BMC Research Notes

RESEARCH NOTE

Open Access

Fe(2)OG: an integrated HMM profile-based web server to predict and analyze putative non-haem iron(II)- and 2-oxoglutarate-dependent dioxygenase function in protein sequences

Siddhartha Kundu*

Abstract

Objective: Non-haem iron(II)- and 2-oxoglutarate-dependent dioxygenases (i2OGdd), are a taxonomically and functionally diverse group of enzymes. The active site comprises ferrous iron in a hexa-coordinated distorted octahedron with the apoenzyme, 2-oxoglutarate and a displaceable water molecule. Current information on novel i2OGdd members is sparse and relies on computationally-derived annotation schema. The dissimilar amino acid composition and variable active site geometry thereof, results in differing reaction chemistries amongst i2OGdd members. An additional need of researchers is a curated list of sequences with putative i2OGdd function which can be probed further for empirical data.

Results: This work reports the implementation of *Fe*(2)*OG*, a web server with dual functionality and an extension of previous work on i2OGdd enzymes (*Fe*(2)*OG* ≡ {*H2OGpred*, *DB2OG*}). *Fe*(2)*OG*, in this form is completely revised, updated (URL, scripts, repository) and will strengthen the knowledge base of investigators on i2OGdd biochemistry and function. *Fe*(2)*OG*, utilizes the superior predictive propensity of HMM-profiles of laboratory validated i2OGdd members to predict probable active site geometries in user-defined protein sequences. *Fe*(2)*OG*, also provides researchers with a pre-compiled list of analyzed and searchable i2OGdd-like sequences, many of which may be clinically relevant. *Fe*(2)*OG*, is freely available (http://204.152.217.16/Fe2OG.html) and supersedes all previous versions, i.e., H2OGpred, DB2OG.

Keywords: Facial triad, Hidden markov model, Non-haem iron(II)- and 2-oxoglutarate-dependent dioxygenases

Introduction

Dioxygenases, unlike monooxygenases are oxidoreductases which can incorporate both atoms of molecular oxygen, one each into a substrate and co-substrate (Fig. 1). These enzymes are classified on the basis of a metal co-factor (iron, cobalt, nickel, copper) and the presence of a haem-prosthetic group. Iron-based dioxygenases are mononuclear (extradiol catechol, *EC*1.13.11, x; 2-oxoglutarate-dependent, *EC*1.14.11, x), possess Rieske clusters (naphthalene 1,2-dioxygenase, *EC*1.14.12, 12) and may utilize haem (indoleamine 2,3-dioxygenase, *EC*1.13.11, 52; tryptophan 2,3-dioxygenase, *EC*1.13.11, 11) [1–5]. Extradiol and 2OG-dependent dioxygenases, possess a triad of catalytically competent $HX_n[DE]X_nH$ residues and comprise one face of a distorted octahedral co-ordination

*Correspondence: siddhartha_kundu@yahoo.co.in
Department of Biochemistry, All India Institute of Medical Sciences, Ansari Nagar, New Delhi 110029, India

© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

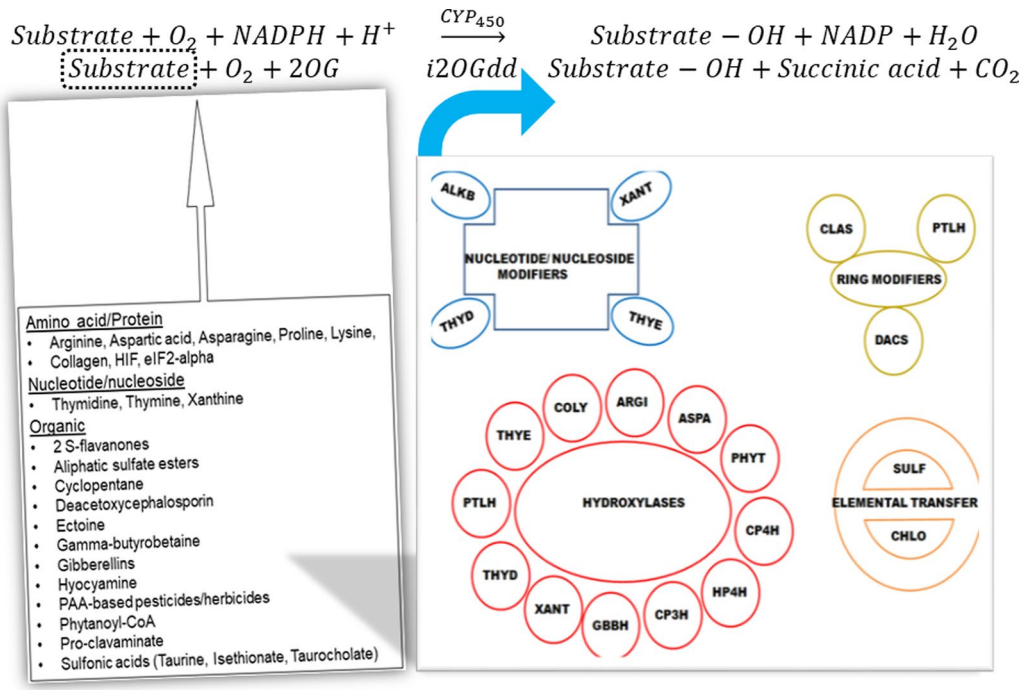

Fig. 1 Non-haem iron(II)- and 2-oxoglutarate-dependent dioxygenase based catalysis. This large superfamily of enzymes is characterized by a variable reaction chemistry, broad spectrum of substrates and is present in all kingdoms of life. The enzymes are characterized by a triad of Histidine, Aspartic-/Glutamic-acid and Histidine residues which co-ordinate ferrous iron. 2OG and the substrate. Enzymes are classified on the basis of the substrate(s) transformed and dominant reaction chemistry. Each cluster is a HMM-profile of at least two members and is derived from enzymes with available empirical data (structure, kinetic, mRNA expression). *ALKB* Alk-B like demethylase, *ARGI* Arginine hydroxylase; *ASPA* Aspartyl:Asparaginyl hydroxylase, *CHLO* Chlorinating enzyme, *CLAS* Clavaminate synthase; *COLY* Collagen lysyl dioxygenase, *CP3H* Collagen prolyl 3-hydroxylase, *CP4H* Collagen prolyl 4-hydroxylase, *CYCL* Cyclization and ring closure, *DACS* Deacetoxycephalosporin-C synthase, *DSAT* Desaturases, *ECTO* Ectoine hydroxylase, *FLAV* 2S-Flavones, *GBBH* γ-butyrobetaine hydroxylase; *GIAC* Gibberellic acid modification, *HILY* Histone lysyl demethylase, *HP4H* Hypoxia prolyl 4-hydroxylase, *HYOS* Hyoscyamine hydroxylase, *NUHY* Nucleotide/nucleoside hydroxylase, *OGFD* Eukaryotic initiation factor 2α, *PHYT* Phytanoyl-CoA hydroxylase, *PTLH* 1-Deoxypentalenic acid 11β-hydroxylase, *SULF* Sulfate cleaving, *TFDA* 2,4-Diphenoxyacetic acid metabolizing, *THYD* Thymdine dioxygenase, *THYE* Thymine dioxygenase, *XANT* Xanthine hydroxylase

sphere with iron(II) [3–6]. The other face is formed by three displaceable water molecules, a factor that contributes significantly to the architecture of the active site [3–6]. The subset that comprises non-haem iron(II)- and 2OG-dependent dioxygenases (i2OGdd) is characterized by $HX[DE]X_nH(n \in [50, 120])$, in a jelly-roll or double stranded beta-helical fold [4–6]. The i2OGdd-superfamily ($EC1.14.11.x$), is characterized by the dependence of catalysis on 2-oxoglutarate and occurs by the bi-dentate co-ordination of $C1$ (carboxylic acid) and $C2$ (2-oxo/alpha-keto) with mononuclear ferrous iron (Fig. 1). The last dative bond is the cognate substrate after the solitary water molecule has been displaced. The reaction chemistry involves a progressive increase in the oxidation state of iron ($II \rightarrow IV$) and is followed by proton abstraction and the formation of a substrate-radical. This in turn leads to catalytic conversion of the substrate by the incorporation of a single oxygen atom (Fig. 1). The transformation is therefore, a de facto oxidative hydroxylation although, this is accompanied in most cases by a concomitant desaturation, cyclization, stereo-isomerization and sulfate cleavage (Fig. 1) [7–9]. The second oxygen atom is incorporated into 2OG with the release of succinic acid and CO_2 (Fig. 1). Members of the i2OGdd-superfamily participate in cell signaling under hypoxic conditions, DNA repair, stress response mechanisms, metabolism (lipids, growth factors) and biodegradation of herbicides (Fig. 1) [5, 10–17].

The work presented revises, updates and integrates the functionality of two servers, i.e., $Fe(2)OG \equiv \{H2OGpred, DB2OG\}$ [18, 19]. $Fe(2)OG$, can be used by researchers as a single-point web resource to screen protein sequence(s) for potential i2OGdd-activity and shortlist putative i2OGdd members from the available pre-compiled sequence repository. The latter is searchable on the basis of taxonomy, cellular compartment and HMM-profiles of the sequences. A novel feature of $Fe(2)OG$ is the inclusion of clinically relevant non-haem iron(II)- and 2OG-dependent dioxygenases. This includes links and preliminary analyses to several

human putative i2OGdd members. The coding and interfacing is done using in-house developed PERL scripts.

