Unity in diversity, a systems approach to regulating plant cell physiology by 2-oxoglutarate-dependent dioxygenases

Siddhartha Kundu *†

School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India

Could a disjoint group of enzymes synchronize their activities and execute a complex multi-step, measurable, and reproducible response? Here, I surmise that the alpha-ketoglutarate dependent superfamily of non-haem iron (II) dioxygenases could influence cell physiology as a cohesive unit, and that the broad spectra of substrates transformed is an absolute necessity to this portrayal. This eclectic group comprises members from all major taxa, and participates in pesticide breakdown, hypoxia signaling, and osmotic stress neutralization. The oxidative decarboxylation of 2-oxoglutarate to succinate is coupled with a concomitant substrate hydroxylation and, in most cases, is followed by an additional specialized conversion. The domain profile of a protein sequence was used as an index of miscellaneous reaction chemistry and interpreted alongside existent kinetic data in a linear model of integrated function. Statistical parameters were inferred by the creation of a novel, empirically motivated flat-file database of over 3800 sequences (DB2OG) with putative 2-oxoglutarate dependent activity. The collated information was categorized on the basis of existing annotation schema. The data suggests that 2OG-dependent enzymes incorporate several desirable features of a systems level player. DB2OG, is free, accessible without a login to all users, and available at the following URL (http://comp-biol.theacms.in/DB2OG.html).

Keywords: dioxygenases, 2-oxoglutarate, facial triad, systems biology, hidden Markov model, generalized linear model

INTRODUCTION

Physiology is an unmeasured outcome of a complex undefined molding of several underlying, disparate, and interdependent molecular level events. A particularly challenging problem is harmonizing these seemingly unrelated micro-steps to accomplish a macroscopic scale observable. The architecture of any such model would have to be sufficiently coarse-grained to take into account the magnitude and dimensions of the predicted response, yet be offset by a small set of regulatory parameters which could function as reaction modifiers. Finally, given the inherent complexity of biological systems and the varied approximations of existent models, a cumulative error term that summarizes inestimable and excluded data points is an obligatory element of any representation under consideration. In biochemical terms this would be analogous to an enzyme system which, as a unit could catalyze the conversion of several different substrates using a library of reaction chemistries, and yet possess similar kinetics with common co-factors, -enzymes, and/or—substrates. Miscellaneous contributors such as subcellular location, feedback mechanisms, number of irreversible reactions, and presence of pathways with common metabolic intermediates would also need to be factored in. A reasonable candidate appears to be the 2OG-dependent superfamily of enzymes.

The alpha-ketoglutarate (AKG) and non-heme Fe (II) dependent dioxygenases catalyze the incorporation of both atoms of molecular dioxygen in a reaction which may be described as an oxidative decarboxylation with associated hydroxylation. In these enzymes, water and 2OG function as uni- and bi-dentate ligands. The residues His-X-Asp/Glu-Xn-His form dative linkages with iron at one face of a regular octahedron, distorting it completely. This arrangement, however, may aptly be described as combinatorial, with other dioxygenases sharing a similar conserved motif-homology (Hegg and Que, 1997). The specific and fixed inter-residue metrics of the HD/EH-motif along with arginine (Arg; R), tryptophan (Trp; W), tyrosine (Tyr; Y), asparagine (Asn; N), glutamine (Gln; Q), serine (Ser; S), threonine (Thr; T), and methionine (Met; M) complete the active site of most 2OG-dependent enzymes. The superfamily is also unique for its substrate driven specialized reaction chemistry, and encompasses chlorination, desaturation, epoxidation, aliphatic- and aromatic-ring closure, and lysine demethylation. The relaxed coordination of iron, formation of an exceptionally reactive, transient ferryl species [Fe(IV)=O], and several sequence specific features are purported to contribute to the substrate versatility of 2OG-dependent enzymes (Price et al., 2003a,b; Hausinger, 2004; Clifton et al., 2006).

The major focus of this work was to numerically derive a mathematical expression that could highlight the ability of this cluster to work in a concerted and organized manner to execute complex cellular function. A critical parameter to be considered is a measure of enzyme variability. The exemplary multiformity

demonstrated by these enzymes is well-documented with biochemical, mutagenesis, and structural data available from several laboratories. Hypothetically, the distribution of AKG-catalytic domains in a particular enzyme(s) when examined against a similar spread in a larger repository of sequences might be a suitable index of this inconsonance. Since, data from characterized enzymes ($N \cong 227$; Kundu, 2012) is limited, a larger database was sought to accomplish this. The profile HMMs from Pfam lists several candidate (>5500; PF02668_TauD) sequences and is exceptionally detailed. However, this data is based on generic models which includes sequences with unvalidated and putative function. A novel database, DB2OG, was then constructed wherein predicted catalytic domains of biochemically validated enzymes were mapped onto a set of sequences which had no supporting laboratory data. This extrapolated parameter, i.e., the simultaneous domain probability with respect to the taxa under consideration, when merged with available kinetic data could offer insights into the system being modeled.

Plant 2OG-dependent enzymatic activity is narratively charted, with genomic and proteomic data, and their variants thereof. They participate principally as hormonal regulators (gibberellic acids, jasmonic acid), general flavonoid metabolism, alkaloid biosynthesis, and maintainence of structural integrity (prolyl hydroxylases) (Kawai et al., 2014). The presence of a rigid cell wall, absence of an extracellular matrix, and the abundance of organelles constitutes a defined experimental system amenable to hypothesis testing with minimal confounders. These features were instrumental in selecting the phytocellular mileau as the system of study.

MATERIALS AND METHODS
CONSTRUCTION AND ANALYSIS OF DB2OG
The database, DB2OG, was created using a generic hidden Markov model (HMM) of alignments (Kundu, 2012) and used to query the UniprotKB database (http://uniprot.org). Briefly, members of the alpha-ketoglutarate superfamily (EC 1.14.11.x, EC 1.14.20.y) with available empirical data were randomly clustered, with at least 2–4 members of each family contributing to the model. In this work, a family is defined as enzymes with the same substrate profile. Several filters were applied to screen this initial set. The final list ($N = 3806$), comprised experimentally unvalidated full-length protein sequences. A detailed chart of the catalytic profile of these sequences was constructed using the server module of H2OGpred (http://comp-biol.theacms.in/H2OGpred.html; Kundu, 2012). This matrix of profiles (Tables T1; S1A, S1B) is used to search DB2OG for suitable matches. Analysis to ascertain the preferred intracellular location was carried out with the PSORT suite of programs (Horton et al., 2007; Yu et al., 2010; Tables T3, T4). The nomenclature of the clusters is with regards to the substrates they catalyze, and is in accordance with previous work (Kundu, 2012). Chemical structures were drawn by combing data from PubChem_Compound (http://ncbi.nlm.nih.gov/pccompound) and ChemSketch 12.00 (freeware) downloaded and installed locally. Phylogenetic trees and alignment files were generated using the STRAP suite of programs (Gille and Frommel,

2001), and all biochemical data was extracted from BRENDA (Schomburg et al., 2002).

IMPLEMENTING DB2OG AS A UNIVERSALLY ACCESSIBLE PORTAL
Unlike the sequence based pairwise-scoring and threshold-exceeding search scheme, DB2OG was populated by formulating the query string as a profile-HMM. The database is organized into three sections: (a) introduction to 2-oxoglutarate dependent dioxygenases with explanations of the types of files that are accessible and general usage, (b) database search. This is subdivided into, (i) a pre-defined sequence based search, and (ii) a user-specific patterned query. The sequences were previously categorized on the basis of their taxonomic spread and their predicted cell locations. The results can be downloaded as standard fasta (*.fasta) files. The specialized query option is exclusively determined real-time by the user. The search criteria, which are logical functions are initially used to probe a flat-file implementation of the computed chart of profiles. The selected domains are used to retrieve the sequences. All coding was done in-house using PERL. The website was implemented with standard PERL- CGI-HTML codes. Consider the following examples for a profile-based search:

Case a: Identify sequences of DB2OG **with** the following SULF-, TFDA- domains, but exclude those with a PHYT-domain ($x \in DB2OG|$ {SULF, TFDA} $\subseteq x$; {PHYT} $\not\subset x$).

Case b: Identify sequences of DB2OG **with** the following SULF-, TFDA- domains ($x \in DB2OG|$ {SULF, TFDA} $\subseteq x$)

Case c: Identify sequences of DB2OG **without** a PHYT-domain ($x \in DB2OG|$ {PHYT} $\not\subset x$)

In these examples cases- b and -c are simple search options using logical "OR," while case- a would constitute a complex search with a logical "AND" function.

MATHEMATICAL MODELS
HMMs are a sub-class of Markov models (a model where the future system state is determined by the current state and is, therefore "memoryless"), wherein, given the presence of observables, algorithms are used to compute the sequential probabilities of occurrence of hidden states (Markov chains), and are part of the online repositories Pfam and H2OGpred among several others (Sonnhammer et al., 1997; Kundu, 2012). HMMs used in this work were constructed using HMMER3.0 (http://hmmer.janelia.org) downloaded locally. The preliminary alignments as fasta (*.fasta) formatted files were converted to the stockholm (*.sto, *stockholm) formats. Generalized linear models (GLMs), are mathematically, linear combinations of more than one independent variable. This abstract representation is able to hypostatize the functioning of several enzymes into a concise numerical summary which can be reverse mapped to well-established distributions such as the gaussian, poisson's, and exponential probability densities. A statistical model was formulated to incorporate known and predefined parameters (Equations 1 and 2) of enzymes that could contribute to a particular reaction. The final computations were done using the generalized linear model

(GLMs) function of R-3.0.0 (S3). An algorithm outlining the steps involved in these calculations is mentioned.

