit Metab Dis. 32, 135-142. https://doi.org/ 10.1007/s10545-008-1042-3.

vander Jagt, D.L., 1989. The glyoxalase system, coenzymes and cofactors. Glutathione: chemical, biochemical, and medical aspects. Part. A (Dolphin, D., Poulson, R., Avromonic, O., eds.) John Wiley & Sons, New York 3, 597-641.

Vianello R and Domenece C and Mavri J (2017) The use of multiscale molecular simulations in understanding a relationship between the structure and function of biological systems of the brain. The application of monoamine oxidase enzymes. In the book: Structure-based drug design for diagnosis and treatment of neurological diseases. Ramsay RR and Di Gionanni G editors. Frontiers in Pharmacology 24. Neuropharmacology. https://doi.org/10.3389/fphar.2017.00013

Withers, S.T., Gottlieb, S.S., Lieu, B., Newman, J.D., and Keasling, J.D., 2007. Identification of isopentenol biosynthetic genes from Bacillus subtilis by a screening method based on isoprenoid precursor toxicity. Appl. Environ. Microbiol. 73, 6277-6283. https://doi.org/10.1128/AEM.00861-07

Wishart, D.S., Jewison, T., Guo, A.C., Wilson, M., Knox, C., Liu, Y., Djoumbou, Y., Mandal, R., Aziat, F., Dong, E., Bouatra, S., Sinelnikov, I., Arndt, D., Xia, J., Liu, P., Yallou, F., Bjorndahl, T., Perez-Pineiro, R., Eisner, R., Allen, F., Neveu, V., Greiner, R., Scalbert, A., 2013. HMDB 3.0--The Human Metabolome Database in 2013. Nucleic Acids Res. 41, D801-807. https://doi.org/10.1093/nar/gks1065

Wu, C., Visscher, A. Gates, I.D., 2017. Reactions of hydroxyl radicals with benzoic acid and benzoate. Royal Society of Chemistry Advances 7, 35776-35785.

Zetler, G., 1988. Neuroleptic-like, anticonvulsant and antinociceptive effects of aporphine alkaloids: bulbocapnine, corytuberine, boldine and glaucine. Arch Int Pharmacodyn Ther. 296, 255-281.

Zhang, X., Tervo, C.J., Reed, J.L., 2016. Metabolic assessment of E. coli as a Biofactory for commercial products. Metab. Eng. 35, 64-74. https://doi.org/ 10.1016/j.ymben.2016.01.007

Zornetzer, G.A., Tanem, J., Fox, B.G., Markley, J.L., 2010. The length of the bound fatty acid influences the dynamics of the acyl carrier protein and the stability of the thioester bond. Biochemistry 49, 470-477. https://doi.org/ 10.1021/bi9014659