

# Iodotyrosine deiodinase: a unique flavoprotein present in organisms of diverse phyla<sup>†</sup>

Cite this: *Mol. BioSyst.*, 2014, 10, 86

Abhishek Phatarphekar,<sup>a</sup> Jennifer M. Buss<sup>b</sup> and Steven E. Rokita\*<sup>a</sup>

Iodide is required for thyroid hormone synthesis in mammals and other vertebrates. The role of both iodide and iodinated tyrosine derivatives is currently unknown in lower organisms, yet the presence of a key enzyme in iodide conservation, iodotyrosine deiodinase (IYD), is suggested by genomic data from a wide range of multicellular organisms as well as some bacteria. A representative set of these genes has now been expressed, and the resulting enzymes all catalyze reductive deiodination of diiodotyrosine with  $k_{cat}/K_m$  values within a single order of magnitude. This implies a physiological presence of iodotyrosines (or related halotyrosines) and a physiological role for their turnover. At least for Metazoa, IYD should provide a new marker for tracing the evolutionary development of iodinated amino acids as regulatory signals through the tree of life.

Received 11th September 2013,  
Accepted 17th October 2013

DOI: 10.1039/c3mb70398c

[www.rsc.org/molecularbiosystems](http://www.rsc.org/molecularbiosystems)

## Introduction

Reductive dehalogenation is rare in aerobic organisms and even more unusual in Metazoa.<sup>1,2</sup> Only two enzymes are known to catalyze reductive dehalogenation in mammals, and both are critical for thyroid function. Iodothyronine deiodinase (ID) catalyzes deiodination of thyroid hormone (TH) to regulate its physiological activity<sup>3</sup> while iodotyrosine deiodinase (IYD) catalyzes deiodination of mono- and diiodotyrosine (MIT and DIT) for iodide salvage from these byproducts of TH synthesis (Fig. 1).<sup>4</sup> Despite similarities in function, their origins are quite diverse. ID is a member of the thioredoxin structural superfamily<sup>5</sup> and promotes deiodination of thyroxine with help of an active site selenocysteine.<sup>6</sup> In contrast, IYD is a member of the nitro-FMN reductase superfamily (formerly referred to as the NADH oxidase/flavin reductase superfamily) and its catalysis relies on a flavin cofactor (FMN).<sup>7,8</sup>

IYD is unique in its recruitment of flavin to promote reductive dehalogenation in mammals.<sup>9</sup> Sequence analysis of mammalian IYD detected an N-terminal membrane anchor, an intermediate domain and a large C-terminal catalytic domain that is homologous to flavoproteins associated with reduction

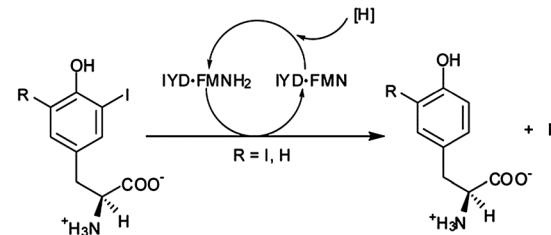


Fig. 1 Reductive deiodination of mono- and diiodotyrosine is promoted by iodotyrosine deiodinase (IYD) and its bound cofactor, flavin mononucleotide (FMN and, in its reduced form, FMNH<sub>2</sub>).

of nitroaromatics among other types of substrates.<sup>10</sup> The crystal structure of IYD<sup>11</sup> revealed most similarity to a bacterial flavin reductase (BluB) responsible for synthesis of the lower ligand of vitamin B<sub>12</sub>.<sup>12</sup> Together, IYD and BluB form a distinct subclass within their superfamily based on the position of their active site lid sequences.<sup>11,12</sup> The remaining two subclasses represented by NADH oxidase (NOX) from *Thermus thermophilus* and flavin reductase P (FRP) from *Vibrio harveyi* utilize different regions of their primary structure to form active site lids.<sup>13,14</sup> Emergence of a deiodinase from this collection of activities is far from obvious and made all the more interesting due to its unusual dependence on a flavin for dehalogenation and its contribution to iodide homeostasis and thyroid function in higher organisms. This enzyme also has the potential to serve as a marker to trace the origins of cell signaling by iodinated compounds that ultimately culminated in the development of the mammalian thyroid.