RESULTS AND DISCUSSION
UNIQUE FEATURES OF DB2OG
Access and usage
The GUI introduces and summarizes the salient features of alpha-KG dependent enzymes, and provides general instructions of use. The user may opt to download various categories of sequences or search the database for combinations of previously computed profiles (twenty-six functional hidden Markov model profiles; Kundu, 2012). Since, most sequences of the database have more than one assigned profile, there are several examples of overlapping data. This may be minimized by usage of the logical OR/AND functions to filter the profiles (**Figure 1**). The matched list is presented as a simple text file, which may be downloaded or viewed online. Users are encouraged to refresh their browsers to visualize the latest search results.

Description and characterization of DB2OG
The taxonomic distribution of the sequences of DB2OG (**Figure 2A**), appropriately mirrors the assortment of biochemical reactions that typifies this superfamily. An interesting observation is the predicted cellular location of some of these sequences. While, the cytosol is the preferred location across all taxa, a

considerable fraction of fungal proteins appear to be associated with the nucleus and the cytoskeleton when compared to animal and plant sequences. In comparison, animal 2OG-dependent enzymes seem to localize to the endosome (E.R. and Golgi apparatus), mitochondria, the plasma membrane, and the extracellular space. This perhaps reflects the need for glycosylation to achieve optimal activity with subsequent secretion (**Figure 2B**; Table T3). Analysis of the sequences specific to plants shows expected patterns. The vast majority are enzymes that participate in flavonoid metabolic pathways, gibberellic acid (GA) catabolism, 1-amino-1-cyclopropane carboxylate oxidase (ACC-oxidase; the terminal enzyme in ethylene synthesis), prolyl hydroxylases, alkaloid biosynthetic enzymes, and senescence related proteins, like the hormone regulator jasmonic acid (**Figures 2C,D**; Table T4). A large group of protein sequences which are either un- or minimally- annotated, or present nominally (PROB, $N = 727$), are grouped together, with probable roles in metal chelation, defense, and detoxification. Some of these proteins are predicted to complex iron directly, notably, the phytosiderophores (Nakanishi et al., 2000).

Distribution of catalytic domains in plant 2OG sequences
A comparative analysis of plant 2-oxoglutarate domains with sequences of other taxa ($N = 1624$; **Figure 3**, Tables T1, T2;

FIGURE 1 | Profile based search of sequences in DB2OG. Users may query the database for putative 2-OG dependent sequences using one or more of the pre-defined HMM profiles ($N = 26$). A structured query of this nature (simple or complex), without an inclusion score will list all sequences with the profile of interest. Other sequences corresponding to variable subcellular locations may also be downloaded by clicking on the appropriate hyperlinks. NOTE: The browser window displaying the search results needs to be refreshed for each query.

FIGURE 2 | Analysis and description of DB2OG. (A) Taxonomic distribution of sequences among all known kingdoms of life, **(B)** Predicted sub-cellular location of sequences using the PSORT suite of programs. **(C)** 2-OG dependent enzyme groups in plants ($n = 1624$), with their **(D)** predicted cellular location. P4H, unlike the other groups has a compartmental preference, with appreciable numbers in the nuclear, mitochondrial, vacuolar, chloroplast, and extracellular regions of the cell. GA-oxidases have an almost comparable cytosolic and nuclear distribution. Data represent normalized values of the highest score for a particular cellular region in either plants, animal, or fungi **(C)** or 2OG-dependent enzyme clusters **(D)**. Abbreviations: ALKB, Alk-B like demethylase; ARGI, Arginine hydroxylase; ASPA,

Aspartyl:Asparaginyl hydroxylase; CHLO, Chlorination; CLAS, Clavaminate synthase; COLY, Collagen lysyl dioxygenase; CP3H, Collagen prolyl 3-hydroxylase; CP4H, Collagen prolyl 4-hydroxylase; CYCL, Cyclization; DACS, Deacetoxycephalosporin-C synthase; DSAT, Desaturase; ECTO, Ectoine hydroxylase; FLAV, 2S-Flavones; GBBH, γ – butyrobetaine hydroxylase; GIAC, Gibberellic acid; HILY, Histone lysyl demethylase; HP4H, Hypoxia prolyl 4-hydroxylase; HYOS, Hyoscyamine; NUHY, Nucleotide/side hydroxylase; OGFD, Eukaryotic initiation factor 2α (eIF2α); PHYT, Phytanoyl-CoA; SULF, Sulfate cleaving; TFDA, 2,4-Diphenoxyacetic acid metabolizing; THYD, Thymidine dioxygenase; THYE, Thymine dioxygenase; XANT, Xanthine hydroxylase.

S2A, S2B), conveys important information on the nature of enzymes that have specialized functions. The domains DACS (Deacetoxycephalosporin-S synthase/expandase), GIAC (Gibberellic acid metabolizing), HYOS/ALKO (Hyoscyamine transforming), NUHY (nucleotide/side hydroxylases), and THYE (Thymine dioxygenase) are exclusively high scoring (q4), in comparison to other organisms. Coupled mono- (DSAT; desaturase) or poly- (FLAV; 2S-flavones) hydrogen- abstracting, i.e., desaturases, have (q4) scores, but are shared with fungal spp. The domains SULF (sulfate cleaving), CHLO (chlorinating), and CYCL (cyclization) are uniformly high (q3, q4), in all sequences examined. In contrast, the low scoring plant

GBBH (γ-butyrobetaine), COLY (collagen lysyl-dioxygenases), and TFDA (enzymes catalyzing the degradation of phenoxyalkanoic acid herbicides) (q1), and XANT domains (Xanthine hydroxylase; q2) have high valued equivalents (q4) in fungal, animal, and bacterial sequences. Homogenous low scoring CP3H (collagen prolyl 3-hydroxylase) and THYD (Thymidine dioxygenases) (q1, q2) were also part of this profile. These results suggest that while collagen lysyl-hydroxylases (COLY) predominate, expectedly in animal tissues, gibberellic acids (GIAC), alkaloids (HYOS), deacetyl/deacetoxycephalopsorin-C synthases/expandase domains (DACS) populate plant sequences. The susceptibility of plant DNA to undergo modifications is

FIGURE 3 | Domain analysis of sequences of DB2OG. (A) Taxonomic spread of catalytic domains in sequences, **(B)** Quartile representation of domains. Abbreviations: F, fungi; A, animals; GP, gram positive; GN, gram negative; P, plants and algae; NP, non-plants. Nomenclature for enzyme families, see abbreviations in **Figure 2**.
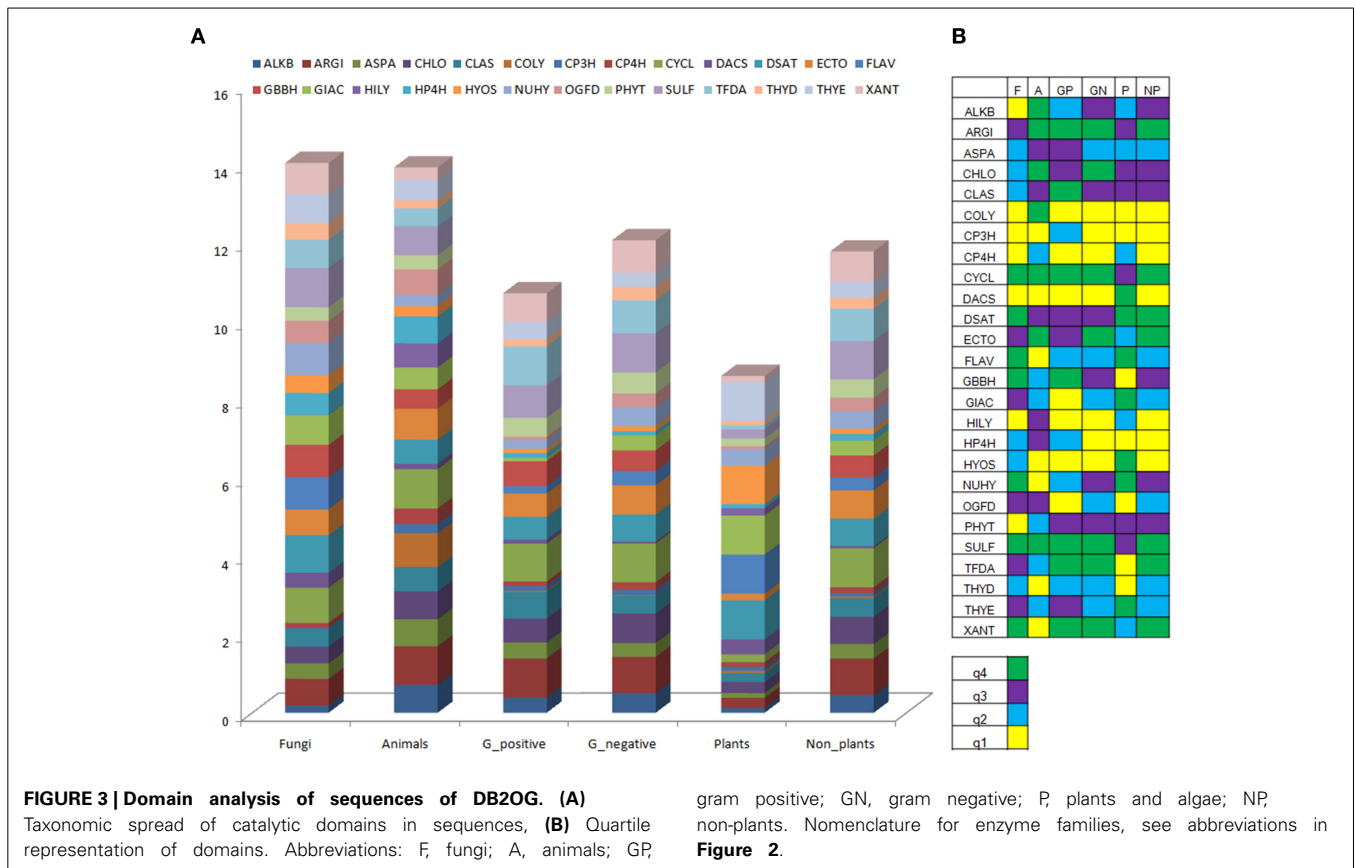
reflected by the presence of the NUHY and related THYE domains.

