Metabolite Damage and Metabolite Damage Control in Plants

Andrew D. Hanson,[1] Christopher S. Henry,[3,4] Oliver Fiehn,[5] and Valérie de Crécy-Lagard[2]

[1]Horticultural Sciences Department and [2]Microbiology and Cell Science Department, University of Florida, Gainesville, Florida 32611; email: adha@ufl.edu, vcrecy@ufl.edu

[3]Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois 60439; email: chenry@mcs.anl.gov

[4]Computation Institute, University of Chicago, Chicago, Illinois 60637

[5]Genome Center, University of California, Davis, California 95616; email: ofiehn@ucdavis.edu

Annu. Rev. Plant Biol. 2016. 67:131–52

First published online as a Review in Advance on November 19, 2015

The *Annual Review of Plant Biology* is online at plant.annualreviews.org

This article's doi: 10.1146/annurev-arplant-043015-111648

Keywords

cheminformatics, comparative genomics, damage preemption, damage repair, directed overflow, metabolomics

Abstract

It is increasingly clear that (*a*) many metabolites undergo spontaneous or enzyme-catalyzed side reactions in vivo, (*b*) the damaged metabolites formed by these reactions can be harmful, and (*c*) organisms have biochemical systems that limit the buildup of damaged metabolites. These damage-control systems either return a damaged molecule to its pristine state (metabolite repair) or convert harmful molecules to harmless ones (damage preemption). Because all organisms share a core set of metabolites that suffer the same chemical and enzymatic damage reactions, certain damage-control systems are widely conserved across the kingdoms of life. Relatively few damage reactions and damage-control systems are well known. Uncovering new damage reactions and identifying the corresponding damaged metabolites, damage-control genes, and enzymes demands a coordinated mix of chemistry, metabolomics, cheminformatics, biochemistry, and comparative genomics. This review illustrates the above points using examples from plants, which are at least as prone to metabolite damage as other organisms.

Contents

INTRODUCTION

Metabolite damage: the conversion of a normal metabolite to an abnormal one as the result of a chemical or enzymatic side reaction

Metabolite damage control: the restoration of damaged metabolites to their undamaged form (repair) or the conversion of potentially harmful compounds to harmless ones (preemption)

The terms "metabolite damage" and "metabolite damage control" may look unfamiliar, but they rest on several simple and uncontroversial ideas. First, chemically reactive metabolites—which is most of them—will undergo spontaneous chemical side reactions in vivo (32). Second, enzymes that catalyze side reactions in vitro—again, most of them—will do the same in vivo (13, 77). Third, damage products formed by metabolic side reactions will be useless, if not toxic, and hence unwanted (7). And last, cells need systems to stop such damage products from accumulating (7, 28). These ideas now seem self-evident, and classical papers (e.g., 50, 69, 81) show that the twentieth-century pioneers of metabolic biochemistry thought they were. However, metabolite damage and its control have been largely overlooked for the last 50 years.

This is all the more surprising because metabolite damage and its control closely parallel DNA and protein damage and repair (**Figure 1*a***), which are well researched (101, 103). Why, then, has metabolite damage (**Figure 1*b***) been so little studied? One reason may be the past prevalence of a paradigm that saw metabolic pathways as purpose driven and essentially error free (32). There being no place in this unfortunate paradigm for disorderly metabolite chemistry, imperfect enzymes, and their metabolic repercussions, it was easy to dismiss these topics as trivia. There were, and still are, also practical disincentives to performing metabolite damage research (64). First, damaged metabolites are often low abundance, labile, and commercially unavailable, making them easy to miss in analytical procedures and hard to identify if found. Nonavailability also discourages their use as biochemical substrates. Second, knocking out damage-control genes

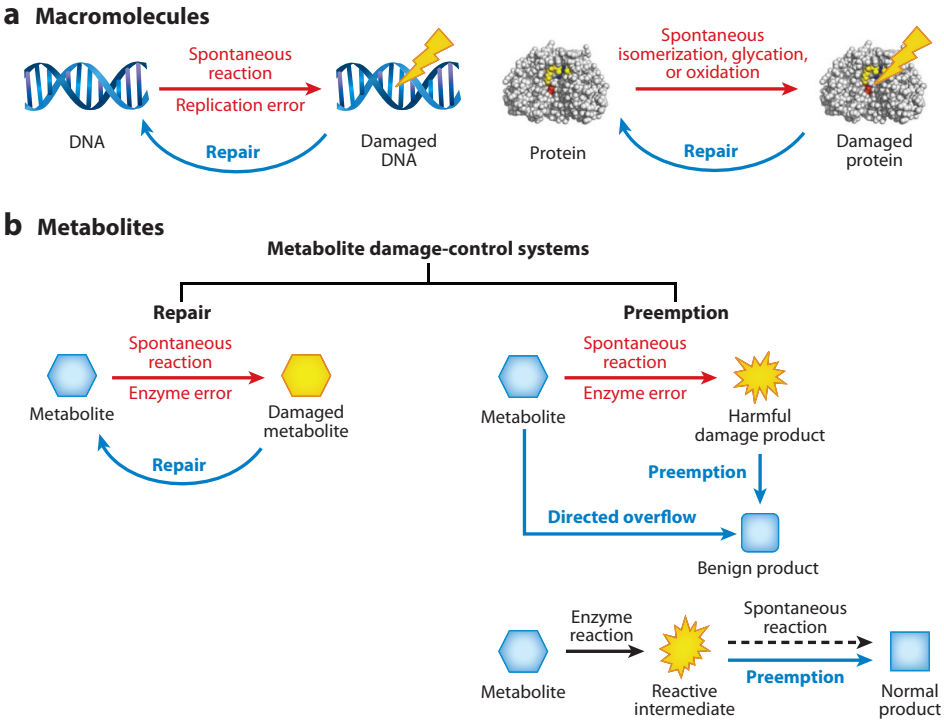


Figure 1

Damage-control systems for metabolites are analogous to those for macromolecules. The red arrows indicate damage reactions, and the blue arrows indicate various types of damage-control reactions; at the bottom, the solid black arrow indicates a normal enzyme reaction, and the dashed black arrow indicates a relatively slow spontaneous reaction. (*a*) It is well known that DNA and proteins are damaged by spontaneous reactions and enzymatic mistakes, and that enzymatic damage-control systems counter this damage. (*b*) The same applies to metabolites. Metabolite damage-control systems either repair the damage or preempt it. Directed overflow is a subcategory of damage preemption. Modified from Reference 64.

often has only subtle effects—i.e., they impact fitness, not survival (64). Subtle phenotypes let mutants slip through classical genetic screens and generally hinder genetic studies.

Whatever its causes, the neglect of metabolite damage is over. Interest is burgeoning in the chemical side reactions of metabolites (15, 52, 64) and even more in the side reactions mediated by metabolic enzymes (10, 54, 75) and in systems that limit metabolite damage (64, 101, 103). Fascinatingly, some damage-control systems have proven to be conserved across kingdoms of life, presumably because organisms share a core set of metabolites, and these suffer the same damage reactions always and everywhere (64). Work involving plants has been at the forefront of much of this progress.

Accordingly, here we first explain what chemical and enzymatic metabolite damage reactions are and how metabolite damage-control systems work, giving specific cases from plants. Next, we trace a connection between metabolite damage and the huge number of unidentified peaks in metabolomics profiles, and describe progress in systematizing information on damage reactions and in predicting damage products. Lastly, we suggest that hundreds of plant genes of unknown function may encode metabolite damage-control functions, and discuss how cross-kingdom comparative genomics can help predict these functions. Plants are good subjects for comparative

Comparative genomics: the comparison across organisms of genomic features such as gene fusions, gene clustering on a chromosome, and gene presence/absence patterns

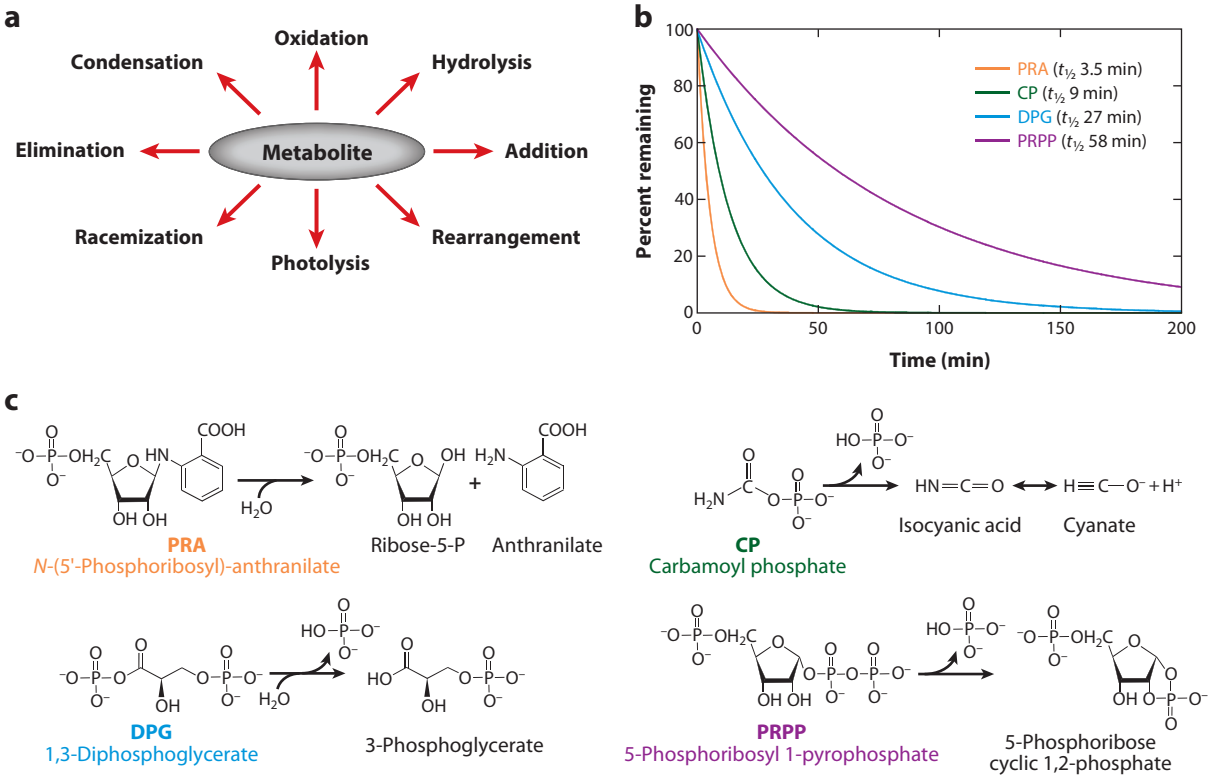


Figure 2

Chemical damage to metabolites. (*a*) Common spontaneous chemical reactions that metabolites may undergo, depending on their structure. (*b*) Illustrative first-order decay curves and half-lives ($t_{1/2}$) under physiological conditions for reactive intermediates in key metabolic pathways: the tryptophan synthesis intermediate *N*-(5′-phosphoribosyl)-anthranilate (PRA); the arginine precursor carbamoyl phosphate (CP); the glycolytic intermediate 1,3-diphosphoglycerate (DPG); and the purine, pyrimidine, and histidine precursor 5-phosphoribosyl 1-pyrophosphate (PRPP). (*c*) The chemical reactions responsible for the rapid decay of these four intermediates. The PRA and DPG reactions are hydrolyses, and the CP and PRPP reactions are eliminations.