Main text

Rationale for incorporating empirical data into a profile-based search application

Non-haem iron(II)- and 2OG-dependent-dioxygenases are characterized by variable reaction chemistry and a broad spectrum of substrates. The reverse mapping of substrate descriptors to the active site of known enzymes is well documented and can be utilized to repurpose pharmacological agents. Several theoretically sound statistical tools such as multi-class support vector machines (SVMs), artificial neural networks (ANNs), and hidden markov models (HMMs) have been utilized to garner insights into the active site geometry of an enzyme in the presence of a pharmacophore [18–22]. Although HMMs, as a predictive modality are non-committal, this can be rectified by mathematical filters. The transformed output can then be utilized by clustering algorithms and ANNs to generate unambiguous predictors [21, 23, 24]. In fact, a rigorously derived integrated HMM-ANN algorithm has been presented and used to characterize sequences which are few and closely related such as those from an enzyme family or sub family [23, 24].

Mathematical basis for the algorithms deployed by *Fe*(2)*OG*

Whilst, a detailed description of the computational pipeline deployed and its relevance has already been published, the mathematical basis for these has not been addressed [18, 19]. Briefly, HMM-profiles of catalytically relevant clusters and laboratory validated enzymes of the i2OGdd-superfamily ($a_i \in A \subseteq \mathcal{H}$) are utilized to score regions of an amino acid sequence. The empirical data that is considered is the presence of one or more 3D-structures, kinetic and mutagenesis data and mRNA expression levels [18]. A suitable mathematical representation is as under:

$$\begin{aligned} A &= \{a_1, a_2, \ldots a_i | \#a_i \geq 2\} \\ &= \bigcup a_i \end{aligned} \tag{1}$$

Theorem: *A unique set of HMM profiles* $(A, B \subseteq \mathcal{H})$ *can exist iff there is at least one unique sub-profile.*

$$A \neq B \ \textit{iff} \ \#(a_i \cap b_i) \leq min(\#a_i, \#b_i) \ \ a_i \in A \subseteq \mathcal{H}, b_i \in B \subseteq \mathcal{H}, \left\{\#a_i, \#b_i\right\} \geq 2$$

Proof:

$$\begin{array}{rcl}
Case1 & if \cap_{i=1} a_i = \emptyset, \cap_{i=1} b_i = \emptyset & \\
SinceA, B \subseteq \mathcal{H}, & A \cap B \neq \emptyset & \\
Rewriting, & (\cup_{i=1} a_i) \cap (\cup_{i=1} b_i) \neq \emptyset & \\
Expanding, & (a_1 \cap b_1) \cup (a_2 \cap b_2) \ldots (a_i \cap b_i) \neq \emptyset & \\
Let, & (a_1 \cap b_1) \cup (a_2 \cap b_2) \ldots (a_i \cap b_i) = k & \\
 & k \in (a_1 \cap b_1) OR k \in (a_2 \cap b_2) \ldots OR k \in (a_i \cap b_i) & \\
Clearly, & k = a_i \cup b_i & \\
Generalizing, & \exists (a_i \cap b_i) | \#(a_i \cap b_i) \leq min(\#a_i, \#b_i) & (a) \\
Case2 & if \cap_{i=1} a_i \neq \emptyset, \cap_{i=1} b_i \neq \emptyset & \\
Let, & \cap_{i=1} a_i = a_i \in A | a_i \geq 2, \cap_{i=1} b_i = b_i \in B | b_i \geq 2 & \\
itfollows, & \#(a_i \cap b_i) \leq min(\#a_i, \#b_i) & (b) \\
 & & \\
From (a), (b) & A \neq B & \\
 & & \\
Conversely, & if A \neq B, & \\
SinceA, B \subseteq \mathcal{H}, & A \cap B \neq \emptyset & \\
Let, & A \cap B = a_i \cup b_i & \\
\Rightarrow & \exists (a_i \cap b_i) \in A \cap B | \#(a_i \cap b_i) \leq min(\#a_i, \#b_i) & \\
 & a_i \neq b_i &
\end{array}$$