The combined domain frequency (α_i; Table T2C) for a particular enzyme is the product of its occurrence within the plant sub-database and DB2OG, and is used in this work as a plausible measure of the catalytic domain spread within plant 2OG-dependent sequences.

STATISTICAL REPRESENTATION OF 2OG-DEPENDENT SEQUENCES
Generalized linear model

A predefined objective function (**F**) is approximated by a linear combination of variables, a value which is determined by the number of enzymes hypothesized (Equation 1). Several case scenarios are listed wherein enzymes may be part of the same or distinct metabolic pathways, catalyze common or related substrates, or recognize specific molecular stereochemistry with the production of structural isomers (**Table 1**).

$$F = \sum_{i=1}^{N=n} C_i X_i + E_i \qquad (1)$$

$$C_i = -\log\left(\alpha_i P_i \left(\prod_{j=1}^{T=t} Km_j\right)\right) \qquad (2)$$

$\alpha_i :=$ combined domain frequency

$P_i :=$ probability of reaction occurring

$Km_j :=$ average substrate affinity

$E_i :=$ error term

$X_i :=$ distinct generic metabolite

$n =$ number of enzymes considered together

$t =$ number of Km values

Algorithm to compute coefficients

Step 1: Formulate the master equation.

Step 2: Take each term of F, i.e., C_i separately and use Equation (3).

$$(\beta_k)(C_i) = \delta_k \qquad (3)$$

$C =$ result from equations 1 and 2

$\beta, \delta :=$ random numbers (0, 1)

k := observation

Step 3: Iterate for 100 observations.

Step 4: Introduce the GLM formula and use the "quasi" family of distributions to compute and analyze the coefficients, significance, and the standard error.

Step 5: Sum the errors.

The formulation (Equations 1 and 2) ensures that the partitioning probability, domain frequency, and Km values to determine values of the F. The following scenario illustrates the significance and utility of this combinatorial approach.

Table 1 | Alpha-ketoglutarate dependent enzymes as mediators of synergistic function.

	Enzyme	Km (mM)						References
		$\alpha^{\ddagger}$	Iron (II)	2-oxoglutarate	Dioxygen	Substrates		
A	Prolyl 4-hydroxylase (HP4H)	0.11	0.0031	0.03	0.186	0.053		De Jong and Kemp, 1984; Hirsila et al., 2003; Kukkola et al., 2003; Myllyharju, 2003, 2008
B	Prolyl 4-hydroxylase (CP4H)	0.12	0.0031	0.021	0.04	0.311		Kivirikko and Myllyla, 1982; Kivirikko et al., 1989; Helaakoski et al., 1995; Lamberg et al., 1995; Annunen et al., 1997, 1998, 1999; Hieta et al., 2003; Kukkola et al., 2003; Myllyharju, 2003, 2008; Kersteen et al., 2004
C	Prolyl 3-hydroxylase (cis)	0.10		0.08		0.3975		Mori et al., 1997; Shibasaki et al., 1999, 2000
D	Prolyl 3-hydroxylase (trans)	0.10	0.0005	0.08	0.03	0.165		Tryggvason et al., 1979; Tiainen et al., 2008
E	Asparagine hydroxylase	0.17	0.0005	0.019	0.09	0.01		Hewitson et al., 2002; Koivunen et al., 2004; Flagg et al., 2012
*Subcase 1	A + E			$F = (3.0266)\mathbf{X1} + (4.3943)\mathbf{X2} + 0.9469$				X1:= 4-hydroxyproline X2:= hydroxyasparagine
*Subcase 2	(A + B) + E			$F = (7.54375)\mathbf{X1} + (4.3943)\mathbf{X2} + 0.58821$				X1:= hydroxyproline X2:= hydroxyasparagine
*Subcase 3	(A + B + C + D) + E			$F = (1.6997)\mathbf{X1} + (4.3943)\mathbf{X2} + 0.6779$				X1:= hydroxyproline X2:= hydroxyasparagine

*Coefficient details (S3): Subcase 1: X1:: se = 0.4515, p = 1.3E-09, df = 99; Subcase 2: X1:: se = 0.09281, p < 2.00E-16, df = 99; Subcase 3: X1:: se = 0.1825, p = 3.7E-15, df = 99; Subcases 1, 2, 3: X2:: se = 0.4954, p = 3.41E-14, df = 99.

$\ddagger$See Table T2C in Data sheet 2.

Case: Role of prolyl 4-hydroxylase and asparagine hydroxylase in plant biology. Both these enzymes, acting together are known to mediate the hypoxia-response in mammalian tissues. Here, the model may be a minimalistic or conversely, represent all functioning enzymes in great detail. These variants are referred to as subcases and lead to the computation of different coefficients (**Table 1**).
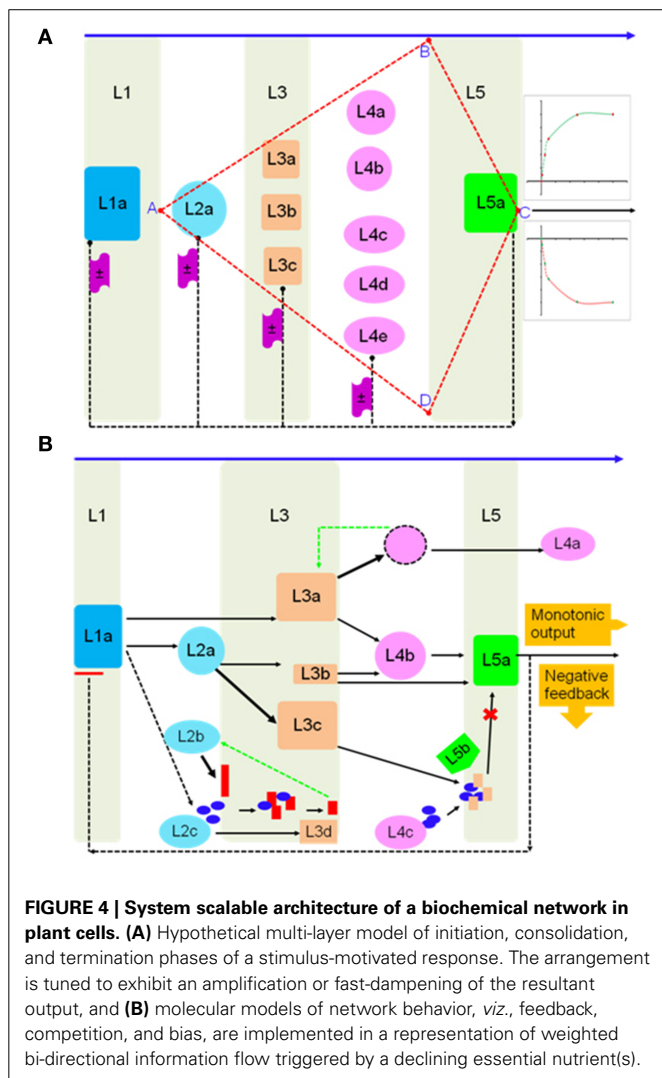
Conformity to the indubitable

There are multiple ways to represent/model this biochemical federalism. Regression models (linear and non-linear) mandate the presence of a large corpus of raw data which could then be fitted to a suitable equation. The fit is assessed by examining the values of the coefficient of determination (R^2). Here, model quality improves as $R^2 \rightarrow 1$. Although computationally intensive, polynomial equations ($f(x) = P$) of high degrees ($n > 5$) perform considerably better. Artificial neural networks (ANNs), which may consist of several layers of hidden nodes and a weight modifying function, need to be initially trained on existent datasets. The statistical model along with the computational algorithm outlined in this work is robust and theoretically implementable, compromising perhaps partially on the specificity for a particular system. The information used for the final calculations subsumes the frequency of catalytic domain occurrence (structural), numerical probability of a reaction, and Michaelis-Menten (Km/Ka) constants for several key substrates and cofactors (reaction chemistry). This basic formulation is followed by using the monte carlo method to predict a possible set of inputs (S4), which in turn is used by the GLM, as inputs to a family of distributions (poissons, gaussian, gamma, inverse gaussian, binomial) for comparative analysis. The data suggest the computed grand mean might serve as accurate estimators of the final coefficients (**Table 1**). The error values for a particular subcase are an aggregate of the standard errors of each coefficient predictor. Equation (1), computes a modified probability function that when used could be approximated by well-known probability density functions (PDFs). The subsequent parameterization could then used to glean salient features of the underlying system. Consider the subcases −1, −2, and −3 (**Table 1**), the enzymatic data suggests a progressive coarse graining (specificity → 0). In these cases studies a normal distribution provides the best fit. The selection of the random number from a given set of values (Gaussian/uniform) clearly highlighting the rationale of this approach. The system under study may then be mapped and investigated numerically.