genomics of metabolite damage because plastids as well as mitochondria are damage hot spots and have inherited damage-control systems from their prokaryotic progenitors.

CHEMICAL DAMAGE

Russian biochemist Alexey Golubev captured the essence of chemical damage to biomolecules in the six words "nothing of chemistry disappears in biology" (31, p. 6). Thus, metabolites that are chemically reactive and unstable in vitro remain so in vivo, and can actually become more unstable in vivo because they are exposed to other reactive molecules (3, 32). Many metabolic intermediates and end products thus constantly face spontaneous chemical damage (32, 52, 64). The damage processes are of three types: unimolecular reactions (racemization, rearrangement, elimination, and photolysis), reactions with water or oxidants, and reactions with other metabolites (addition and condensation) (**Figure 2*a***). Of these, only the reactions of certain metabolites with oxidants have been much studied in plants (21, 26).

It may come as a shock to realize just how chemically reactive a metabolite can be. **Figure 2*b*** gives four striking examples of highly unstable intermediates from central metabolic pathways (there are many more such cases). Under physiological conditions (neutral pH, ~37°C), these intermediates have half-lives of only 3.5–58 min (12, 25, 70, 105). Such half-lives become shorter still in leaves hitting temperatures of ~50°C (61) because chemical reaction rates roughly double per 10°C rise in temperature.

Although many different classes of metabolites are chemically reactive, some—the usual suspects—are more so than others and are hence particularly damage prone. These metabolites are also potentially damaging to other molecules. The examples shown in **Figure 2*b*** represent two such metabolite classes: acyl phosphates (carbamoyl phosphate and 1,3-diphosphoglycerate) and sugar phosphate derivatives, both of which can undergo elimination or hydrolysis (**Figure 2*c***). Note that the cyanate produced by breakdown of carbamoyl phosphate is an indiscriminate carbamoylating agent, so the initial chemical damage event leads to others. Such knock-on effects are common in metabolite damage chemistry. Other usual-suspect damage-prone metabolites include reactive carbonyls (e.g., methylglyoxal, triose phosphates, and fumarate) (93); imines (e.g., 2-aminoacrylate) (17); thiols (e.g., cysteine and coenzyme A) (109); and cofactors such as flavins, pyridine nucleotides, thiamin phosphates, pyridoxal and pyridoxamine phosphates, folates, and *S*-adenosylmethionine (64). That cofactors are seriously damage prone reflects that they are, by nature, molecules whose normal role is to react with other metabolites.

ENZYMATIC DAMAGE

It is now clear that most if not all enzymes make mistakes (i.e., that they are "promiscuous" or "sloppy") and that some make more mistakes than others (10, 13, 54). Promiscuity—the ability to catalyze slow, secondary reactions differing from the normal ones—is not the same as broad specificity; enzymes that are broadly specific (e.g., alkaline phosphatase) have the ability to catalyze the same reaction on various physiological substrates at similar rates (10, 102). The side activities of promiscuous enzymes are of two types. In the first type, an enzyme catalyzes its normal reaction on an abnormal (noncanonical) substrate. This is termed substrate promiscuity (right reaction, wrong substrate). In the second type, which is less common, an enzyme catalyzes an abnormal reaction on its normal substrate. This is termed catalytic promiscuity (right substrate, wrong reaction). Both types of side activity damage metabolites by converting them to aberrant products; recent examples of both types are given below.

The side reactions of promiscuous enzymes proceed at far lower rates (~10^2–10^7-fold slower) than their normal, physiological reactions (102). However, this slowness is deceptive because damage-product formation can be significant over time, especially if the enzyme is in a high-flux pathway or the noncanonical substrate is at high concentration. Two examples from mammals underscore these points. The glycolytic enzyme triose phosphate isomerase forms the toxic side product methylglyoxal at only 1/5,000 of the rate of the isomerase reaction, but daily methylglyoxal production in rats can nevertheless reach 0.4 μmol per gram of tissue (86). Likewise, mitochondrial malate dehydrogenase catalyzes α-ketoglutarate reduction to L-2-hydroxyglutarate ~10^7 times less efficiently than its normal reaction, oxaloacetate reduction. But because malate dehydrogenase is abundant and the mitochondrial α-ketoglutarate level is ~100 times that of oxaloacetate, the side activity can form several grams of L-2-hydroxyglutarate per day in a human adult (100).

Lastly, it should be noted that enzymatic and chemical damage reactions are not mutually exclusive—i.e., some enzymatic side reactions of metabolites can also occur spontaneously. The example of NAD(P)H hydrate formation covered below is one such case.

DAMAGE-CONTROL MECHANISMS

Damage-control systems can be classified in various ways and have been given various names. A classification based on that of Linster et al. (64) is shown in **Figure 1*b*** and outlined in the remainder of this section, along with some commonly used terms. Illustrative examples follow in the next section.

Repair

Repair deals with damage after it has happened. It simply returns a chemically or enzymatically damaged molecule to its original state via one or more enzymatic reactions (**Figure 1*b***). When the damage is enzymatic—i.e., an enzyme makes an unwanted side product—and the repair system reconverts this side product to the metabolite from which it came, the term metabolite proofreading can be used, by analogy to the proofreading activities of DNA polymerases and aminoacyl-tRNA synthetases (100).

Damage Preemption

Preemption stops damage from happening. It involves enzymes that either (*a*) convert a potentially harmful chemical or enzymatic side product to a benign one or (*b*) convert a normal but reactive and damaging intermediate to its normal product much faster than can occur spontaneously (**Figure 1*b***). The first case, which has been called "house cleaning" (28), is the best-known type of metabolite damage control. The second case—in which a preemption enzyme speeds up an insufficiently fast chemical reaction—is turning out to be crucial in the hydrolysis (quenching) of reactive enamine/imine intermediates (17, 71). The hydrolysis of these intermediates was formerly thought to be purely spontaneous because their half-life for chemical hydrolysis is 4 min or less (62). A lesson here is that reactions that proceed quite fast chemically may still require acceleration by a damage preemption enzyme.

Directed Overflow

Directed overflow also stops damage from happening. It is a form of preemption in which an excess of a normal metabolite that can potentially give rise to harmful products is enzymatically diverted to one or more benign products before it does so (27, 85) (**Figure 1*b***). The benign product can be as simple as the dephosphorylated version of a phosphometabolite. Directed overflow heads off damage even earlier in the event chain than other preemptive mechanisms do, and may be concentrated in pathways that lack feedback regulation mechanisms to prevent the buildup of intermediates (85).

EXEMPLARS FROM PLANTS

This section describes four iconic cases of metabolite damage and its repair or preemption, all drawn from plants. These examples are not confined to plants, however; the fact that they are also present in other organisms was crucial to their discovery, which in each case involved cross-kingdom comparative genomics. The comparative genomics trails that led to these discoveries are retraced below (see the section Metabolite Damage and Genomics: The Unknown Gene Problem). **Table 1** summarizes 14 more cases of metabolite damage and damage control in plants, most of which also occur in other kingdoms of life. It is worth noting that all four cases described below were recognized in the past two years, and that 10 of the cases in **Table 1** are similarly recent. This emphasizes the accelerating tempo of work on metabolite damage.

Table 1 Additional cases of metabolite damage and damage-control enzymes in plants

Metabolite	Damage reaction	Damage-control enzymes	Reference(s)
Repair			
S-Adenosylmethionine (AdoMet)	Natural (*S*,*S*)-AdoMet is spontaneously racemized to the inactive diastereomer (*R*,*S*)-AdoMet.	(*R*,*S*)-AdoMet-selective homocysteine methyltransferases yield methionine plus *S*-adenosylhomocysteine, which is reconverted to (*S*,*S*)-AdoMet via the methyl cycle.	6, 104
Methionine	Free methionine is chemically oxidized to (*R*)- and (*S*)-methionine sulfoxides.	The methionine sulfoxide reductases MSRA and MSRB return the *R* and *S* forms, respectively, to methionine.	63
α-Amino adipic semialdehyde (α-AAS)	The lysine catabolite α-AAS spontaneously and reversibly cyclizes to Δ^1-piperideine-6-carboxylate, which is adventitiously reduced to the dead-end product pipecolate.	Pipecolate oxidase reconverts pipecolate to Δ^1-piperideine-6-carboxylate, which decyclizes to α-AAS and then reenters the lysine catabolic pathway.	35, 100
α-Tocopherol	α-Tocopherol is chemically oxidized to α-tocopherol quinone.	Quinone oxidoreductase (NDC1), tocopherol cyclase (VTE1), and unknown dehydratases reconvert the quinone to α-tocopherol.	19
5,10-Methenyltetrahydrofolate (5,10-CH = THF)	5,10-CH = THF is hydrolyzed to the dead-end, inhibitory folate 5-formyltetrahydrofolate by a side reaction of serine hydroxymethyltransferase, or spontaneously at low pH.	5-Formyltetrahydrofolate cycloligase reconverts 5-formyltetrahydrofolate to 5,10-CH = THF in an irreversible, ATP-dependent reaction.	33, 88
2-Oxoglutarate	A side reaction of mitochondrial malate dehydrogenases reduces 2-oxoglutarate to the dead-end metabolite L-2-hydroxyglutarate.	A flavin adenine dinucleotide (FAD)–containing mitochondrial L-2-hydroxyglutarate dehydrogenase converts L-2-hydroxyglutarate back to 2-oxoglutarate.	46
Isopentenyl and dimethylallyl diphosphates (IPP and DMAPP)	Phosphatase action converts IPP and DMAPP to the corresponding monophosphates, which inhibit farnesyl diphosphate synthase.	An isopentenyl phosphate kinase reconverts the monophosphates to IPP and DMAPP.	42
Preemption			
Thiamin diphosphate (ThDP)	ThDP is oxidized spontaneously or by a side reaction of ThDP enzymes to oxo-ThDP, or spontaneously hydrolyzed to oxy-ThDP; oxy- and oxo-ThDP inhibit ThDP enzymes.	A Nudix diphosphatase selectively hydrolyzes oxo- and oxy-ThDP to their monophosphates, which are harmless to ThDP enzymes.	34