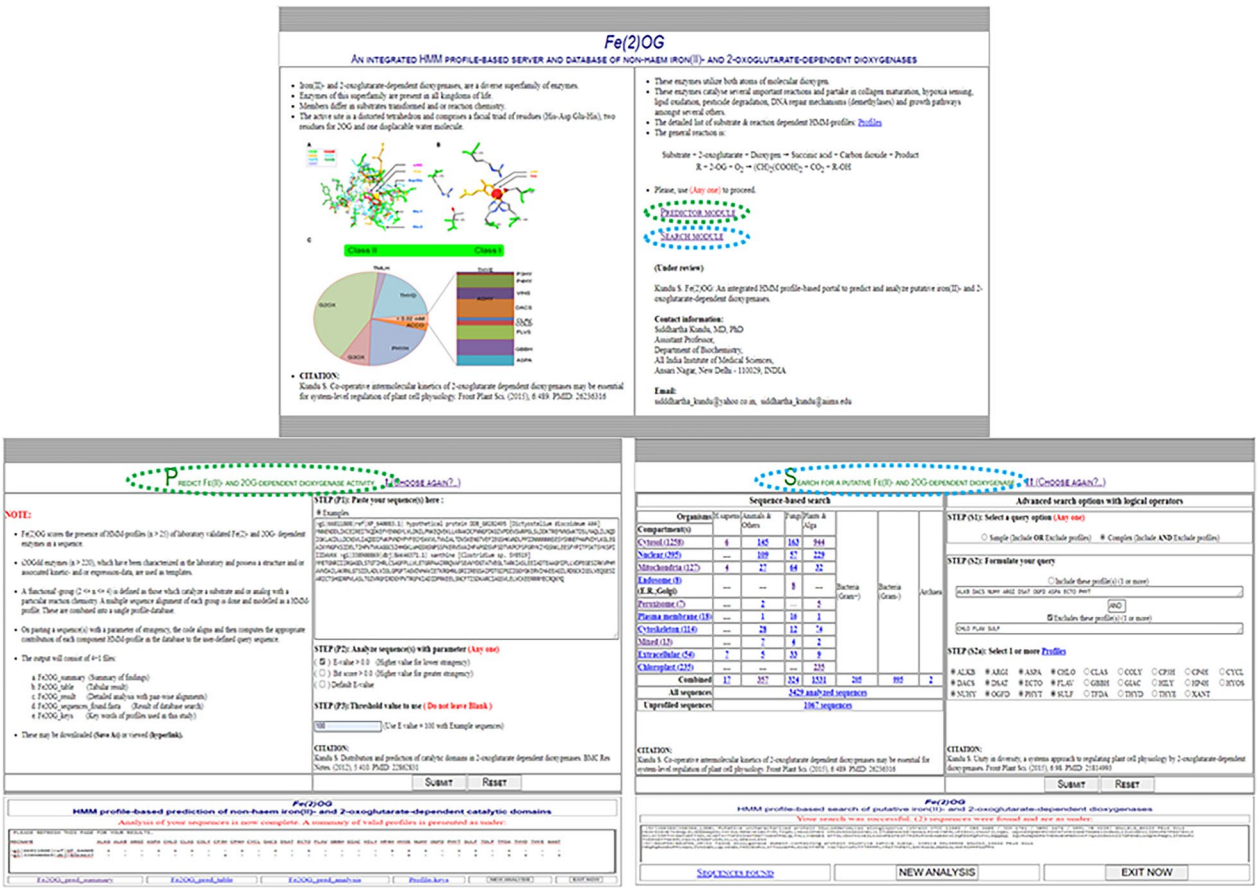

Fig. 2 *Fe*(2)*OG*, an integrated HMM-profile based web server of non-haem iron(II)- and 2OG-dependent dioxygenases. *Fe*(2)*OG*, is a revised and updated web resource which completely integrates the functionality of two independent servers, *H2OGpred* and *DB2OG*. A brief introduction with regards to active site geometry and reaction chemistry is provided. Clear directives on the usage, precautions and processing of data are provided. The predictor and search modules are distinct from each other and can be submitted independently of each other. The predictor module of *Fe*(2)*OG* is modified and will analyze a user-defined sequence(s) for the complete set of HMM-profiles. Users may use the "Examples" option to add sample sequences or "Paste" their own selection in the text area. These, when submitted will result in a chart of closely-matched HMM-profiles across the full length of the sequence. Since, the output is based on optimally chosen parameters (*Evalue*, *Bitscore*), these need to be manually added and carefully chosen before submission. The database, too, is updated and has been extensively curated. Here, putative i2OGdd members are classified and arranged on the basis of their taxonomy and dominant cellular compartment. These are arranged in an easy-to-select matrix format. The sequences of this database can also be queried and recursively analyzed using the logical operators (AND, OR) for the presence of combinations of HMM-profiles. The results of these searches can be downloaded and investigated further. The list of sequences not analyzed is also presented along with links to several human putative i2OGdd members. Abbreviations: 2OG, 2-oxoglutarate; HMM, Hidden Markov Models; i2OGdd, non-haem iron(II)- and 2OG-dependent dioxygenases; *H2OGpred* HMM-based prediction of putative 2-oxoglutarate function; *DB2OG* database of sequences that results from a generic-2OG HMM-based query of UniprotKB