2OG-DEPENDENT DIOXYGENASE ACTIVITY IN REGULATING SYSTEMS-LEVEL FUNCTION IN PLANTS

Deploying a top-down strategy, a generalized system response can be fractionated into an interconnected sequence of distinct steps ($S = \int dx$; **Figure 4A**), arranged as part of a multilayered complex (L1–L5; **Figure 4A**). Here, the expression F (Equation 1) might be a thought of as a solution to the system-model ($S \cong F + c_0$). If, the constraint of a cascade were imposed ($AB \cup AD$), downstream processes, which are disproportionate to the initiating stimuli can be embedded. Here, a cascade is defined as being scalable (λ), with factors for amplification or damping ($S = \pm\lambda \int dx$). Fine-steplets can be programmed by incorporating positive- (L2 ⇌ L3 ⇌ L4; **Figure 4A**) and negative-feedback mechanisms (L2b + L2c ⇌ L3d; L3a ⇌ L4a; **Figure 4B**). These would allow an element of robustness into the system making it less sensitive and more threshold dependent. Horizontal interconnectivity between individual steps may vary from competition for a single resource (shared metabolic intermediates; L2a → L3b ∨ L2a → L3c), to shared storage allocation nodes (nutritional reservoir; L3b → L4b ∧ L3c → L4b) (**Figure 4B**).

FIGURE 4 | System scalable architecture of a biochemical network in plant cells. (A) Hypothetical multi-layer model of initiation, consolidation, and termination phases of a stimulus-motivated response. The arrangement is tuned to exhibit an amplification or fast-dampening of the resultant output, and **(B)** molecular models of network behavior, *viz.*, feedback, competition, and bias, are implemented in a representation of weighted bi-directional information flow triggered by a declining essential nutrient(s).

Reversibility (sub-threshold), can be a function of nutrient deprivation/consumption or including negation equivalents of key parameters, while, plasticity or an irreversible change may be molded into the system by switching to amplification/termination mode.
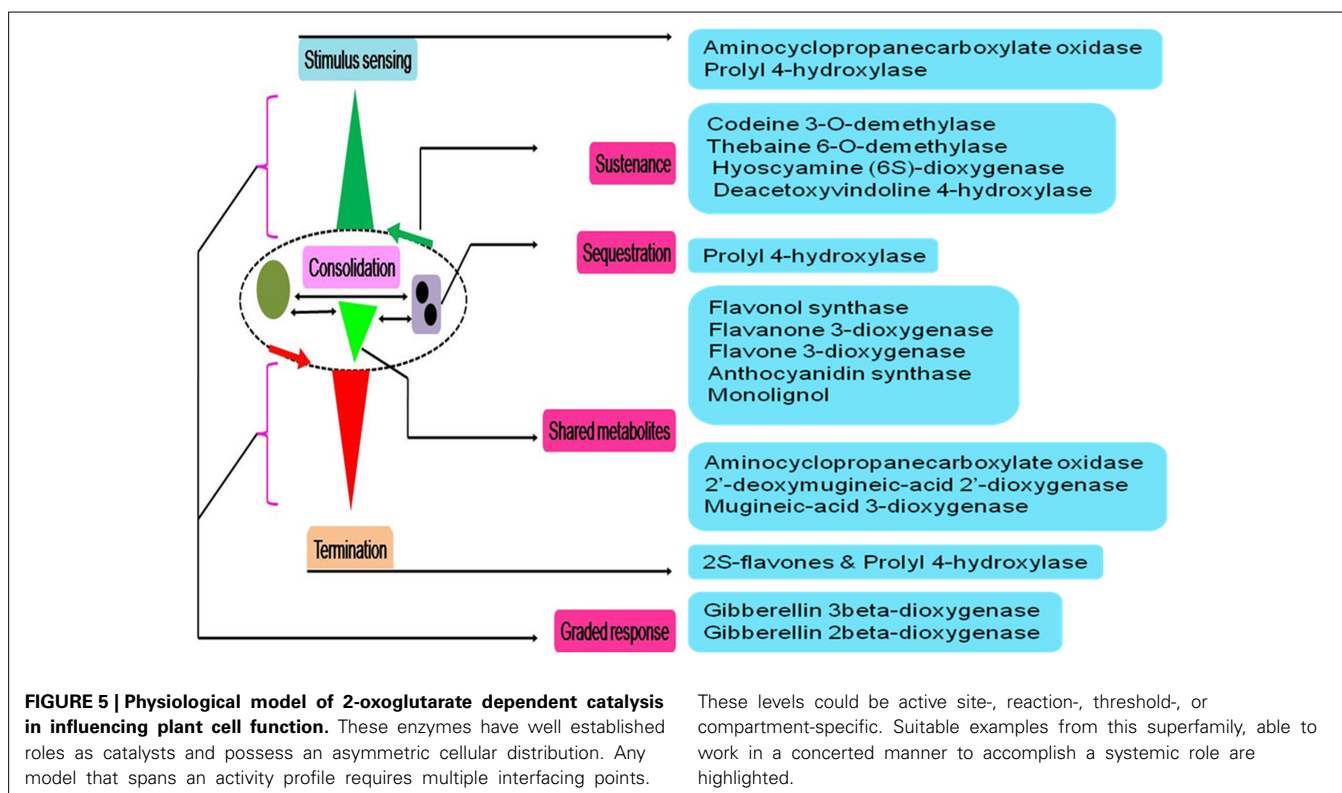
The biological role of individual members of this superfamily of enzymes is unequivocal. However, given its taxonomic spread, conserved requirement of co-factors and—substrates, substrate and reaction chemistry heterogeneity, the systems level relevance of these enzymes remains speculative. In mammals, these enzymes might work to coordinate a transcriptional response to hypoxia, fatty-acid metabolism, and structural integrity, while in bacteria and fungi, antibiotic synthesis and pesticide breakdown may be the dominant roles. In plants the hormonal regulation of growth and development involving ethylene and gibberellins are well-characterized, as is the maintenance of rigid cell wall and transport of critical nutrients.

Monotonicity and substrate heterogeneity as pre-requisites for systems play

The archetypal melange of 2OG-dependent substrate turnover is crucial to their ascribed role as a systems player. Here, the distribution of catalytic domains in DB2OG has been utilized as an index of this variability. Other discrete measures take into account active site geometry and sequence homology as possible indices (Hausinger, 2004; Clifton et al., 2006; Kundu, 2012). The definition and prediction of domains used in this work conjoins structure-sequence data with the theoretically rigorous HMMs (Kundu, 2012), and is therefore a suitable measure of substrate and reaction heterogeneity. Manifestations of this could be a facultative requirement for 2OG (ACCO, EC 1.14.17.4; Rocklin et al., 2004; Zhang et al., 2004; **Figure 6A**), or an incomplete facial triad (SyrB2; Blasiak et al., 2006). However, 2OG-dependent dioxygenases are inherently bipolar. While at one extreme members possess a common jelly-roll fold, arranged as major ($N = 7$–10) and minor ($N = 3$–5) antiparallel β-strands, at the other is the breadth of substrate modifying action, secondary to an equally divergent reaction chemistry. This unifying structural signature suffices, in so far as the dependence on 2OG as an adjunct donor and requirement for Fe (II) is concerned, permitting the regulatory arm of this superfamily to be conserved, a factor that is essential to executing a systems-level response.