(*Continued*)

Table 1 (*Continued*)

Metabolite	Damage reaction	Damage-control enzymes	Reference(s)
Guanidine	Guanidine is chemically oxidized to 8-oxo-guanidine, which is erroneously converted to 8-oxo-dGTP and 8-oxo-GTP; these lead to replication and transcription errors.	A Nudix diphosphatase hydrolyzes 8-oxo-dGTP and 8-oxo-GTP to the corresponding monophosphates, preventing their incorporation into DNA and RNA.	78, 114
Homocysteine thiolactone (HCTL)	Error editing of mischarged homocysteine by certain aminoacyl-tRNA synthetases forms HCTL, which damages proteins by forming adducts with amino groups.	A selective thiolactonase hydrolyzes HCTL to homocysteine, preventing protein damage.	47
Methylglyoxal	Triose phosphates are converted to the reactive carbonyl compound methylglyoxal by a side reaction of triose phosphate isomerase and also spontaneously.	The tandem action of glyoxalase I, glyoxalase II, and D-lactate dehydrogenase disposes of methylglyoxal by converting it to pyruvate, which reenters central carbon metabolism.	18, 94, 113
Ribulose 1,5-bisphosphate (RuBP)	A Rubisco side reaction (misfire) converts RuBP to pentodiulose 1,5-bisphosphate, a tight-binding inhibitor of Rubisco.	A selective phosphatase (CA1Pase) dephosphorylates pentodiulose 1,5-bisphosphate (and related Rubisco misfire products), maintaining Rubisco activity.	2, 82
Directed overflow			
Flavin mononucleotide (FMN)	FMN overproduction from riboflavin by riboflavin kinase can potentially occur owing to a lack of feedback regulation of riboflavin synthesis. FMN homeostasis appears to be crucial in plants.	A haloacid dehalogenase (HAD) family hydrolase domain fused to riboflavin kinase dephosphorylates FMN to riboflavin, thus limiting FMN buildup.	68, 90
Coenzyme A (CoA) precursors	The CoA precursors phosphopantothenate and phosphopantetheine (which forms oxidation products) may accumulate; the oxidation products are converted to inactive CoA forms.	A DUF89 phosphatase domain fused to pantothenate kinase dephosphorylates phosphopantothenate and phosphopantetheine and its oxidation products, thereby maintaining the integrity of the CoA pool.	45, 108

An Epimerase-Dehydratase Duo Repairs NAD(P)H Hydrates

The damage reaction in this case is both enzymatic and chemical. NADH and NADPH are converted to hydrates (NADHX and NADPHX, respectively) either via a side reaction of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) or spontaneously (1). NADHX and NADPHX exist as a mixture of C-6 *R* and *S* epimers (**Figure 3*a***), and because they inhibit various dehydrogenases, they cannot be left to accumulate in cells (67).

The repair of NAD(P)HX is a two-part process. An ATP-dependent NAD(P)HX dehydratase reconverts the *S* epimer of the hydrates to NAD(P)H, and an NAD(P)HX epimerase facilitates interconversion of the *R* and *S* epimers (9, 67, 73) (**Figure 3*a***). Knocking out the dehydratase in

Arabidopsis leads to massive (up to 40-fold) accumulation of the hydrates (73). Both the dehydratase and the epimerase are targeted in plants to multiple subcellular compartments via use of alternative translation start sites (9, 73). This multiple targeting of the repair enzymes presumably reflects the cell-wide nature of NAD(P)HX formation and the need to counter it wherever it occurs.

A Selective Bisphosphatase Preempts Blockage of Rubisco Activity

The potential damage source here is a minor side reaction (misfire reaction) catalyzed by Rubisco (**Figure 3*b***). This reaction epimerizes the Rubisco substrate ribulose 1,5-bisphosphate at the 3 position to give xylulose 1,5-bisphosphate, which is a potent Rubisco inhibitor and must consequently be disposed of (5, 82). Potential harm to Rubisco is preempted by a selective sugar phosphatase that hydrolyzes xylulose 1,5-bisphosphate to xylulose 5-phosphate, a normal Calvin-Benson cycle intermediate (5) (**Figure 3*b***).

An *N*-Glycosidase Safely Directs Excess Intermediates Out of the Riboflavin Pathway

In this case, the potential sources of damage are the first two intermediates of the riboflavin synthesis pathway (**Figure 3*c***), which may accumulate because riboflavin synthesis is not feedback regulated (27). Both intermediates are highly reactive and break down spontaneously and rapidly to create harmful compounds that likely include Maillard products and 5-phosphoribosylamine (24, 27, 66).

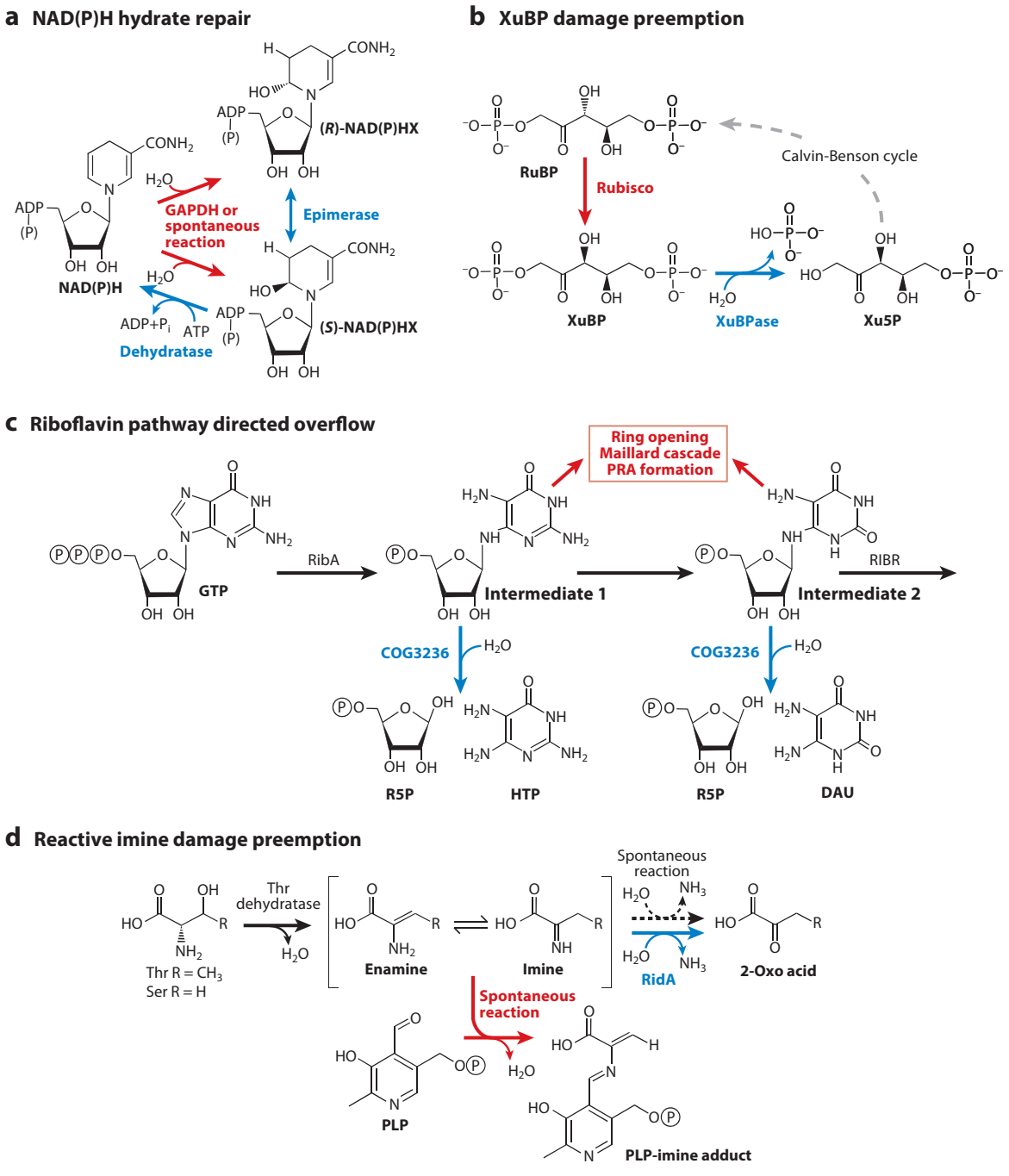
The damage that these products could otherwise inflict is preempted by an *N*-glycosidase (COG3236) that hydrolyzes excess intermediates 1 and 2 to ribose 5-phosphate and pyrimidine moieties, which are relatively harmless (27). The *N*-glycosidase thus mediates a directed overflow type of damage preemption—i.e., it safely diverts a dangerous excess of intermediates out of the pathway.