Fe(2)*OG*, then, is an implementation of a particular instance of the combined HMM of sequences and available structures $(A = a_i | 1 \leq i \leq 28, 2 \leq \#a_i \leq 4)$ [19]; URL-http://janelia.org. The lower limit of number of the sequences in each profile ($\min(\#a_i)$) Eq. (1) is implied by definition. The upper limit, however, is estimated as a proportion of the total number of sequences,

$$\max(\#a_i) = \frac{\#\mathcal{H}}{\#A} \tag{2}$$

Description and utilization of *Fe*(2)*OG*

i) Fe(2)*OG, a predictor of the catalytic spectrum of an unknown or single function enzyme*

The algorithm and code that *Fe*(2)*OG* utilizes to predict the dominant profile, in a user-defined sequence(s), has been described in detail [18]. Briefly, i2OGdd enzymes ($n > 220$) with available empirical data (structure, kinetic, mRNA expression) are clustered on the basis of the substrates catalyzed and/or the reaction chemistry (Fig. 1) [18]. The enzymes present in each

Table 1 Comparative analysis and biomedical relevance of non-haem iron(II)- and 2OG-dependent dioxygenases

A		*Fe*(2)*OG*	*H2OGpred*, *DB2OG*
	Nature and composition	Integrated webserver	Independent servers
	Mode of prediction and impact	Generic, improved and extendible	Specific; restricted
	Putative i2OGdd sequences	Recent and relevant	NA
	Display and accessibility	Matrix format (compartments, taxonomy)	Simple tabular format
	"Unclassified" sequences	Yes; Amenable to further investigation	No
	"All sequences" search criteria	Omitted; extended profiles of sequences with known function	NA
	Clinically relevant members	Yes	No

B	Enzymes (*EC*1.14.11.x)	Physiological role; Disease biology		Profile(s); References
1	Phytanoyl-CoA hydroxylase ($x = 18$)	Phytanic acid hydroxylation;	Refsum's disease	PHYT; [18, 27]
2	(HIF) Prolyl hydroxylases ($x = 2, 29$)	Hypoxia-induced Proline hydroxylations	Tumor suppression, ubiquitin-mediated proteasomal degradation of HIF via the Von Hippel-Lindau complex	HP3H, HP4H; [13, 14, 18, 28]
3	(HIF) Asparaginyl hydroxylases ($x = 16, 30$)	Hypoxia-induced Aspartic acid/Asparagine hydroxylations	Traboulsi's syndrome	ASPA; [18, 29, 30]
4	Collagen Prolyl hydroxylases ($x = 2, 28$)	Assembly of mature collagen	Connective tissue disorders, promoter of metastasis	CP3H, CP4H, OGFD; [18, 31, 32]
5	Procollagen-lysine 5-dioxygenase ($x = 4$)		Kyphoscoliotic Ehlers–Danlos syndrome Bruck syndrome	COLY; [18, 33, 34]
6	JMJD6 (Jumonji-domain containing protein) ($x = 66, 67$) Histone H3 trimethyl L lysine (4|9) demethylase Histone H3 dimethyl L lysine 36 demethylase ($x = 27$)	Arginine demethylation, Lysine hydroxylation Lysine demethylation	Promoters of metastasis, Developmental disorders,	HILY; [18, 35, 36]
7	DNA demethylases ALKBH-1, 2, 3, 4, 5, 8 ($x = 33, 51, 53, 54$)	DNA/mRNA-repair after alkylation	Heightened propensity for malignant transformation, disorders of growth and development Decreased atherosclerosis	ALKB; [18, 37, 38]

'functional'-group, ($2 \leq \#a_i \leq 4$) Eqs. (1) and (2) are then aligned and assigned a HMM-profile (Figs. 1 and 2) [18]. A database of these HMM-profiles is used to probe the catalytic spectrum of a user-defined sequence as per the stringency specified. Unlike *H2OGpred*, *Fe*(2)*OG*, compares a query sequence(s) with all, rather than isolated HMM-profiles (Fig. 2) [18]. The rationale for this alteration is that since the catalytic profile of an unknown sequence(s) is debatable, a generic analysis rather than a specific one is a better indicator of i2OGdd-like activity. Furthermore, sequences with known function can also be investigated for other reaction chemistries. Clearly, in both cases the analysis with individual profiles is superfluous and may be omitted (Table 1A). The tabulated list of relevant cognate substrates, for each profile is also available and may be used as a reference (Figs. 1 and 2). In addition, to the overt directives of use, users can also sample the functionality of *Fe*(2)*OG* by clicking the "Examples" button (*StepP*1) (Fig. 2). This loads bonafide i2OGdd sequences into the text area which can be analyzed in accordance with the steps that are outlined subsequently. These include choice of threshold parameter (*Evalue*, *Bitscore*) and assignment of a suitable numerical value (*StepsP*2, *P*3) (Fig. 2). The output comprises a tabular summary of suitably matched profiles with detailed statistics and exhaustive pair-wise alignments of all supra-threshold matches (Fig. 2). Since, *Fe*(2)*OG* has dual functionality, the user can submit this independently (*StepsP*1 − *P*3 → *Submit*) (Fig. 2).