Sensing and triggering a response

A pan-systemic role for AKG-dependent dioxygenases may only be envisaged if members populate, and thereby, influence the cascade at all possible levels (**Figures 4**, **5**). This would, in turn only be feasible if, supplemental to the variations discussed, *vide supra*, there were examples of graded responders to a particular stimulus or substrate. The enzymes prolyl 4-hydroxylase (P4HY; **Figures 5**, **7B**) and ACCO might serve to facilitate these effects as sensors and terminators of a particular response. An analysis of kinetic data for oxygen in P4Hs (EC 1.14.11.2) suggests that at Km values greater than 0.065 mM, the enzyme may function as a index for declining cytosolic oxygen (De Jong and Kemp, 1984; Hirsila et al., 2003; Myllyharju, 2008), while the higher affinity

FIGURE 5 | Physiological model of 2-oxoglutarate dependent catalysis in influencing plant cell function. These enzymes have well established roles as catalysts and possess an asymmetric cellular distribution. Any model that spans an activity profile requires multiple interfacing points. These levels could be active site-, reaction-, threshold-, or compartment-specific. Suitable examples from this superfamily, able to work in a concerted manner to accomplish a systemic role are highlighted.

forms, i.e., <0.065 mM function as efficient catalysts (Hutton et al., 1967; Kivirikko and Myllyla, 1980; Tanaka et al., 1980; Chrispeels, 1984; Myllyharju, 2008). Another salient feature of P4Hs, at least in plants is the almost complete sequestration (≈85%; $N = 22$) of this enzyme in various organelles. Since, proline/hydroxyproline-rich glycoproteins (PRPs) are abundant (≈20%; Hijazi et al., 2014) in cell walls, this may impose a measure of sensitivity to 4-hydroxyproline levels. Any insult to P4H activity in the form of insufficient co-factors or transporters could weaken the cell wall, and initiate a downstream response.
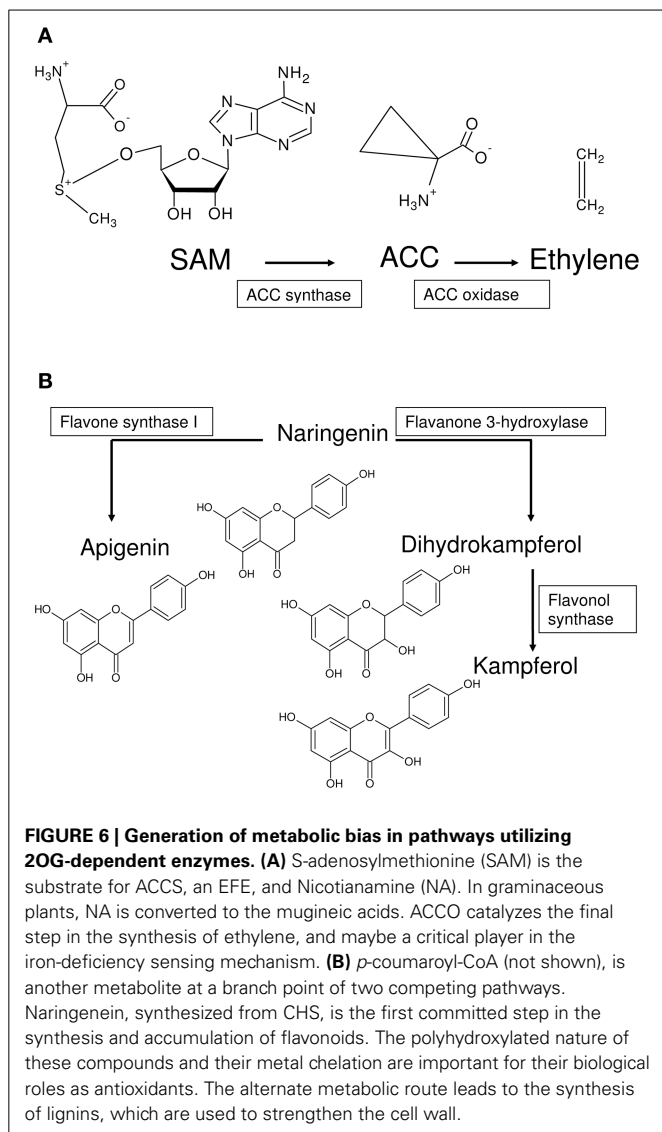
The role of ACCO appears more complex. The presence of non-enzymatic product formation or facile electron transfers (FETs) ensures a steady production of ethylene gas (≅0.35 mol; Rocklin et al., 2004). The facultative role for 2OG and ascorbic acid in ACCO is puzzling given the conserved active site residues. An examination of the active site of the crystal structure (Zhang et al., 2004; PDB_ID: 1WA6), suggests that despite a relatively unconstrained and open structural core, narrow conformational pockets in the crystal structure of ACCO might favor binding of the more compact and abundant ACC (MM $\sim$ 101.1 g mol^{-1}), rather than AKG (MM $\sim$ 146 g mol^{-1}) to critical amino acids. ACCO also possesses a low Km (0.111–0.125 mM; Kosugi et al., 1997) for its cognate substrate (ACC). These factors may ensure a high active-site occupancy by ACC at all times with a resultant decreased need for 2OG.

Consolidating the response by sharing metabolic intermediates
To comprehend the importance of this stage a clear definition of what it entails needs to be outlined. Consolidation, here, describes a decrease in the probability of a reverse step. Viewed from thermodynamic principles, a decrease in entropy might be a suitable analogy. As in the case of supra-threshold stimuli, sub-threshold states, too may be considered bound by small intervals (δS). This would imply that progression to the next sub-interval $dS(t) \rightarrow dS(t + 1)$ is dependent on achieving this unidirectionality. Given, the multitude of overlapping reactions in a cellular mileau, this dynamic bias is a pre-requistite to any product formation or its consequences thereof. Absence of this reaction vector would render the system static and unresponsive, an undesirable outcome if stressor neutralization were the objective.

Four kinds of bias may be identified in 2OG-dependent plant dioxygenases (**Figures 4**, **5**): (a) competing routes of reactions with shared substrate preferences. In the event of a flux toward a particular pathway, metabolites of the other decline reciprocally. The major biosynthetic channel of the flavonoids (**Figure 6B**), is the routing of *p*-coumaroyl-CoA through chalcone synthase (CHS, EC 2.3.1.74), in competition with the monolignol synthesizing hydroxycinnamoyl-CoA: shikimate/quinate hydroxycinnamoyl-CoA transferase (HCT, EC 2.3.1.133) (Burbulis et al., 1996; Hoffmann et al., 2004; **Figure 6**), (b) continued product utilization is an alternate strategy to favor certain reactions, thereby, introducing a dominant direction (*L3a* ⇌ *L4a*; **Figure 4B**). Activated L-methionine (S-adenosylmethionine, SAM) is the precursor for both ACC and nicotianamine (NA), which in the presence of Fe-deficiency, leads to the unhindered synthesis and efflux of the mugineic acids (MAs) which function as phytosiderophores (Nakanishi et al., 2000; Nozoye et al., 2011; **Figure 6**), (c) co-factor driven altered enzymatic activity of gibberellic acid (GA) metabolizing

FIGURE 6 | Generation of metabolic bias in pathways utilizing 2OG-dependent enzymes. (A) S-adenosylmethionine (SAM) is the substrate for ACCS, an EFE, and Nicotianamine (NA). In graminaceous plants, NA is converted to the mugineic acids. ACCO catalyzes the final step in the synthesis of ethylene, and maybe a critical player in the iron-deficiency sensing mechanism. **(B)** p-coumaroyl-CoA (not shown), is another metabolite at a branch point of two competing pathways. Naringenein, synthesized from CHS, is the first committed step in the synthesis and accumulation of flavonoids. The polyhydroxylated nature of these compounds and their metal chelation are important for their biological roles as antioxidants. The alternate metabolic route leads to the synthesis of lignins, which are used to strengthen the cell wall.

FIGURE 7 | 2-oxoglutarate-dependent dioxygenases as stimulus reinforcers. (A) Anabolic (GA20ox, GA3ox) and catabolic (GA2ox) alpha-ketoglutarate dependent enzymes function to prepare a cell mileau conducive to growth and development (increased formation and storage of alkaloids, expansins, glycosylases, and decreased formation of lignins) and a modulatory antioxidant process (formation of flavonoids). **(B)** Plant P4Hs catalyze the hydroxylation of proline-rich peptide segments of cell wall glycoproteins, stabilizing the macromolecule. Conversely, decreased activity might result in loosely networked unstable components, that are required for continued growth.