Hydrolases Speed Up Conversion of Reactive Imines to 2-Oxo Acids

In this instance, the potential damage agents are reactive enamine/imine tautomers (**Figure 3*d***). These are the normal initial products of the pyridoxal 5′-phosphate (PLP) enzyme threonine dehydratase acting on threonine or serine, of other PLP enzymes, and of amino acid oxidase (17, 71); they were formerly thought to hydrolyze spontaneously to the final 2-oxo acid products (62). The enamines/imines, particularly 2-aminoacrylate, can inactivate PLP enzymes via adduct formation with the PLP cofactor (17, 72). Although enamines/imines spontaneously hydrolyze very quickly, this reaction (half-time of 4 min or less) is insufficient to avoid damage to the aminotransferase required for branched-chain amino acid synthesis and other PLP enzymes (17, 72). RidA enzymes preempt damage from enamines/imines by facilitating their hydrolysis to 2-oxo acids (**Figure 3*d***). RidA enzymes can also preempt damage from the extremely reactive compound carbamoyl phosphate (**Figure 2*b*,*c***), but the biochemical mechanism has yet to be determined (71).

METABOLITE DAMAGE, CHEMICAL SPACE, AND METABOLOMICS

The concept of chemical space—the ensemble of all possible molecules—is very important for thinking about metabolite damage and the analysis of damage products. The first point to grasp is that chemical space is vast, and metabolic pathways occupy very little of it. For instance, current chemical databases comprise more than 40 million structures (53), but there are only 15,000


a NAD(P)H hydrate repair
(R)-NAD(P)HX
GAPDH or spontaneous reaction
Epimerase
NAD(P)H
(S)-NAD(P)HX
ADP+Pi
ATP
Dehydratase
b XuBP damage preemption
RuBP
Rubisco
Calvin-Benson cycle
XuBP
XuBPase
Xu5P
c Riboflavin pathway directed overflow
Ring opening
Maillard cascade
PRA formation
GTP
RibA
Intermediate 1
Intermediate 2
RIBR
COG3236
R5P
HTP
DAU
d Reactive imine damage preemption
Thr dehydratase
Thr R = CH3
Ser R = H
Enamine
Imine
Spontaneous reaction
RidA
2-Oxo acid
PLP
PLP-imine adduct

metabolic substrates and products in the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (51). Metabolic pathways are thus narrow, well-lit tunnels of known chemistry boring through a huge, dark matrix of unknown chemistry (**Figure 4**). Damage reactions are excursions from the light of the tunnels out into the adjacent darkness. The damage reactions of central carbon and energy pathways, which are far better characterized than most (but nowhere near completely), illustrate how much damage traffic may cross the tunnel walls of metabolism as a whole (**Figure 4**). One way to identify damaged metabolites should therefore be to search the chemical space outside pathways using metabolomics, looking for unknown peaks and then identifying them. This is much easier said than done (11).

Cheminformatics: computational approaches to chemical problems, such as using a molecule's structure to predict its physicochemical properties and potential reactions

Metabolomics Peaks May Well Be Damage Products but Are Hard to Identify

Plant metabolomics profiles from accurate mass liquid chromatography–mass spectrometry (LC-MS) instruments can include thousands of unknown (i.e., unannotated) signals (30), and many of these could be damage products (23, 64). But it is extremely hard to establish the identities of the unknown molecules that give rise to these signals. Simply matching their observed masses against the predicted masses of the more than 40 million chemical database structures is of little use because, even when the experimental data are sufficient to define a unique elemental composition (which they often are not), many molecules within chemical space have the same elemental composition and hence the same mass (55). Nor is mass fragment spectral matching a catch-all solution, because (*a*) it relies on comparisons to spectra of authentic standards and (*b*) the major mass spectral libraries (ReSpect, MassBank, Metlin, and NIST14) collectively contain only 40,000 experimentally determined electrospray tandem mass spectra (56, 59)—and these are obviously of known compounds, not unknown damage products. Fortunately, cheminformatics can help.

Cheminformatics Can Predict Metabolite Damage Products

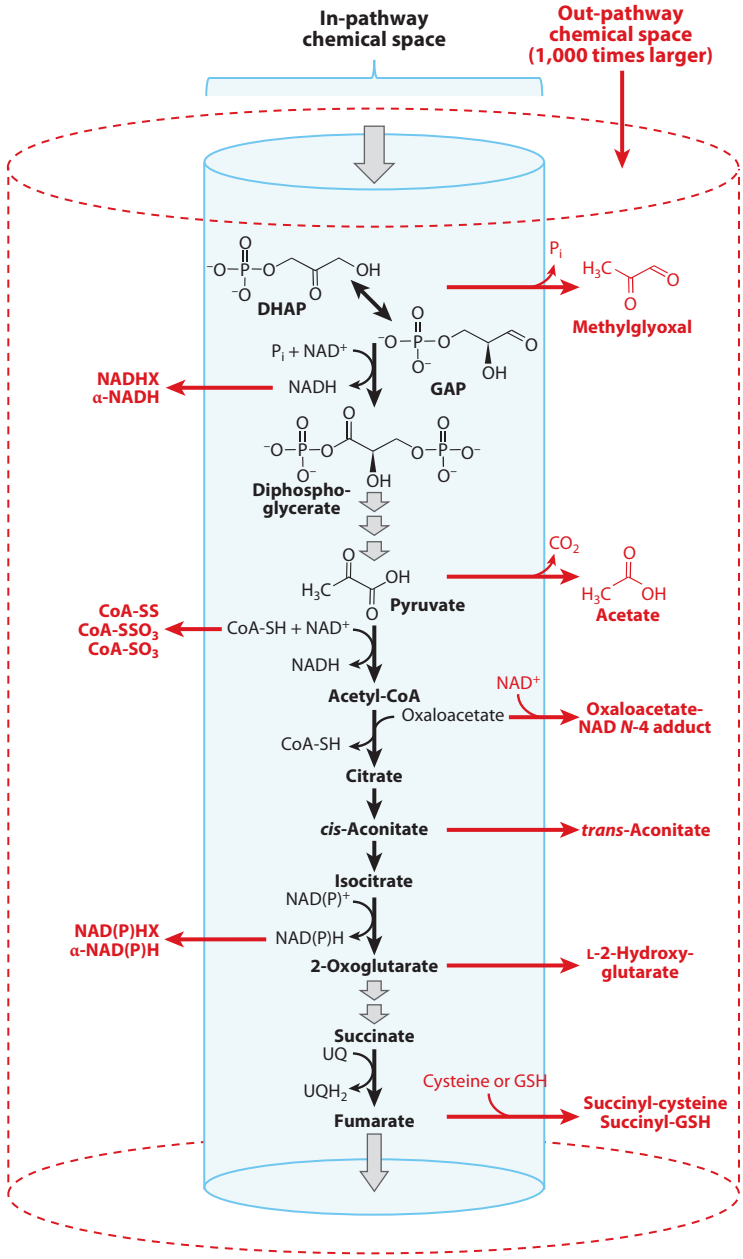
Cheminformatic approaches can help identify metabolite damage products in two basic ways: (*a*) by predicting the mass spectra and retention times of hypothetical database structures that can be matched with those of unknown metabolomics peaks, and (*b*) by predicting potential damage products of known metabolites using (bio)chemical principles (i.e., by exploring chemical space).

For the first approach, heuristic rules for bond cleavages in specific compound classes can be derived from experimental spectra (56, 57, 87, 107, 111), and retention times can be modeled for

Figure 3

Examples of metabolite damage and its repair or preemption. Damage reactions are shown in red; repair or preemption reactions are shown in blue. (*a*) NADH and NADPH are converted to hydrates (NADHX and NADPHX, respectively) either via a side reaction of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) or spontaneously. An ATP-dependent NAD(P)HX dehydratase returns the *S* form of the hydrates to NAD(P)H, and an NAD(P)HX epimerase interconverts the *R* and *S* forms. (*b*) Ribulose 1,5-bisphosphate (RuBP) is epimerized at the 3 position by a side reaction of Rubisco, yielding xylulose 1,5-bisphosphate (XuBP), a Rubisco inhibitor. A selective phosphatase (XuBPase) hydrolyzes XuBP to xylulose 5-phosphate (Xu5P), a normal Calvin-Benson cycle intermediate. The dashed gray arrow indicates multiple steps in the Calvin-Benson cycle. (*c*) Intermediates 1 and 2 of riboflavin synthesis are reactive and spontaneously form Maillard products, 5-phosphoribosylamine (PRA), and other harmful compounds. An *N*-glycosidase (COG3236) hydrolyzes excess intermediates 1 and 2 to ribose 5-phosphate (R5P) and a pyrimidine (HTP or DAU), which are fairly harmless. (*d*) Threonine dehydratase converts threonine or serine to highly reactive enamine/imine tautomers. The serine product (2-aminoacrylate) can inactivate pyridoxal 5′-phosphate (PLP) enzymes via adduct formation with the PLP cofactor. RidA enzymes preempt such damage by hydrolyzing the imines to 2-oxo acids; the hydrolysis also occurs spontaneously (*dashed black arrow*), but too slowly to prevent enamine/imine damage to enzymes.

a specific chromatographic method via quantitative structure-property relationships (98). However, the great diversity of damaged metabolites may challenge these structural cheminformatic approaches, making it necessary to use more powerful strategies such as accurate calculation of bond energies and the likelihood that bond dissociations will actually occur using first-principle or quantum chemistry approaches. At least for electron ionization mass spectra, this is now feasible (36). Such first-principle computational advances now need to be extended to electrospray collision-induced fragmentation, as used in LC-MS-based metabolomics.