ii) Fe(2)*OG, a repository of i2OGdd-like sequences*

The second component of *Fe*(2)*OG* is a flat-file database. This comprises a pre-compiled and updated list of i2OGdd-like sequences ($n_{AB} = 4496$) (Fig. 2). This is accomplished by constructing a generic-HMM after combining representative ($n \sim 80$) i2OGdd enzymes from each 'functional'-group. This is then used to query UniprotKB for probable matches (n_{AB}) [19]. The downloaded sequences are analyzed and assigned a dominant cellular compartment (n_A) [19]. Sequences, which are not amenable to these preliminary investigations are annotated as such (n_B). Users can download updated lists of these sequences ($n_A = 3429, n_B = 1067$) (Fig. 2). This is facilitated by arranging the sequences as a matrix of compartments (p) and taxonomy

$(q)\left(AB = \left\{y_{pqr} \in \left(ab_{pq}\right); p = 10, q = 7, r \in \mathbb{N}\right\}\right)$. *Fe*(2)*OG*, also uses the logical operators ($\{AND, OR\}$) to formulate an advanced HMM profile-based query to partition the sequences (*StepS*1; Fig. 2) [19]. Another modification introduced in *Fe*(2)*OG* is the omission of the "All sequences"-option (*StepS*1) (Fig. 2). The rationale for this amendment, is that users may require sequences specific to one or more HMM-profiles (Figs. 1 and 2). Since, each profile is based on a specific reaction chemistry, users will also possess, a priori, a definitive list of probable ligands to characterize the kinetics of their search result with (Fig. 1, Table 1A). Furthermore, the entire database (n_A) is accessible with the "OR" and "Include these profile(s)", if the user so chooses (*StepsS*1, *S*2) (Fig. 2). The other fraction could not be further classified and is presented only in terms of their respective taxonomies ($n_B = 1067$). Here, too, the user can submit this independently (*StepsS*1, *S*2 $\rightarrow$ *Submit*) (Fig. 2).

Comparative analysis and biomedical relevance of *Fe*(2)*OG*

Despite the similarity in algorithms and general usage, *Fe*(2)*OG*, offers several new and upgraded features (Table 1). These include links to i2OGdd members which are uncharacterized and clinically relevant, whilst offering researchers a tool to extend the catalytic profiles of known enzymes. Additionally, the list of sequences with putative i2OGdd function is updated and non-redundant. The i2OGdd are amongst the largest group of non-haem dioxygenases and can arguably compete in importance with the more established cytochrome P450 (CYP_{450}) superfamily of haem monooxygenases (Fig. 1). The differential activity of i2OGdd members in response to fluctuating concentrations of oxygen and iron also suggest a system-level function in sensing and thence regulating the uptake, utilization and release of these micronutrients [25, 26]. In fact, clinical data is available for several i2OGdd enzymes. This includes phytanoyl-CoA hydroxylase, hypoxia-inducible Proline hydroxylases, collagen modifiers (Proline- and Lysine-hydroxylases) and DNA/mRNA-demethylases (Table 1B) [27–38]. The analysis by *Fe*(2)*OG* results in a small subset ($\approx 24\%, n = 17$) of enzymes and are grouped into mitochondrial, cytosolic and extracellular fractions (Additional file 1: Text S1a). However, a larger proportion ($\approx 76\%, n = 53$) remains unclassified and merits a deeper investigation (Additional file 1: Text S1b).

Limitations

Fe(2)*OG*, is an online web resource that is dedicated to expanding the knowledge base of non-haem iron(II)- and 2OG-dependent-dioxygenase superfamily of enzymes amongst scientists and clinicians. *Fe*(2)*OG*, can predict whether an unknown protein sequence(s) possesses i2OGdd-activity. It also provides preliminary analyses (taxonomy, cellular compartment) and an analytic tool (sequence-based, logical) to shortlist enzyme candidates from a pre-compiled list of sequences. Since, newer sequences are constantly becoming available, *Fe*(2)*OG* will require constant updates to its core of HMM-profiles and the raw sequences that are queried for putative function, thereof, to remain relevant to the biomedical community. However, since this information is dependent on available empirical data, an annual update might suffice. *Fe*(2)*OG*, is also not exhaustive and lacks structural-models and simulation data for its members. These short comings will be addressed in future studies.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13104-021-05477-z.