enzymes (GA −20, −3, −2 oxidases; EC 1.14.11.x, x = −, 15, 13) (**Figure 7A**). The kinetics (GA3O, $Km_{Fe} \cong 0.2$ mM; GA2O, $Km_{Fe} \cong 1.0$ mM) (Smith and MacMillan, 1984; Kwak et al., 1988), suggest that if cytosolic iron levels drop to levels below 1mM, the activity of the catabolic GA2ox approximates a null value. This balance between the highly active GA −1, −3, −4, −5, and −7 and the inactive GAs −8, −34, −97, −110 (Hedden and Phillips, 2000a,b) is critical to seed dormancy, root and shoot development, flowering, and generalized cell elongation. GAs serve as master regulators of other enzyme systems as well, and (d) initiate a cascade of reactions utilizing high affinity AKG-dependent enzymes for small molecule modifiers. Activated molecular dioxygen is a key component of these enzymes and could diffuse out if catalysis was compromised. The active site geometry of ACCO ($Km_{Fe} \cong 0.059$ mM; Nagahama et al., 1991) not just permits FETs, but may mitigate the effects of co-substrate withdrawal as well. The free radicals generated subsequently by O_2 (ROS and RNS) amplify the response manifold. The rapid
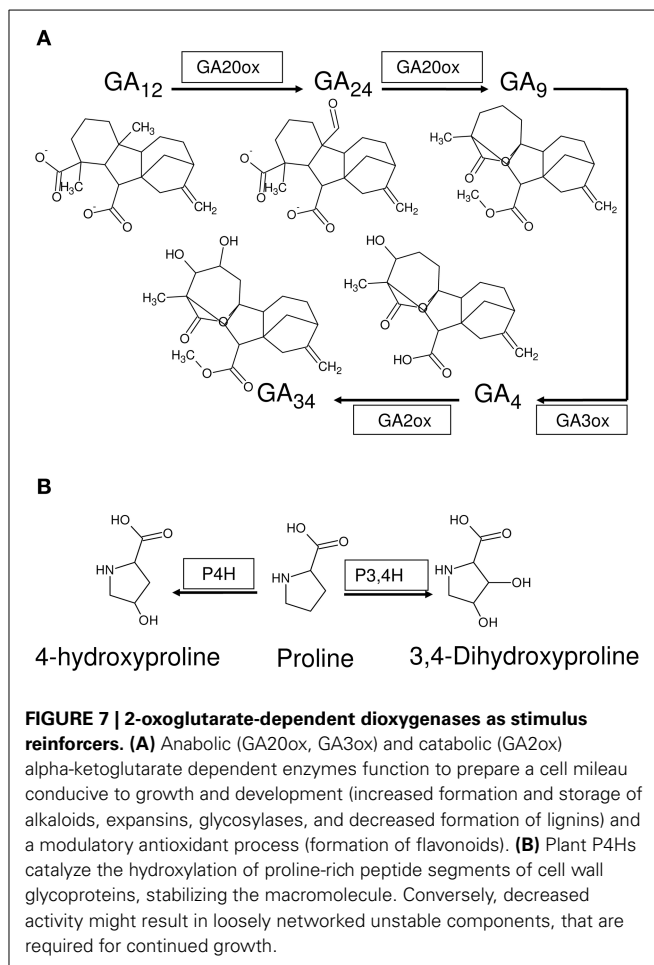
generation and involvement of substrate radicals (lipid peroxidation of polyenes), could translate into an equally swift consumption of potential substrates, which indirectly, could introduce the requisite bias.

The aforementioned discussion results in a systemic response that is committed and increasingly ordered. Preservation of the division between system-elasticity and -plasticity is co-terminus with evolution of complex systems and translates into a reversibility factor whose mathematical limit approximates zero. This lower bound can be realized biochemically by sets of reactions that negate each other or simply exhaust an important resource.

Sustaining/terminating the response for/after a finite duration

The reversibility of any path is integral to regulating a response proportionate to the stimulus. AKG-dependent proteins working together, and with other exponents of cell function can facilitate the near-cessation of a response (**Figures 4, 5**). In most cases this necessitates sophisticated feedback mechanisms embedded in the implementing nodes.

These may be intermittent (fluctuate) or persistent (uniform). A fluctuating system is dependent on the presence of a connection between the output and input. Here, the magnitude of difference between the observed and desired outputs is funneled back to the input signal, which may be modeled as
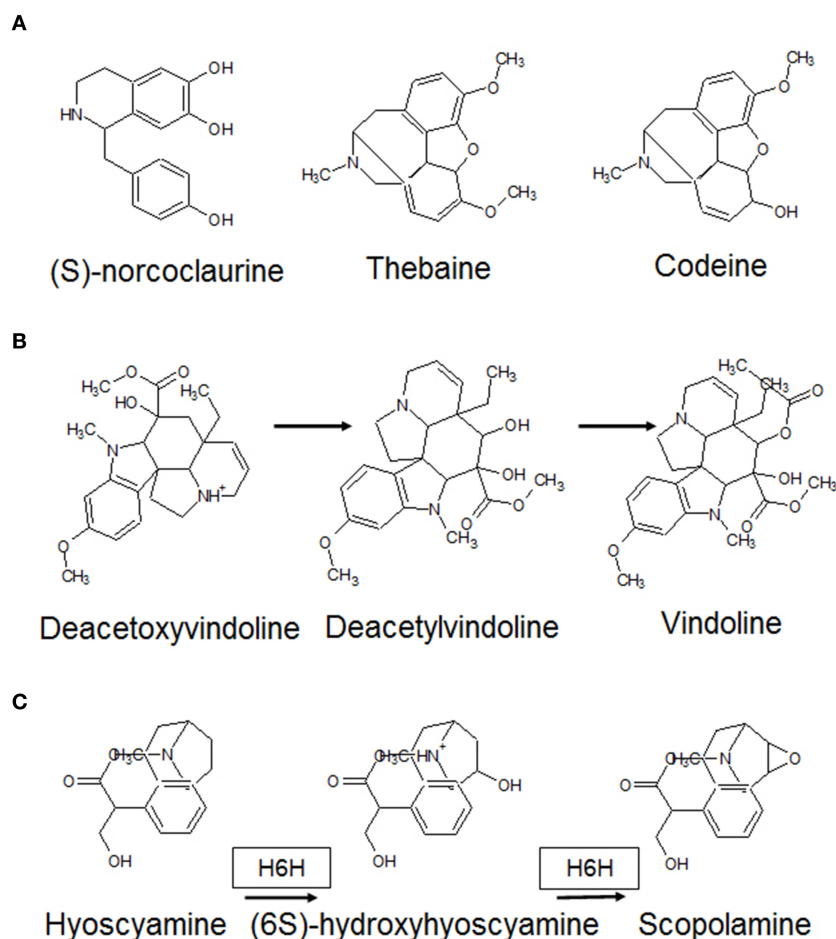
FIGURE 8 | Alkaloid classes that require 2-oxoglutarate as a co-substrate. These N-heterocyclic molecules are often synthesized in preparation for growth of new tissue and serve as an on-demand nitrogen source accessible at all times. As the tissues grow and differentiate, alkaloid reserves have been shown to progressively deplete. **(A)** (S)-norcoclaurine is the precursor of the isoquinoline classs of alkaloids (codeine, thebaine). These require demethylases, which interconvert them. **(B)** The vinca alkaloids (precursor: deacetoxyvindoline), are inhibitors of microtubule transport and terminally synthesized compounds (vincristine, vinblastine) are used in various chemotherapy regimens. **(C)** The tropanes are best represented by hyoscyamine and scopolamine both of which require activity of H6H (Hyoscyamine 6β-hydroxylase).

a weighted variable corresponding to the factor(s) in question (ANNs, artificial neural networks) with backpropagation- (negative) modulation. The efflux of MAs with the resultant solubilization of Fe (III) chelates in the rhizosphere and transport-protein mediated divalent cation import would automatically constitute a self-limited closed loop by restoring cytosolic iron levels. A resilient framework (ANN with positive- or negative-feedforward moderation), however, would be dependent on *de novo* synthesis or conversely an exhaustible pool of compounds. These events might be expected to dominate proceedings at a later stage in the response. Major contributors within the plant 2OG-dependent dioxygenases conforming to this model are alkaloid biosynthesis, the free radical pathway, and prolyl 4-hydroxylases (**Figure 5**). Alkaloids are nitrogenous compounds produced by many organisms, and possess distinct pharmacological profiles. Analysis of DB2OG for probable alkaloid-synthesizing enzymes in plants suggests that the isoquinoline (thebaine, codeine), monoterpenoid indole (vinblastine, vincristine), and tropane

(hyoscyamine, scopoalmine) families of compounds could be deployed as a stable, local, and perishable pool of nutrients. (Matsuda et al., 1991; Vazquez-Flota et al., 1997; Hagel and Facchini, 2010a,b) (**Figure 8**). As mentioned earlier, oxygen, post-activation is a potent generator of free radicals, capable of scaling-up a response (Muller et al., 2009). Unchecked, these may cause an irreversible and unrecoverable loss of function. Besides the inter-radical neutralization, the upregulated poly-hydroxylated 2S-flavonols previously synthesized ($L3c \rightarrow L5b$; **Figure 4B**), could reprise their roles as powerful antioxidants (Kumar and Pandey, 2013).

The single-most critical end-point is the integrity of the cell wall. Interestingly, the local concentration of the principal components of both the primary and secondary cell wall can be assumed to be minimally supra-threshold, if not, borderline. This paradoxical vulnerability to biochemical factors is in contrast to the structural stability exhibited in transporting water and nutrients to various parts of the plant. Whilst, the sequestration of

the P4HY could potentially effect the PRPs in the primary form, the sub-optimal monolignol synthesizing pathway weakens the molecular framework of the secondary cell wall (Gallego-Giraldo et al., 2014). Endoperoxidation of the membranous unsaturated fatty acids, elevated transcription levels of expansion proteins (EXPs; Cho and Cosgrove, 2002) and hydrolytic enzymes (hydrolases/transglycosylases/mannoses; Carpita and Kanabus, 1988; Cui et al., 2005) complement these to dismantle the cell wall. Senescence triggered by ethylene too, is purported to be the result of low pH conditions and/or elevated expression of certain expansin proteins (Vreeburg et al., 2005), and its continued formation, in tandem with other factors imparts a predisposition to cell wall weakening.

CURRENT STATUS AND FUTURE DIRECTIONS

There is a vast body of literature highlighting individual enzymes of this superfamily, and is genomic, biochemical, and structural in character. This work underscores a novel treatment for 2-oxoglutarate-dependent dioxygenases, i.e., as a systems player and digresses, considerably from the single enzyme single function norm. From this analysis, there seems to clear indication that this superfamily possesses the necessary credentials to transcend their well-characterized roles as supercatalysts. The evolution of this collegial behavior, from single molecules to mediators of complex function, however remains unanswered. The redundancy exhibited, in terms of absolute numbers may contribute to the development of biological hysteresis, and thresholds, on one hand, as well as permit fine tuning of response pathways to neutralize/mitigate noxious stimuli. Whether, these bounds could be used as predictors for the emergence of a complex system needs to be investigated further.