In the second approach, BNICE (Biochemical Network Integrated Computational Explorer) (38) or other cheminformatics tools (44, 58, 79) enable high-throughput generation of hypothetical metabolite damage chemistry. The first step is to encode generic reaction rules that capture (bio)chemical damage mechanisms (e.g., decarboxylation or nonenzymatic carbamoylation) using databases of example reactions (79) or expert chemical knowledge (38, 44, 58). These rules are then applied to databases of known metabolites, e.g., PlantCyc (115) or AraCyc (116), to propose all possible damage reactions and products, as is done in the MINE (Metabolic In Silico Network Expansion) databases for enzymatic reactions (49) and the Chemical Damage MINE for spontaneous reactions (C. Lerma-Ortiz, J.G. Jeffryes, A.J.L. Cooper, T.D. Niehaus, A.M. Thamm, et al., manuscript in review). The rules can also be applied in a targeted way to individual metabolites to assess which damage reactions they are most likely to undergo (27).

Damage predictions can be sharpened by integrating predicted metabolite and reaction properties into genome-scale metabolic models. For example, group contribution methods to estimate the standard Gibbs free energy change of reactions (48) can identify damage reactions that are far from thermodynamic equilibrium in cells and thus likely to go toward completion (41, 60), and thermodynamics-based metabolic flux analysis can predict flux through reactions and hence possible concentrations of damage products (40). More generally, in conjunction with genome-scale metabolic models (22, 92), thermodynamics-based metabolic flux analysis and other flux balance analysis methods can identify indispensable metabolites (C. Lerma-Ortiz, J.G. Jeffryes, A.J.L. Cooper, T.D. Niehaus, A.M. Thamm, et al., manuscript in review). The damage reactions of such pivotal metabolites are particularly high-priority targets for research.

METABOLITE DAMAGE AND GENOMICS: THE UNKNOWN GENE PROBLEM

Many Genes of Unknown Function Likely Encode Metabolite Damage-Control Enzymes

Given that a multitude of metabolites can undergo chemical or enzymatic damage, and that the search for damage-control enzymes has only just begun, it is likely that many such enzymes—and the genes that specify them—remain to be discovered. Here, we estimate how many of these genes are in the *Arabidopsis* genome based on the number of genes of unknown function, the percentage of these genes that code for enzymes, and the percentage of these that perform damage control.

Although the genomes of plants and other organisms clearly encode many enzymes of unknown function (37, 43), establishing the exact percentage of unknown genes in a genome is problematic

Figure 4

The chemical space occupied by central carbon and energy pathways (in-pathway chemical space) and part of the vast chemical space surrounding these pathways (out-pathway chemical space) that metabolite damage reactions begin to penetrate. Note that the total out-pathway space is at least 1,000 times the size of the in-pathway space. The pathway intermediates and cofactors are shown in black within the blue cylinder of the in-pathway space; chemical or enzymatic damage reactions and the resulting damage products are in red in the out-pathway space. Abbreviations: CoA-SH, coenzyme A; CoA-SO_3, CoA sulfonate; CoA-SS, CoA disulfide; CoA-SSO_3, CoA *S*-sulfonate; DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde 3-phosphate; GSH, glutathione; α-NAD(P)H, α anomer of NAD(P)H; NAD(P)HX, NAD(P)H hydrate; UQ, ubiquinone. The chemical and enzymatic damage reactions come from References 20, 45, 46, 86, 96, and 117 and from C. Lerma-Ortiz, J.G. Jeffryes, A.J.L. Cooper, T.D. Niehaus, A.M. Thamm, et al., manuscript in review.

because the definitions of known and unknown functions vary. In early genome annotations, functions as vague as "ATPase" or "transporter" were counted as known, whereas best current practice as implemented by UniProt (Universal Protein Resource) (99) is more stringent and uses various types of evidence (sequence similarity or experimental data) (83, 84). However, some of the functions called by UniProt would still be classified as unknown by the most stringent definitions, in which gene function must be experimentally validated in vitro or in vivo. The problem of function calling is compounded by the overannotation of paralogous families in which a gene that has been erroneously given a specific function is taken as known when in fact it is not (91). Estimates of unknown gene percentages are thus necessarily rough.

With this disclaimer in place, to estimate how many enzymes of unknown function are present in *Arabidopsis* and how many of these could be involved in metabolite damage control, we analyzed the genomes of both *Arabidopsis* and *Escherichia coli*. By various estimates, the *Arabidopsis* genome encodes ~10,000 enzymes (74) (**Supplemental Table 1**; follow the **Supplemental Materials link** from the Annual Reviews home page at **http://www.annualreviews.org**). To estimate how many of these are of unknown function, we combined different UniProt filters (**Supplemental Table 1**). This gave numbers between 1,300 and 8,600, depending on the strictness of the criteria used to define a function as known. Note that the lower bound (1,300) is certainly an underestimate because it unreasonably assumes that Gene Ontology molecular function assignments always correspond to biological function unless the enzyme is labeled as uncharacterized.

Supplemental Material

In 1998, *E. coli* genes whose functions were at that time unknown were given "Y" names, e.g., *ygfZ* (89). In May 2015, we extracted 1,441 Y genes from EcoGene (118), of which 1,324 remained after eliminating pseudogenes. By combining UniProt and SEED (80) database annotations with manual analysis, we found that 231 of these Y genes have had solid functions assigned to them since 1998. Of these 231, manual analysis indicated that 93 (or 40%) encode enzymes, which is a little above the automated prediction of ~30% enzymes in the current protein space (106). Of the 93, 14 were damage-control enzymes, all from conserved families (**Supplemental Table 2**). Thus, 15% of the previously unknown enzymes in *E. coli* whose functions have recently been elucidated have roles in metabolite damage control.

Projecting this 15% value from *E. coli* onto the estimated 1,300–8,600 unknown-function enzymes in *Arabidopsis* implies that there are between 200 and 1,300 undiscovered metabolite damage-control enzymes in this and other plants. Even taking the conservative lower bound of ~200 still gives a sense of how much damage-control machinery is still "missing" and, consequently, why a focused effort to identify metabolite damage-control genes is needed. Fortunately, as discussed next, metabolite damage-control enzymes are often conserved between plants and other organisms, so comparative genomics provides a powerful way to predict their functions.

Discovering Damage-Control Functions via Comparative Genomics

Comparative genomics, particularly between bacteria and plants, has been a royal road to discovery of metabolite damage-control genes and enzymes and is likely to remain so in future (64). Below, we illustrate how this discovery process works by revisiting the four exemplars of metabolite damage and its repair or preemption described above.

NAD(P)HX epimerase. This case demonstrates the value of protein-protein fusions. The partners in such fusions—which have aptly been called "Rosetta stones" (95)—nearly always have related functions. After the yeast NAD(P)HX dehydratase gene was found via classical biochemistry and cloning (67), its homologs in bacteria and mammals (67) and then plants (9, 73) were readily identified. In *E. coli* and most other bacteria, the dehydratase gene proved to be fused to an

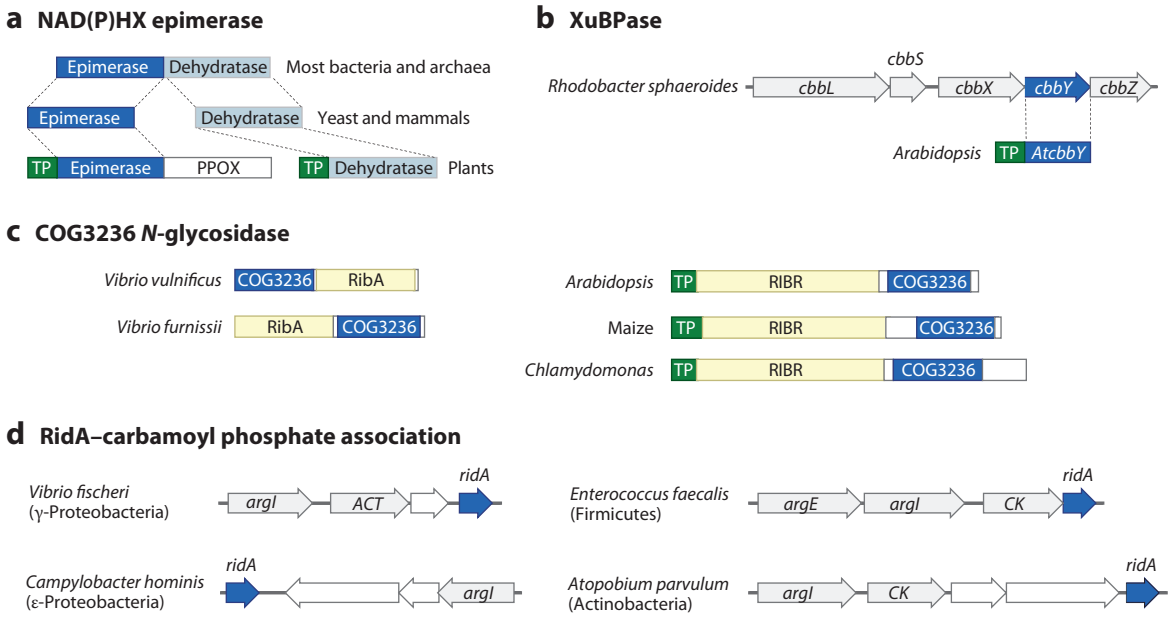


Figure 5

Examples of comparative genomic evidence predicting damage-control functions. (*a*) The yeast NAD(P)HX dehydratase gene was cloned via classical biochemistry and found to have homologs in bacteria, mammals, and plants. In most bacteria, the dehydratase gene is fused to an N-terminal domain of previously unknown function. This domain was predicted and then shown to be NAD(P)HX epimerase. Epimerase homologs were then identified in eukaryotes, including plants. (*b*) The *cbb* operon of the photosynthetic bacterium *Rhodobacter sphaeroides* and similar operons in other bacteria encode Rubisco subunits (*cbbL* and *cbbS*), Rubisco activase (*cbbX*), phosphoglycolate phosphatase (*cbbZ*), and a haloacid dehalogenase (HAD) family phosphatase (*cbbY*) of previously unknown function. CbbY and its *Arabidopsis* homolog (AtCbbY) were predicted and shown to have a function related to the Calvin-Benson cycle (hydrolysis of the Rubisco misfire product xylulose 1,5-bisphosphate). (*c*) The first enzyme of riboflavin synthesis (RibA) in *Vibrio* species and the third enzyme (RIBR) in plants are fused to a domain of previously unknown function (COG3236). This domain was predicted and shown to have a damage-control role (hydrolysis of excess riboflavin intermediates). (*d*) In taxonomically diverse bacteria, genes encoding RidA family proteins cluster with genes of carbamoyl phosphate metabolism, including those encoding aspartate carbamoyltransferase (*ACT*), ornithine carbamoyltransferase (*argI*), acetylornithine deacetylase (*argE*), and carbamate kinase (*CK*). This clustering predicted a damage-control role for RidA in carbamoyl phosphate metabolism, which was subsequently confirmed by genetic evidence. Additional abbreviation: TP, targeting peptide.