Additional file 1: Text S1. a HMM profiles of analyzed human i2OGdd-like sequences (n = 17). **b** Uniprot IDs of unprofiled human i2OGdd-like sequences (n = 53)

Abbreviations

Fe: Iron; 2OG: 2-Oxoglutarate; HMM: Hidden Markov Model; i2OGdd: Non-haem iron(II)- and 2-oxoglutarate-dependent dioxygenases.

Acknowledgements

Not Applicable.

Authors' contributions

SK outlined and designed the study, designed and conceptualized the algorithm(s) and formulae for prediction, wrote the mathematical proofs, manually collated all the sequences, and their references, carried out the computational analysis, wrote all the code and the manuscript. All authors read and approved the final manuscript.

Funding

This work is funded by an early career intramural grant awarded to SK (Code A-766) by the All India Institute of Medical Sciences (AIIMS, New Delhi, INDIA).

Availability of data and materials

Data is available as supporting material with the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Yes.

Competing interests

The authors declare that they have no competing interests.

Received: 11 July 2020 Accepted: 3 February 2021
Published online: 01 March 2021

References

1. Koehntop KD, Emerson JP, Que L Jr. The 2-His-1-carboxylate facial triad: a versatile platform for dioxygen activation by mononuclear non-heme iron(II) enzymes. J Biol Inorg Chem. 2005;10(2):87–93.