It is generally accepted that the presence or absence of a few residues can markedly alter the catalytic profile of an enzyme, both, abrogating as well as accentuating activity. This minimalistic notion has numerous proponents and a detailed treatment of this nano-/femto-level stochasticity could offer insights into enzyme association/dissociation and the consequent kinetics. The creation of a carefully curated resource of enzymes with probable 2-oxoglutarate-dependency will aid workers in profiling new members of this remarkable superfamily.

AUTHOR CONTRIBUTIONS

SK collated the data, carried out the computational analysis, formulated and refined the models, constructed DB2OG and the GUI, wrote the code, and the manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpls.2015.00098/abstract

REFERENCES

Annunen, P., Autio-Harmainen, H., and Kivirikko, K. I. (1998). The novel type II prolyl 4-hydroxylase is the main enzyme form in chondrocytes and capillary endothelial cells, whereas the type I enzyme predominates in most cells. *J. Biol. Chem.* 273, 5989–5992. doi: 10.1074/jbc.273.11.5989

Annunen, P., Helaakoski, T., Myllyharju, J., Veijola, J., Pihlajaniemi, T., and Kivirikko, K. I. (1997). Cloning of the human prolyl 4-hydroxylase alpha subunit isoform alpha(II) and characterization of the type II enzyme tetramer. The alpha(I) and alpha(II) subunits do not form a mixed alpha(I)alpha(II)beta2 tetramer. *J. Biol. Chem.* 272, 17342–17348. doi: 10.1074/jbc.272.28.17342

Annunen, P., Koivunen, P., and Kivirikko, K. I. (1999). Cloning of the alpha subunit of prolyl 4-hydroxylase from Drosophila and expression and characterization of the corresponding enzyme tetramer with some unique properties. *J. Biol. Chem.* 274, 6790–6796. doi: 10.1074/jbc.274.10.6790

Blasiak, L. C., Vaillancourt, F. H., Walsh, C. T., and Drennan, C. L. (2006). Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* 440, 368–371. doi: 10.1038/nature04544

Burbulis, I. E., Iacobucci, M., and Shirley, B. W. (1996). A null mutation in the first enzyme of flavonoid biosynthesis does not affect male fertility in Arabidopsis. *Plant Cell* 8, 1013–1025. doi: 10.1105/tpc.8.6.1013

Carpita, N. C., and Kanabus, J. (1988). Chemical structure of the cell walls of dwarf maize and changes mediated by gibberellin. *Plant Physiol.* 88, 671–678. doi: 10.1104/pp.88.3.671

Cho, H. T., and Cosgrove, D. J. (2002). Regulation of root hair initiation and expansin gene expression in Arabidopsis. *Plant Cell* 14, 3237–3253. doi: 10.1105/tpc.006437

Chrispeels, M. J. (1984). Prolyl hydroxylase in plants. *Meth. Enzymol.* 107, 361–369. doi: 10.1016/0076-6879(84)07024-5

Clifton, I. J., McDonough, M. A., Ehrismann, D., Kershaw, N. J., Granatino, N., and Schofield, C. J. (2006). Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J. Inorg. Biochem.* 100, 644–669. doi: 10.1016/j.jinorgbio.2006.01.024

Cui, D., Neill, S. J., Tang, Z., and Cai, W. (2005). Gibberellin-regulated XET is differentially induced by auxin in rice leaf sheath bases during gravitropic bending. *J. Exp. Bot.* 56, 1327–1334. doi: 10.1093/jxb/eri133

De Jong, L., and Kemp, A. (1984). Stoicheiometry and kinetics of the prolyl 4-hydroxylase partial reaction. *Biochim. Biophys. Acta.* 787, 105–111. doi: 10.1016/0167-4838(84)90113-4

Flagg, S. C., Martin, C. B., Taabazuing, C. Y., Holmes, B. E., and Knapp, M. J. (2012). Screening chelating inhibitors of HIF-prolyl hydroxylase domain 2 (PHD2) and factor inhibiting HIF (FIH). *J. Inorg. Biochem.* 113, 25–30. doi: 10.1016/j.jinorgbio.2012.03.002

Gallego-Giraldo, L., Bhattarai, K., Pislariu, C. I., Nakashima, J., Jikumaru, Y., Kamiya, Y., et al. (2014). Lignin modification leads to increased nodule numbers in alfalfa. *Plant Physiol.* 164, 1139–1150. doi: 10.1104/pp.113.232421

Gille, C., and Frommel, C. (2001). STRAP: editor for STRuctural alignments of proteins. *Bioinformatics* 17, 377–378. doi: 10.1093/bioinformatics/17.4.377

Hagel, J. M., and Facchini, P. J. (2010a). Biochemistry and occurrence of o-demethylation in plant metabolism. *Front. Physiol.* 1:14. doi: 10.3389/fphys.2010.00014

Hagel, J. M., and Facchini, P. J. (2010b). Dioxygenases catalyze the O-demethylation steps of morphine biosynthesis in opium poppy. *Nat. Chem. Biol.* 6, 273–275. doi: 10.1038/nchembio.317

Hausinger, R. P. (2004). FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit. Rev. Biochem. Mol. Biol.* 39, 21–68. doi: 10.1080/10409230490440541

Hedden, P., and Phillips, A. L. (2000a). Manipulation of hormone biosynthetic genes in transgenic plants. *Curr. Opin. Biotechnol.* 11, 130–137. doi: 10.1016/S0958-1669(00)00071-9

Hedden, P., and Phillips, A. L. (2000b). Gibberellin metabolism: new insights revealed by the genes. *Trends Plant Sci.* 5, 523–530. doi: 10.1016/S1360-1385(00)01790-8

Hegg, E. L., and Que, L. Jr. (1997). The 2-His-1-carboxylate facial triad–an emerging structural motif in mononuclear non-heme iron(II) enzymes. *Eur. J. Biochem. FEBS.* 250, 625–629. doi: 10.1111/j.1432-1033.1997.t01-1-00625.x

Helaakoski, T., Annunen, P., Vuori, K., MacNeil, I. A., Pihlajaniemi, T., and Kivirikko, K. I. (1995). Cloning, baculovirus expression, and characterization of a second mouse prolyl 4-hydroxylase alpha-subunit isoform: formation of an

alpha 2 beta 2 tetramer with the protein disulfide-isomerase/beta subunit. *Proc. Natl. Acad. Sci. U.S.A.* 92, 4427–4431. doi: 10.1073/pnas.92.10.4427

Hewitson, K. S., McNeill, L. A., Riordan, M. V., Tian, Y. M., Bullock, A. N., Welford, R. W., et al. (2002). Hypoxia-inducible factor (HIF) asparagine hydroxylase is identical to factor inhibiting HIF (FIH) and is related to the cupin structural family. *J. Biol. Chem.* 277, 26351–26355. doi: 10.1074/jbc.C200273200

Hieta, R., Kukkola, L., Permi, P., Pirila, P., Kivirikko, K. I., Kilpelainen, I., et al. (2003). The peptide-substrate-binding domain of human collagen prolyl 4-hydroxylases. Backbone assignments, secondary structure, and binding of proline-rich peptides. *J. Biol. Chem.* 278, 34966–34974. doi: 10.1074/jbc.M303624200

Hijazi, M., Roujol, D., Nguyen-Kim, H., Del Rocio Cisneros Castillo, L., Saland, E., Jamet, E., et al. (2014). Arabinogalactan protein 31 (AGP31), a putative network-forming protein in Arabidopsis thaliana cell walls? *Ann. Bot.* 114, 1087–1097. doi: 10.1093/aob/mcu038

Hirsila, M., Koivunen, P., Gunzler, V., Kivirikko, K. I., and Myllyharju, J. (2003). Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. *J. Biol. Chem.* 278, 30772–30780. doi: 10.1074/jbc.M304982200

Hoffmann, L., Besseau, S., Geoffroy, P., Ritzenthaler, C., Meyer, D., Lapierre, C., et al. (2004). Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *Plant Cell* 16, 1446–1465. doi: 10.1105/tpc.020297

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259

Hutton, J. J., Tappel, A. L., and Udenfried, S. (1967). Cofactor and substrate requirements of collagen proline hydroxylase. *Arch. Biochem. Biophys.* 118, 231–240. doi: 10.1016/0003-9861(67)90302-5

Kawai, Y., Ono, E., and Mizutani, M. (2014). Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants. *Plant J.* 78, 328–343. doi: 10.1111/tpj.12479

Kersteen, E. A., Higgin, J. J., and Raines, R. T. (2004). Production of human prolyl 4-hydroxylase in *Escherichia coli*. *Protein Expr. Purif.* 38, 279–291. doi: 10.1016/j.pep.2004.09.008

Kivirikko, K. I., and Myllyla, R. (1980). "The hydroxylation of prolyl and lysyl residues," Enzymology and Post-Translational Modification of Proteins, Vol. 1, eds R. B. Freedman, H. C. Hawkins (New York, NY: Academic press), 53–104.