N-terminal domain of unknown function (**Figure 5*a***). Because the dehydratase was known to act on the *S* epimer of NAD(P)HX but not on the *R* epimer, the N-terminal domain was predicted to be an NAD(P)HX epimerase (67). Epimerase homologs were then identified in yeast, mammals, and plants (9, 67, 73).

Xylulose 1,5-bisphosphate phosphatase. This case shows how the chromosomal context of a bacterial gene can suggest its function and that of its plant homolog. Prokaryotic genes that consistently cluster together on the chromosome are likely to have related functions (37). The *cbb* operon of the photosynthetic bacterium *Rhodobacter sphaeroides* and similar operons in other bacteria were shown to encode Rubisco subunits, Rubisco activase, phosphoglycolate phosphatase, and a haloacid dehalogenase (HAD) family phosphatase (CbbY) of unknown function (29) (**Figure 5*b***). The clustering of CbbY with genes related to the Calvin-Benson cycle and the chloroplastic location of the *Arabidopsis* CbbY homolog suggested a role in the Calvin-Benson cycle, and membership in the HAD family indicated that this role would involve dephosphorylation (5). CbbY

was subsequently shown to hydrolyze the Rubisco misfire product xylulose 1,5-bisphosphate to xylulose 5-phosphate (5).

COG3236 *N*-glycosidase. This case reemphasizes how informative protein fusions can be. The first enzyme of riboflavin synthesis (RibA) in *Vibrio* species and related bacteria was found to be fused (either N or C terminally) to a domain of unknown function, COG3236 (27) (**Figure 5*c***). Remarkably, a COG3236 protein was also found to be fused to a different enzyme of riboflavin synthesis (RIBR) in green plants from algae to angiosperms (**Figure 5*c***). These fusions from different kingdoms of life pointed strongly to a function related to the riboflavin pathway, and both cheminformatic evidence (27) and distant homology to proteins predicted to depurinate ADP-ribose (16) suggested that COG3236 is an *N*-glycosidase. COG3236 proteins were subsequently shown to hydrolyze the *N*-glycosidic bond in the first two intermediates of riboflavin synthesis (27).

RidA and carbamoyl phosphate. This is a second case in which gene clustering was crucial to a functional prediction. It was first observed that, besides clustering with PLP-related genes, genes encoding RidA family proteins also clustered in taxonomically diverse bacteria with pyrimidine and arginine metabolism genes, particularly those related to carbamoyl phosphate, notably aspartate and ornithine carbamoyltransferases, acetylornithine deacetylase, and carbamate kinase (71) (**Figure 5*d***). Given the extreme instability and reactivity of carbamoyl phosphate and its breakdown product cyanate (105), this clustering pattern predicted a damage-control role for RidA in carbamoyl phosphate metabolism. Genetic support for such a role was subsequently obtained (71).

THE ENVIRONMENTAL AND GENOMIC CONTEXT

Two final key points are that plants are at least as prone to metabolite damage as other organisms, and that metabolite damage can potentially reach up into the genome and epigenome.

Plant proneness to metabolite damage stems from their exposure to solar radiation, to sharp and sometimes large temperature changes, and to varying water content, all of which can promote both chemical and enzymatic side reactions. Thus, the temperature, concentration, pH, and ionic-strength dependences of chemical and enzymatic reactions tend to drive up damage reaction rates during heat stress and water deficit. Furthermore, because enzyme specificity can also depend on physicochemical parameters such as temperature and type of metal cofactor (8, 65, 112) and metabolite pool sizes change drastically in stressed plants (97, 110), enzyme error rates can rise as enzymes become sloppier and encounter higher concentrations of noncanonical substrates in stressed cells. The homeothermic and homeohydric lifestyle of mammals largely insulates them from this type of problem.

A few connections between metabolite damage and the genome or epigenome have been evident for some time, but other, less obvious ones are now emerging. A classical connection is oxidative damage to nucleotides, giving rise to, e.g., 8-oxo-dGTP, which causes mispairing mutations when incorporated into DNA. This damage to the genome is preempted by the Nudix enzyme MutT, a nucleoside-triphosphate pyrophosphohydrolase that is specific for 8-oxo-dGTP and thus "sanitizes" the nucleotide pool (4, 114). Two new examples concern the respiratory metabolite fumarate and its spontaneous adduct with glutathione. Fumarate itself is an analog of 2-oxoglutarate and can competitively inhibit 2-oxoglutarate-dependent enzymes, including those that demethylate DNA and histones and hence regulate the epigenome (76). Formation of the fumarate-glutathione adduct can deplete the glutathione pool, triggering oxidative stress and consequent DNA damage (39, 96, 117). Such small-molecule damage knock-on to the DNA

level has prompted speculation that metabolism is far more important in evolution (natural and engineered) than traditional thinking suggests (14). If DNA is only stable when metabolite damage is contained, then metabolite damage-control systems would rank alongside the classical DNA damage-repair systems in significance to the genome.

SUMMARY POINTS

1. Many metabolites are constantly subject to damage in vivo by spontaneous chemical reactions and by side reactions of promiscuous enzymes.
2. Chemical and enzymatic metabolite damage products are wasteful or harmful and are dealt with by biochemical systems that repair or preempt damage; these systems substantially affect fitness and are often widely conserved.
3. Metabolite damage reactions and damage-control systems have long been overlooked but are now being uncovered at an increasing rate, not least in plants.
4. Many of the thousands of unidentified peaks in plant metabolomic profiles are likely to be damaged metabolites.
5. It is likely that hundreds of genes of unknown function in plant genomes encode enzymes involved in metabolite damage-control systems.
6. Understanding metabolite damage depends on the systematic integration of insights from chemistry, metabolomics, cheminformatics, biochemistry, and comparative genomics.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors' research on metabolite damage and damage control is supported by US National Science Foundation grants MCB-1153413 (to A.D.H.), MCB-1153357 (to C.S.H.), and MCB-1153491 (to O.F.).

LITERATURE CITED

1. Acheson SA, Kirkman HN, Wolfenden R. 1988. Equilibrium of 5,6-hydration of NADH and mechanism of ATP-dependent dehydration. *Biochemistry* 27:7371–75
2. Andralojc PJ, Madgwick PJ, Tao Y, Keys A, Ward JL, et al. 2012. 2-Carboxy-D-arabinitol 1-phosphate (CA1P) phosphatase: evidence for a wider role in plant Rubisco regulation. *Biochem. J.* 442:733–42
3. Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD. 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* 5:593–99
4. Bessman MJ, Frick DN, O'Handley SF. 1996. The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. *J. Biol. Chem.* 271:25059–62
5. Bracher A, Sharma A, Starling-Windhof A, Hartl FU, Hayer-Hartl M. 2015. Degradation of potent Rubisco inhibitor by selective sugar phosphatase. *Nat. Plants* 1:14002
6. Bradbury LM, Ziemak MJ, El Badawi-Sidhu M, Fiehn O, Hanson AD. 2014. Plant-driven repurposing of the ancient *S*-adenosylmethionine repair enzyme homocysteine *S*-methyltransferase. *Biochem. J.* 463:279–86