2. Barry SM, Challis GL. Mechanism and catalytic diversity of rieske non-heme iron-dependent oxygenases. ACS Catal. 2013. https://doi.org/10.1021/cs400087p.
3. Lipscomb JD. Mechanism of extradiol aromatic ring-cleaving dioxygenases. Curr Opin Struct Biol. 2008;18(6):644–9.
4. Islam MS, Leissing TM, Chowdhury R, Hopkinson RJ, Schofield CJ. 2-oxoglutarate-dependent oxygenases. Annu Rev Biochem. 2018;87:585–620.
5. McDonough MA, Loenarz C, Chowdhury R, Clifton IJ, Schofield CJ. Structural studies on human 2-oxoglutarate dependent oxygenases. Curr Opin Struct Biol. 2010;20(6):659–72.
6. Martinez S, Hausinger RP. Catalytic mechanisms of Fe(II)- and 2-oxoglutarate-dependent oxygenases. J Biol Chem. 2015;290(34):20702–11.
7. Clifton IJ, Doan LX, Sleeman MC, Topf M, Suzuki H, Wilmouth RC, et al. Crystal structure of carbapenem synthase (CarC). J Biol Chem. 2003;278(23):20843–50.
8. Eichhorn E, van der Ploeg JR, Kertesz MA, Leisinger T. Characterization of alpha-ketoglutarate-dependent taurine dioxygenase from *Escherichia coli*. J Biol Chem. 1997;272(37):23031–6.
9. Janc JW, Egan LA, Townsend CA. Purification and characterization of clavaminate synthase from *Streptomyces antibioticus*. A multifunctional enzyme of clavam biosynthesis. J Biol Chem. 1995;270(10):5399–404.
10. Fukumori F, Hausinger RP. Purification and characterization of 2,4-dichlorophenoxyacetate/alpha-ketoglutarate dioxygenase. J Biol Chem. 1993;268(32):24311–7.
11. Holland PJ, Hollis T. Structural and mutational analysis of Escherichia coli AlkB provides insight into substrate specificity and DNA damage searching. PLoS ONE. 2010;5(1):e8680.
12. Kershaw NJ, Mukherji M, MacKinnon CH, Claridge TD, Odell B, Wierzbicki AS, et al. Studies on phytanoyl-CoA 2-hydroxylase and synthesis of phytanoyl-coenzyme A. Bioorg Med Chem Lett. 2001;11(18):2545–8.
13. Koivunen P, Hirsila M, Gunzler V, Kivirikko KI, Myllyharju J. Catalytic properties of the asparaginyl hydroxylase (FIH) in the oxygen sensing pathway are distinct from those of its prolyl 4-hydroxylases. J Biol Chem. 2004;279(11):9899–904.
14. Bruick RK, McKnight SL. A conserved family of prolyl-4-hydroxylases that modify HIF. Science. 2001;294(5545):1337–40.
15. Lester DR, Phillips A, Hedden P, Andersson I. Purification and kinetic studies of recombinant gibberellin dioxygenases. BMC Plant Biol. 2005;5:19.
16. Reuter K, Pittelkow M, Bursy J, Heine A, Craan T, Bremer E. Synthesis of 5-hydroxyectoine from ectoine: crystal structure of the non-heme iron(II) and 2-oxoglutarate-dependent dioxygenase EctD. PLoS ONE. 2010;5(5):e10647.
17. Wehner KA, Schutz S, Sarnow P. OGFOD1, a novel modulator of eukaryotic translation initiation factor 2alpha phosphorylation and the cellular response to stress. Mol Cell Biol. 2010;30(8):2006–16.
18. Kundu S. Distribution and prediction of catalytic domains in 2-oxoglutarate dependent dioxygenases. BMC Res Notes 5. 2012. https://doi.org/10.1186/1756-0500-5-410.
19. Kundu S. Unity in diversity, a systems approach to regulating plant cell physiology by 2-oxoglutarate-dependent dioxygenases. Front Plant Sci. 2015;6:98.
20. Khater S, Mohanty D. In silico identification of AMPylating enzymes and study of their divergent evolution. Sci Rep. 2015;5:10804.
21. Kundu S, Sharma R. In silico identification and taxonomic distribution of plant class C GH9 endoglucanases. Front Plant Sci. 2016;7:1185.
22. Kundu S. Insights into the mechanism(s) of digestion of crystalline cellulose by plant class C GH9 endoglucanases. J Mol Model. 2019;25(8):240.
23. Kundu S. Mathematical basis of predicting dominant function in protein sequences by a generic HMM-ANN algorithm. Acta Biotheor. 2018;66(2):135–48.
24. Kundu S. Mathematical basis of improved protein subfamily classification by a HMM-based sequence filter. Math Biosci. 2017;293:75–80.
25. Wilson JW, Shakir D, Batie M, Frost M, Rocha S. Oxygen-sensing mechanisms in cells. FEBS J. 2020. https://doi.org/10.1111/febs.15374.
26. Kundu S. Co-operative intermolecular kinetics of 2-oxoglutarate dependent dioxygenases may be essential for system-level regulation of plant cell physiology. Front Plant Sci. 2015;6:489.
27. Mukherji M, Chien W, Kershaw NJ, Clifton IJ, Schofield CJ, Wierzbicki AS, et al. Structure-function analysis of phytanoyl-CoA 2-hydroxylase mutations causing Refsum's disease. Hum Mol Genet. 2001;10(18):1971–82.
28. Hirsila M, Koivunen P, Gunzler V, Kivirikko KI, Myllyharju J. Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. J Biol Chem. 2003;278(33):30772–80.
29. Patel N, Khan AO, Mansour A, Mohamed JY, Al-Assiri A, Haddad R, et al. Mutations in ASPH cause facial dysmorphism, lens dislocation, anterior-segment abnormalities, and spontaneous filtering blebs, or Traboulsi syndrome. Am J Hum Genet. 2014;94(5):755–9.
30. Pfeffer I, Brewitz L, Krojer T, Jensen SA, Kochan GT, Kershaw NJ, et al. Aspartate/asparagine-beta-hydroxylase crystal structures reveal an unexpected epidermal growth factor-like domain substrate disulfide pattern. Nat Commun. 2019;10(1):4910.
31. Guo H, Tong P, Liu Y, Xia L, Wang T, Tian Q, et al. Mutations of P4HA2 encoding prolyl 4-hydroxylase 2 are associated with nonsyndromic high myopia. Genet Med. 2015;17(4):300–6.
32. Kim JH, Lee SM, Lee JH, Chun S, Kang BH, Kwak S, et al. OGFOD1 is required for breast cancer cell proliferation and is associated with poor prognosis in breast cancer. Oncotarget. 2015;6(23):19528–41.
33. Pirskanen A, Kaimio AM, Myllyla R, Kivirikko KI. Site-directed mutagenesis of human lysyl hydroxylase expressed in insect cells. Identification of histidine residues and an aspartic acid residue critical for catalytic activity. J Biol Chem. 1996;271(16):9398–402.
34. Abdalla EM, Rohrbach M, Burer C, Kraenzlin M, El-Tayeby H, Elbelbesy MF, et al. Kyphoscoliotic type of Ehlers-Danlos Syndrome (EDS VIA) in six Egyptian patients presenting with a homogeneous clinical phenotype. Eur J Pediatr. 2015;174(1):105–12.
35. Webby CJ, Wolf A, Gromak N, Dreger M, Kramer H, Kessler B, et al. Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. Science. 2009;325(5936):90–3.
36. Chang B, Chen Y, Zhao Y, Bruick RK. JMJD6 is a histone arginine demethylase. Science. 2007;318(5849):444–7.
37. Gerken T, Girard CA, Tung YC, Webby CJ, Saudek V, Hewitson KS, et al. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. Science. 2007;318(5855):1469–72.
38. Daoud H, Zhang D, McMurray F, Yu A, Luco SM, Vanstone J, et al. Identification of a pathogenic FTO mutation by next-generation sequencing in a newborn with growth retardation and developmental delay. J Med Genet. 2016;53(3):200–7.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.