Kivirikko, K. I., and Myllyla, R. (1982). Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. Methods in enzymology. 82(Pt A), 245–304. doi: 10.1016/0076-6879(82)82067-3

Kivirikko, K. I., Myllyla, R., and Pihlajaniemi, T. (1989). Protein hydroxylation: prolyl 4-hydroxylase, an enzyme with four cosubstrates and a multifunctional subunit. *FASEB J.* 3, 1609–1617.

Koivunen, P., Hirsila, M., Gunzler, V., Kivirikko, K. I., and Myllyharju, J. (2004). Catalytic properties of the asparaginyl hydroxylase (FIH) in the oxygen sensing pathway are distinct from those of its prolyl 4-hydroxylases. *J. Biol. Chem.* 279, 9899–9904. doi: 10.1074/jbc.M312254200

Kosugi, Y., Oyamada, N., Satoh, S., Yoshioka, T., Onodera, E., and Yamada, Y. (1997). Inhibition by 1-aminocyclobutane-1-carboxylate of the activity of 1-aminocyclopropane-1-carboxylate oxidase obtained from senescing petals of carnation (*Dianthus caryophyllus* L.) flowers. *Plant Cell Physiol.* 38, 312–318. doi: 10.1093/oxfordjournals.pcp.a029168

Kukkola, L., Hieta, R., Kivirikko, K. I., and Myllyharju, J. (2003). Identification and characterization of a third human, rat, and mouse collagen prolyl 4-hydroxylase isoenzyme. *J. Biol. Chem.* 278, 47685–47693. doi: 10.1074/jbc.M306806200

Kumar, S., and Pandey, A. K. (2013). Chemistry and biological activities of flavonoids: an overview. *ScientificWorldJournal*. 2013:162750. doi: 10.1155/2013/162750

Kundu, S. (2012). Distribution and prediction of catalytic domains in 2-oxoglutarate dependent dioxygenases. *BMC Res. Notes*.5:410. doi: 10.1186/1756-0500-5-410

Kwak, S. S., Kamiya, Y., Sakurai, A., Takahashi, N., and Graebe, J. (1988). Partial purification and characterization of gibberellins 3beta-hydroxylase from immature seeds of Phaseolus vulgaris. *Plant Cell Physiol.* 29, 935–943.

Lamberg, A., Pihlajaniemi, T., and Kivirikko, K. I. (1995). Site-directed mutagenesis of the alpha subunit of human prolyl 4-hydroxylase. Identification of three histidine residues critical for catalytic activity. *J. Biol. Chem.* 270, 9926–9931. doi: 10.1074/jbc.270.17.9926

Matsuda, J., Okabe, S., Hashimoto, T., and Yamada, Y. (1991). Molecular cloning of hyoscyamine 6 beta-hydroxylase, a 2-oxoglutarate-dependent dioxygenase, from cultured roots of *Hyoscyamus niger*. *J. Biol. Chem.* 266, 9460–9464.

Mori, H., Shibasaki, T., Yano, K., and Ozaki, A. (1997). Purification and cloning of a proline 3-hydroxylase, a novel enzyme which hydroxylates free L-proline to cis-3-hydroxy-L-proline. *J. Bacteriol.* 179, 5677–5683.

Muller, K., Linkies, A., Vreeburg, R. A., Fry, S. C., Krieger-Liszkay, A., and Leubner-Metzger, G. (2009). *In vivo* cell wall loosening by hydroxyl radicals during cress seed germination and elongation growth. *Plant Physiol.* 150, 1855–1865. doi: 10.1104/pp.109.139204

Myllyharju, J. (2003). Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol.* 22, 15–24. doi: 10.1016/S0945-053X(03)00006-4

Myllyharju, J. (2008). Prolyl 4-hydroxylases, key enzymes in the synthesis of collagens and regulation of the response to hypoxia, and their roles as treatment targets. *Ann. Med.*40, 402–417. doi: 10.1080/07853890801986594

Nagahama, K., Ogawa,T., Fujii, T., Tazaki, M., Tanase, S., Morino, Y., et al. (1991). Purification and properties of an ethylene-forming enzyme from Pseudomonas syringae pv. phaseolicola PK2. *J. Gen. Microbiol.* 137, 2281–2286. doi: 10.1099/00221287-137-10-2281

Nakanishi, H., Yamaguchi, H., Sasakuma, T., Nishizawa, N. K., and Mori, S. (2000). Two dioxygenase genes, Ids3 and Ids2, from *Hordeum vulgare* are involved in the biosynthesis of mugineic acid family phytosiderophores. *Plant Mol. Biol.* 44, 199–207. doi: 10.1023/A:1006491521586

Nozoye, T., Nagasaka, S., Kobayashi, T., Takahashi, M., Sato, Y., Sato, Y., et al. (2011). Phytosiderophore efflux transporters are crucial for iron acquisition in graminaceous plants. *J. Biol. Chem.* 286, 5446–5454. doi: 10.1074/jbc.M110.180026

Price, J. C., Barr, E. W., Glass, T. E., Krebs, C., and Bollinger, J. M. Jr. (2003b). Evidence for hydrogen abstraction from C1 of taurine by the high-spin Fe(IV) intermediate detected during oxygen activation by taurine:alpha-ketoglutarate dioxygenase (TauD). *J. Am. Chem. Soc.* 125, 13008–13009. doi: 10.1021/ja037400h

Price, J. C., Barr, E. W., Tirupati, B., Bollinger, J. M. Jr., and Krebs, C. (2003a). The first direct characterization of a high-valent iron intermediate in the reaction of an alpha-ketoglutarate-dependent dioxygenase: a high-spin FeIV complex in taurine/alpha-ketoglutarate dioxygenase (TauD) from *Escherichia coli*. *Biochemistry* 42, 7497–7508. doi: 10.1021/bi030011f

Rocklin, A. M., Kato, K., Liu, H. W., Que, L. Jr., and Lipscomb, J. D. (2004). Mechanistic studies of 1-aminocyclopropane-1-carboxylic acid oxidase: single turnover reaction. *J. Biol. Inorg. Chem.* 9, 171–182. doi: 10.1007/s00775-003-0510-3

Schomburg, I., Chang, A., and Schomburg, D. (2002). BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.* 30, 47–49. doi: 10.1093/nar/30.1.47

Shibasaki, T., Mori, H., and Ozaki, A. (2000). Cloning of an isozyme of proline 3-hydroxylase and its purification from recombinant *Escherichia coli*. *Biotechnol. Lett.* 22, 1967–1973. doi: 10.1023/A:1026792430742

Shibasaki, T., Sakurai, W., Hasegawa, A., Uosaki, Y., Mori, H., Yoshida, M., et al. (1999). Substrate selectivities of proline hydroxylases. *Tetrahedron Lett.* 40, 5227–5230. doi: 10.1016/S0040-4039(99)00944-2

Smith, V. A., and MacMillan, J. (1984).Purification and partial characterization of a gibberellin 2beta-hydroxylase from Phaseolus vulgaris. *J. Plant Growth Regul.* 2, 251–264. doi: 10.1007/BF02042254

Sonnhammer, E. L., Eddy, S. R., and Durbin, R. (1997). Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* 28, 405–420.

Tanaka, M., Shibata, H., and Uchida, T. (1980). A new prolyl hydroxylase acting on poly-L-proline, from suspension cultured cells of *Vinca rosea*. *Biochim. Biophys. Acta* 616, 188–198. doi: 10.1016/0005-2744(80)90137-0

Tiainen, P., Pasanen, A., Sormunen, R., and Myllyharju, J. (2008). Characterization of recombinant human prolyl 3-hydroxylase isoenzyme 2, an enzyme modifying the basement membrane collagen IV. *J. Biol. Chem.* 283, 19432–19439. doi: 10.1074/jbc.M802973200

Tryggvason, K., Majamaa, K., Risteli, J., and Kivirikko, K. I. (1979). Partial purification and characterization of chick-embryo prolyl 3-hydroxylase. *Biochem. J.* 183, 303–307.

Vazquez-Flota, F., De Carolis, E., Alarco, A. M., and De Luca, V. (1997). Molecular cloning and characterization of desacetoxyvindoline-4-hydroxylase,

a 2-oxoglutarate dependent-dioxygenase involved in the biosynthesis of vin-doline in Catharanthus roseus (L.) G. *Don. Plant Mol. Biol.* 34, 935–948. doi: 10.1023/A:1005894001516

Vreeburg, R. A., Benschop, J. J., Peeters, A. J., Colmer, T. D., Ammerlaan, A. H., Staal, M., et al. (2005). Ethylene regulates fast apoplastic acidification and expansin A transcription during submergence-induced petiole elonga-tion in Rumex palustris. *Plant J.* 43, 597–610. doi: 10.1111/j.1365-313X.2005. 02477.x

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., et al. (2010). PSORTb 3.0: improved protein subcellular localization predic-tion with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26, 1608–1615. doi: 10.1093/bioinformatics/ btq249

Zhang, Z., Ren, J. S., Clifton, I. J., and Schofield, C. J. (2004). Crystal structure and mechanistic implications of 1-aminocyclopropane-1-carboxylic acid oxidase–the ethylene-forming enzyme. *Chem. Biol.* 11, 1383–1394. doi: 10.1016/j.chembiol.2004.08.012