7. Chan CM, Danchin A, Marlière P, Sekowska A. 2014. Paralogous metabolism: S-alkyl-cysteine degradation in *Bacillus subtilis*. *Environ. Microbiol.* 16:101–17
8. Chesters C, Wilding W, Goodall M, Micklefield J. 2012. Thermal bifunctionality of bacterial phenylalanine aminomutase and ammonia lyase enzymes. *Angew. Chem.* 124:4420–24
9. Colinas M, Shaw HV, Loubéry S, Kaufmann M, Moulin M, Fitzpatrick TB. 2014. A pathway for repair of NAD(P)H in plants. *J. Biol. Chem.* 289:14692–706
10. Copley SD. 2015. An evolutionary biochemist's perspective on promiscuity. *Trends Biochem. Sci.* 40:72–78
11. Creek DJ, Dunn WB, Fiehn O, Griffin JL, Hall RD, et al. 2014. Metabolite identification: Are you sure? And how do your peers gauge your confidence? *Metabolomics* 10:350–53
12. Creighton TE. 1968. The nonenzymatic preparation in solution of *N*-(5′-phosphoribosyl) anthranilic acid, an intermediate in tryptophan biosynthesis. *J. Biol. Chem.* 243:5605–9
13. D'Ari R, Casadesús J. 1998. Underground metabolism. *BioEssays* 20:181–86
14. **de Lorenzo V. 2014. From the selfish gene to selfish metabolism: revisiting the central dogma. *BioEssays* 36:226–35**
15. de Lorenzo V, Sekowska A, Danchin A. 2015. Chemical reactivity drives spatiotemporal organisation of bacterial metabolism. *FEMS Microbiol. Rev.* 39:96–119
16. de Souza RF, Aravind L. 2012. Identification of novel components of NAD-utilizing metabolic pathways and prediction of their biochemical functions. *Mol. Biosyst.* 8:1661–77
17. **Downs DM, Ernst DC. 2015. From microbiology to cancer biology: The Rid protein family prevents cellular damage caused by endogenously generated reactive nitrogen species. *Mol. Microbiol.* 96:211–19**
18. Engqvist M, Drincovich MF, Flügge UI, Maurino VG. 2009. Two D-2-hydroxy-acid dehydrogenases in *Arabidopsis thaliana* with catalytic capacities to participate in the last reactions of the methylglyoxal and β-oxidation pathways. *J. Biol. Chem.* 284:25026–37
19. Eugeni Piller L, Glauser G, Kessler F, Besagni C. 2014. Role of plastoglobules in metabolite repair in the tocopherol redox cycle. *Front. Plant Sci.* 5:298
20. Everse J, Zoll EC, Kahan L, Kaplan NO. 1971. Addition products of diphosphopyridine nucleotides with substrates of pyridine nucleotide-linked dehydrogenases. *Bioorg. Chem.* 1:207–33
21. Farmer EE, Mueller MJ. 2013. ROS-mediated lipid peroxidation and RES-activated signaling. *Annu. Rev. Plant Biol.* 64:429–50
22. Feist AM, Palsson BO. 2008. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat. Biotechnol.* 26:659–67
23. **Fiehn O, Barupal DK, Kind T. 2011. Extending biochemical databases by metabolomic surveys. *J. Biol. Chem.* 286:23637–43**
24. Fischer M, Römisch W, Saller S, Illarionov B, Richter G, et al. 2004. Evolution of vitamin B_2 biosynthesis: structural and functional similarity between pyrimidine deaminases of eubacterial and plant origin. *J. Biol. Chem.* 279:36299–308
25. Flaks JG. 1963. 5-Phosphoribosylpyrophosphate. *Methods Enzymol.* 6:473–79
26. Foyer CH, Noctor G. 2011. Ascorbate and glutathione: the heart of the redox hub. *Plant Physiol.* 155:2–18
27. **Frelin O, Huang L, Hasnain G, Jeffryes JG, Ziemak MJ, et al. 2015. A directed-overflow and damage-control *N*-glycosidase in riboflavin biosynthesis. *Biochem. J.* 466:137–45**
28. Galperin MY, Moroz OV, Wilson KS, Murzin AG. 2006. House cleaning, a part of good housekeeping. *Mol. Microbiol.* 59:5–19
29. Gibson JL, Tabita FR. 1997. Analysis of the *cbbXYZ* operon in *Rhodobacter sphaeroides*. *J. Bacteriol.* 179:663–69
30. Glauser G, Veyrat N, Rochat B, Wolfender JL, Turlings TC. 2013. Ultra-high pressure liquid chromatography–mass spectrometry for plant metabolomics: a systematic comparison of high-resolution quadrupole-time-of-flight and single stage Orbitrap mass spectrometers. *J. Chromatogr. A* 1292:151–59
31. Golubev AG. 2009. How could the Gompertz-Makeham law evolve. *J. Theor. Biol.* 258:1–17
32. **Golubev AG. 1996. The other side of metabolism: a review. *Biochemistry* 61:2018–39**
33. Goyer A, Collakova E, Díaz de la Garza R, Quinlivan EP, Williamson J, et al. 2005. 5-Formyltetrahydrofolate is an inhibitory but well tolerated metabolite in *Arabidopsis* leaves. *J. Biol. Chem.* 280:26137–42

14. A thoughtfully provocative article that argues that, largely because of metabolite damage reactions, metabolism governs genes rather than the other way around.

17. A short review on the hydrolysis of reactive nitrogen species by RidA proteins that illustrates the concept of metabolite damage preemption.

23. A one-stop reference on the many unidentified compounds in metabolomics and their relationship to enzyme promiscuity, pathways, and databases.

27. Describes a novel protein family that hydrolyzes riboflavin intermediates in plants and bacteria and provides a cross-kingdom example of directed overflow metabolism.

32. A must-read classic on nonenzymatic reactions, with an incisively brilliant commentary on why biochemistry has overlooked them.

34. Goyer A, Hasnain G, Frelin O, Ralat MA, Gregory JF III, Hanson AD. 2013. A cross-kingdom Nudix enzyme that pre-empts damage in thiamin metabolism. *Biochem. J.* 454:533–42
35. Goyer A, Johnson TL, Olsen LJ, Collakova E, Shachar-Hill Y, et al. 2004. Characterization and metabolic function of a peroxisomal sarcosine and pipecolate oxidase from *Arabidopsis*. *J. Biol. Chem.* 279:16947–53
36. Grimme S. 2013. Towards first principles calculation of electron impact mass spectra of molecules. *Angew. Chem.* 52:6306–12
37. Hanson AD, Pribat A, Waller JC, de Crécy-Lagard V. 2009. "Unknown" proteins and "orphan" enzymes: the missing half of the engineering parts list—and how to find it. *Biochem. J.* 425:1–11
38. Hatzimanikatis V, Li C, Ionita JA, Henry CS, Jankowski MD, Broadbelt LJ. 2005. Exploring the diversity of complex metabolic networks. *Bioinformatics* 21:1603–9
39. Held KD, Epp ER, Clark EP, Biaglow JE. 1988. Effect of dimethyl fumarate on the radiation sensitivity of mammalian cells in vitro. *Radiat. Res.* 115:495–502
40. Henry CS, Broadbelt LJ, Hatzimanikatis V. 2007. Thermodynamics-based metabolic flux analysis. *Biophys. J.* 92:1792–805
41. Henry CS, Jankowski MD, Broadbelt LJ, Hatzimanikatis V. 2006. Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophys. J.* 90:1453–61
42. Henry LK, Gutensohn M, Thomas ST, Noel JP, Dudareva N. 2015. Orthologs of the archaeal isopentenyl phosphate kinase regulate terpenoid production in plants. *PNAS* 112:10050–55
43. Horan K, Jang C, Bailey-Serres J, Mittler R, Shelton C, et al. 2008. Annotating genes of known and unknown function by large-scale coexpression analysis. *Plant Physiol.* 147:41–57
44. Hou BK, Ellis LB, Wackett LP. 2004. Encoding microbial metabolic logic: predicting biodegradation. *J. Ind. Microbiol. Biotechnol.* 31:261–72
45. Huang L, Khusnutdinova A, Nocek B, Brown G, Xu X, et al. 2016. A diverse family of metal-dependent phosphatases implicated in metabolite damage-control. *Nat. Chem. Biol.* In press
46. Hüdig M, Maier A, Scherrers I, Seidel L, Jansen EEW, et al. 2015. Plants possess a cyclic mitochondrial metabolic pathway similar to the mammalian metabolic repair mechanism involving malate dehydrogenase and L-2-hydroxyglutarate dehydrogenase. *Plant Cell Physiol.* 56:1820–30
47. Jakubowski H, Guranowski A. 2003. Metabolism of homocysteine-thiolactone in plants. *J. Biol. Chem.* 278:6765–70
48. Jankowski MD, Henry CS, Broadbelt LJ, Hatzimanikatis V. 2008. Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys. J.* 95:1487–99
49. Jeffryes JG, Colestani RL, Elbadawi-Sidhu M, Kind T, Niehaus TD, et al. 2015. MINEs: open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics. *J. Cheminform.* 7:44
50. Jones ME, Lipmann F. 1960. Chemical and enzymatic synthesis of carbamyl phosphate. *PNAS* 46:1194–205
51. 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
52. Keller MA, Piedrafita G, Ralser M. 2015. The widespread role of non-enzymatic reactions in cellular metabolism. *Curr. Opin. Biotechnol.* 34:153–61
53. Kelly R, Kidd R. 2015. ChemSpider—a tool for natural products research. *Nat. Prod. Rep.* 32:1163–64
54. Khersonsky O, Tawfik DS. 2010. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* 79:471–505
55. Kind T, Fiehn O. 2006. Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinform.* 7:234
56. Kind T, Liu KH, Lee DY, DeFelice B, Meissen JK, Fiehn O. 2013. LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nat. Methods* 10:755–58
57. Kind T, Okazaki Y, Saito K, Fiehn O. 2014. LipidBlast templates as flexible tools for creating new in-silico tandem mass spectral libraries. *Anal. Chem.* 86:11024–27
58. Klopman G, Dimayuga M, Talafous J. 1994. META. 1. A program for the evaluation of metabolic transformation of chemicals. *J. Chem. Inf. Comput. Sci.* 34:1320–25

54. A landmark review establishing enzyme promiscuity as the norm, not the exception, and explaining its mechanistic basis and evolutionary implications.

59. Kueger S, Steinhauser D, Willmitzer L, Giavalisco P. 2012. High-resolution plant metabolomics: from mass spectral features to metabolites and from whole-cell analysis to subcellular metabolite distributions. *Plant J.* 70:39–50
60. Kummel A, Panke S, Heinemann M. 2006. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol. Syst. Biol.* 2:2006.0034
61. Kuraishi S, Nito N. 1980. The maximum leaf surface temperatures of the higher plants observed in the inland sea area. *Bot. Mag.* 93:209–20
62. Lambrecht JA, Schmitz GE, Downs DM. 2013. RidA proteins prevent metabolic damage inflicted by PLP-dependent dehydratases in all domains of life. *mBio* 4:e00033-13
63. Le DT, Tarrago L, Watanabe Y, Kaya A, Lee BC, et al. 2013. Diversity of plant methionine sulfoxide reductases B and evolution of a form specific for free methionine sulfoxide. *PLOS ONE* 8:e65637
64. Linster CL, Van Schaftingen E, Hanson AD. 2013. Metabolite damage and its repair or pre-emption. *Nat. Chem. Biol.* 9:72–80
65. Lutz S, Lichter J, Liu L. 2007. Exploiting temperature-dependent substrate promiscuity for nucleoside analog activation by thymidine kinase from *Thermotoga maritima*. *J. Am. Chem. Soc.* 129:8714–15
66. Magalhães ML, Argyrou A, Cahill SM, Blanchard JS. 2008. Kinetic and mechanistic analysis of the *Escherichia coli ribD*-encoded bifunctional deaminase-reductase involved in riboflavin biosynthesis. *Biochemistry* 47:6499–507

67. A foundational paper in metabolite repair that combines biochemistry and comparative genomics to identify the repair enzymes for NAD(P)H hydrates.

67. Marbaix AY, Noël G, Detroux AM, Vertommen D, Van Schaftingen E, Linster CL. 2011. Extremely conserved ATP- or ADP-dependent enzymatic system for nicotinamide nucleotide repair. *J. Biol. Chem.* 286:41246–52

68. Maruta T, Yoshimoto T, Ito D, Ogawa T, Tamoi M, et al. 2012. An Arabidopsis FAD pyrophosphohydrolase, AtNUDX23, is involved in flavin homeostasis. *Plant Cell Physiol.* 53:1106–16
69. Meinhart JO, Chaykin S, Krebs EG. 1956. Enzymatic conversion of a reduced diphosphopyridine nucleotide derivative to reduced diphosphopyridine nucleotide. *J. Biol. Chem.* 220:821–29
70. Negelein E. 1957. Synthesis, determination, analysis, and properties of 1,3-diphosphoglyceric acid. *Methods Enzymol.* 3:216–20
71. Niehaus TD, Gerdes S, Hodge-Hanson K, Zhukov A, Cooper AJ, et al. 2015. Genomic and experimental evidence for multiple metabolic functions in the RidA/YjgF/YER057c/UK114 (Rid) protein family. *BMC Genom.* 16:382
72. Niehaus TD, Nguyen TN, Gidda SK, ElBadawi-Sidhu M, Lambrecht JA, et al. 2014. *Arabidopsis* and maize RidA proteins preempt reactive enamine/imine damage to branched-chain amino acid biosynthesis in plastids. *Plant Cell* 26:3010–22
73. Niehaus TD, Richardson LG, Gidda SK, ElBadawi-Sidhu M, Meissen JK, et al. 2014. Plants utilize a highly conserved system for repair of NADH and NADPH hydrates. *Plant Physiol.* 165:52–61
74. Niehaus TD, Thamm AM, de Crécy-Lagard V, Hanson AD. 2015. Proteins of unknown biochemical function: a persistent problem and a roadmap to help overcome it. *Plant Physiol.* 169:1436–42
75. Notebaart RA, Szappanos B, Kintses B, Pál F, Györkei Á, et al. 2014. Network-level architecture and the evolutionary potential of underground metabolism. *PNAS* 111:11762–67
76. Nowicki S, Gottlieb E. 2015. Oncometabolites: tailoring our genes. *FEBS J.* 282:2796–805
77. O'Brien PJ, Herschlag D. 1999. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* 6:R91–105
78. Ogawa T, Ueda Y, Yoshimura K, Shigeoka S. 2005. Comprehensive analysis of cytosolic Nudix hydrolases in *Arabidopsis thaliana*. *J. Biol. Chem.* 280:25277–83
79. Oh M, Yamada T, Hattori M, Goto S, Kanehisa M. 2007. Systematic analysis of enzyme-catalyzed reaction patterns and prediction of microbial biodegradation pathways. *J. Chem. Inf. Model.* 47:1702–12
80. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, et al. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42:D206–14
81. Parks LW, Schlenk F. 1958. The stability and hydrolysis of *S*-adenosylmethionine; isolation of *S*-ribosylmethionine. *J. Biol. Chem.* 230:295–305
82. Pearce FG. 2006. Catalytic by-product formation and ligand binding by ribulose bisphosphate carboxylases from different phylogenies. *Biochem. J.* 399:525–34

83. Pedruzzi I, Rivoire C, Auchincloss AH, Coudert E, Keller G, et al. 2015. HAMAP in 2015: updates to the protein family classification and annotation system. *Nucleic Acids Res.* 43:D1064–70
84. Poux S, Magrane M, Arighi CN, Bridge A, O'Donovan C, et al. 2014. Expert curation in UniProtKB: a case study on dealing with conflicting and erroneous data. *Database* 2014:bau016
85. Reaves ML, Young BD, Hosios AM, Xu YF, Rabinowitz JD. 2013. Pyrimidine homeostasis is accomplished by directed overflow metabolism. *Nature* 500:237–41
86. Richard JP. 1991. Kinetic parameters for the elimination reaction catalyzed by triosephosphate isomerase and an estimation of the reaction's physiological significance. *Biochemistry* 30:4581–85
87. Ridder L, van der Hooft JJ, Verhoeven S. 2014. Automatic compound annotation from mass spectrometry data using MAGMa. *Mass Spectrom.* 3:S0033
88. Roje S, Janave MT, Ziemak MJ, Hanson AD. 2002. Cloning and characterization of mitochondrial 5-formyltetrahydrofolate cycloligase from higher plants. *J. Biol. Chem.* 277:42748–54
89. Rudd KE. 1998. Linkage map of *Escherichia coli* K-12, edition 10: the physical map. *Microbiol. Mol. Biol. Rev.* 62:985–1019
90. Sandoval FJ, Roje S. 2005. An FMN hydrolase is fused to a riboflavin kinase homolog in plants. *J. Biol. Chem.* 280:38337–45
91. Schnoes AM, Brown SD, Dodevski I, Babbitt PC. 2009. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLOS Comput. Biol.* 5:e1000605
92. Seaver SM, Gerdes S, Frelin O, Lerma-Ortiz C, Bradbury LM, et al. 2014. High-throughput comparison, functional annotation, and metabolic modeling of plant genomes using the PlantSEED resource. *PNAS* 111:9645–50
93. Semchyshyn HM. 2014. Reactive carbonyl species in vivo: generation and dual biological effects. *Sci. World J.* 2014:417842
94. Singla-Pareek SL, Reddy MK, Sopory SK. 2003. Genetic engineering of the glyoxalase pathway in tobacco leads to enhanced salinity tolerance. *PNAS* 100:14672–77
95. Suhre K. 2007. Inference of gene function based on gene fusion events: the Rosetta-stone method. *Methods Mol. Biol.* 396:31–41
96. Sullivan LB, Martinez-Garcia E, Nguyen H, Mullen AR, Dufour E, et al. 2013. The proto-oncometabolite fumarate binds glutathione to amplify ROS-dependent signaling. *Mol. Cell* 51:236–48
97. Sun CX, Li MQ, Gao XX, Liu LN, Wu XF, Zhou JH. 2016. Metabolic response of maize plants to multi-factorial abiotic stresses. *Plant. Biol.* 18(Suppl. S1):120–29
98. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, et al. 2015. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* 12:523–26
99. UniProt Consort. 2014. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 42:D191–98
100. Van Schaftingen E, Rzem R, Marbaix A, Collard F, Veiga-da-Cunha M, Linster CL. 2013. Metabolite proofreading, a neglected aspect of intermediary metabolism. *J. Inherit. Metab. Dis.* 36:427–34
101. Van Schaftingen E, Rzem R, Veiga-da-Cunha M. 2009. L-2-Hydroxyglutaric aciduria, a disorder of metabolite repair. *J. Inherit. Metab. Dis.* 32:135–42
102. Van Schaftingen E, Veiga-da-Cunha M, Linster CL. 2015. Enzyme complexity in intermediary metabolism. *J. Inherit. Metab. Dis.* 38:721–27
103. Vinci CR, Clarke SG. 2010. Homocysteine methyltransferases Mht1 and Sam4 prevent the accumulation of age-damaged (*R*,*S*)-AdoMet in the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* 285:20526–31
104. Vinci CR, Clarke SG. 2010. Yeast, plants, worms, and flies use a methyltransferase to metabolize age-damaged (*R*,*S*)-AdoMet, but what do mammals do? *Rejuvenation Res.* 13:362–64
105. Wang Q, Xia J, Guallar V, Krilov G, Kantrowitz ER. 2008. Mechanism of thermal decomposition of carbamoyl phosphate and its stabilization by aspartate and ornithine transcarbamoylases. *PNAS* 105:16918–23
106. Wang T, Mori H, Zhang C, Kurokawa K, Xing XH, Yamada T. 2015. DomSign: a top-down annotation pipeline to enlarge enzyme space in the protein universe. *BMC Bioinform.* 16:96
107. Wang Y, Kora G, Bowen BP, Pan C. 2014. MIDAS: a database-searching algorithm for metabolite identification in metabolomics. *Anal. Chem.* 86:9496–503

85. The foundational paper for the concept of directed overflow metabolism, which it introduces in relation to bacterial pyrimidine biosynthesis.

108. Webb ME, Smith AG. 2011. Pantothenate biosynthesis in higher plants. *Adv. Bot. Res.* 58:203–55
109. Winterbourn CC, Hampton MB. 2008. Thiol chemistry and specificity in redox signaling. *Free Radic. Biol. Med.* 45:549–61
110. Witt S, Galicia L, Lisec J, Cairns J, Tiessen A, et al. 2012. Metabolic and phenotypic responses of greenhouse-grown maize hybrids to experimentally controlled drought stress. *Mol. Plant* 5:401–17
111. Wolf S, Schmidt S, Müller-Hannemann M, Neumann S. 2010. In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinform.* 11:148
112. Wu L, Serpersu EH. 2009. Deciphering interactions of the aminoglycoside phosphotransferase(3′)-IIIa with its ligands. *Biopolymers* 91:801–9
113. Yadav SK, Singla-Pareek SL, Ray M, Reddy MK, Sopory SK. 2005. Methylglyoxal levels in plants under salinity stress are dependent on glyoxalase I and glutathione. *Biochem. Biophys. Res. Commun.* 337:61–67
114. Yoshimura K, Ogawa T, Ueda Y, Shigeoka S. 2007. AtNUDX1, an 8-oxo-7,8-dihydro-2′-deoxyguanosine 5′-triphosphate pyrophosphohydrolase, is responsible for eliminating oxidized nucleotides in *Arabidopsis*. *Plant Cell Physiol.* 48:1438–49
115. Zhang P, Dreher K, Karthikeyan A, Chi A, Pujar A, et al. 2010. Creation of a genome-wide metabolic pathway database for *Populus trichocarpa* using a new approach for reconstruction and curation of metabolic pathways for plants. *Plant Physiol.* 153:1479–91
116. Zhang P, Foerster H, Tissier CP, Mueller L, Paley S, et al. 2005. MetaCyc and AraCyc. Metabolic pathway databases for plant research. *Plant Physiol.* 138:27–37
117. Zheng L, Cardaci S, Jerby L, MacKenzie ED, Sciacovelli M, et al. 2015. Fumarate induces redox-dependent senescence by modifying glutathione metabolism. *Nat. Commun.* 6:6001
118. Zhou J, Rudd KE. 2013. EcoGene 3.0. *Nucleic Acids Res.* 41:D613–24