A prerequisite for tracing IYD through the tree of life is a robust set of structural determinants. Automated sequence analysis will typically assign a protein to the correct superfamily

<sup>a</sup> Department of Chemistry, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD, 21218, USA. E-mail: rokita@jhu.edu; Fax: +1-410-516-8420; Tel: +1-410-516-5793

<sup>b</sup> Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA

† Electronic supplementary information (ESI) available: NCBI accession numbers for IYD homologs chosen for expression; optimized conditions for expressing IYD homologs; SDS-PAGE gel images of purified IYD homologs; identification of active IYD from hydra (hmIYD); deiodination rate dependence on DIT concentration for IYD homologs. See DOI: 10.1039/c3mb70398c



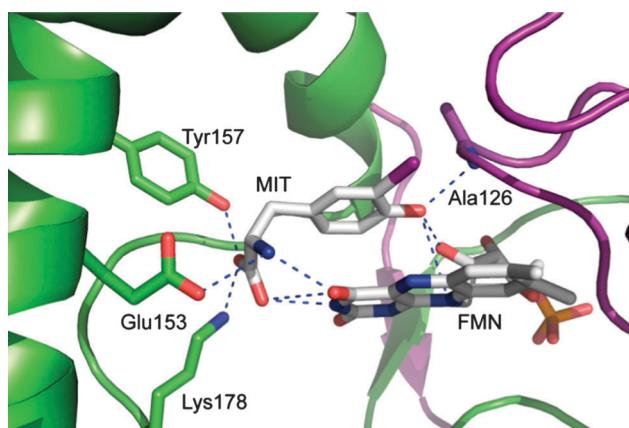


Fig. 2 Active site region of iodotyrosine deiodinase from mouse is established by both subunits of its  $\alpha_2$  dimer (green and purple) and contains flavin (FMN), monoiodotyrosine (MIT) and the key residues stabilizing their complex (PDB ID 3GFD).

but is much less reliable in predicting its function.<sup>15</sup> Additions to the nitro-FMN reductase superfamily are often broadly annotated as nitroreductases without regard to their potential catalytic diversity. The co-crystal structure of mouse (mm) IYD and MIT identified key residues that are expected to be diagnostic of deiodinase activity and not required by other activities within the superfamily.<sup>11</sup> The side chains of E153, Y157, and K178 interact directly with the zwitterion portion of MIT and appear critical for function (Fig. 2).<sup>16</sup> Although the mutation Y157F only diminished the  $k_{cat}/K_m$  of mmIYD for turnover of DIT by  $\sim 35\%$ , the affinity for MIT was suppressed by 20-fold.<sup>16</sup> More dramatic was the complete loss of catalytic activity and substrate affinity for the mutation E153Q.<sup>16</sup> Comparable analysis of K178G was not possible since the resulting protein could not be expressed in a soluble form.

The amide nitrogen of A126 provides yet another interaction to the substrate of IYD by coordinating to its phenolic OH group. BluB contains a Gly at an equivalent site (G61) but lacks all key residues responsible for coordinating MIT.<sup>12</sup> Another distinction is the T235 in IYD and the S167 in BluB that provide a side chain hydrogen bond to the O<sup>4</sup> and N<sup>5</sup> of FMN.<sup>11,12</sup> No equivalent activation of FMN is evident in the remaining members of the superfamily. As described below, the residues corresponding to A126, E153, Y157, K178 and T235 of mmIYD now accurately predict deiodinase activity (Fig. 3). The presence of IYD extends well beyond initial expectations and includes diverse organisms not previously associated with a need for iodide conservation.

## Methods

### Cloning of homologous genes

IYD homologs from zebrafish, lancelet and honeybee were truncated to remove their N-terminal membrane domains predicted by TMHMM (ESI,† Table S1).<sup>17</sup> Full length sequences were used for IYD homologs from daphnia, sea anemone, hydra and the bacterium *Haliscomenobacter hydrossis* (DSM 1100).

Each gene included a C-terminal His<sub>6</sub> tag and flanking 5' BamHI and 3' XhoI restriction sites. Genes were synthesized (Blue Heron Biotechnology and GenScript) and subcloned to create SUMO-fusions in the pSMT3 plasmid<sup>18</sup> kindly provided, along with the expression vector for the SUMO-specific protease Ulp1, by Dr C. Lima (Memorial Sloan-Kettering Institute).

### Expression of IYD homologs

Rosetta 2(DE3) *E. coli* cells were transformed with the appropriate pSMT3 plasmid and grown in LB media containing kanamycin and chloramphenicol at 37 °C with shaking to an OD<sub>600</sub> of 0.6–0.8. The media was then cooled and supplemented with isopropyl-β-D-thiogalacto-pyranoside as summarized in ESI,† Table S2. Cells were harvested, resuspended in lysis buffer (500 mM NaCl, 0.1 mM DTT, 10% glycerol and 50 mM sodium phosphate pH 8.0) and lysed by three passages through a French press cell at 1000 psi. The lysates were centrifuged at 40 000g for 2 h.

### Purification of IYD homologs

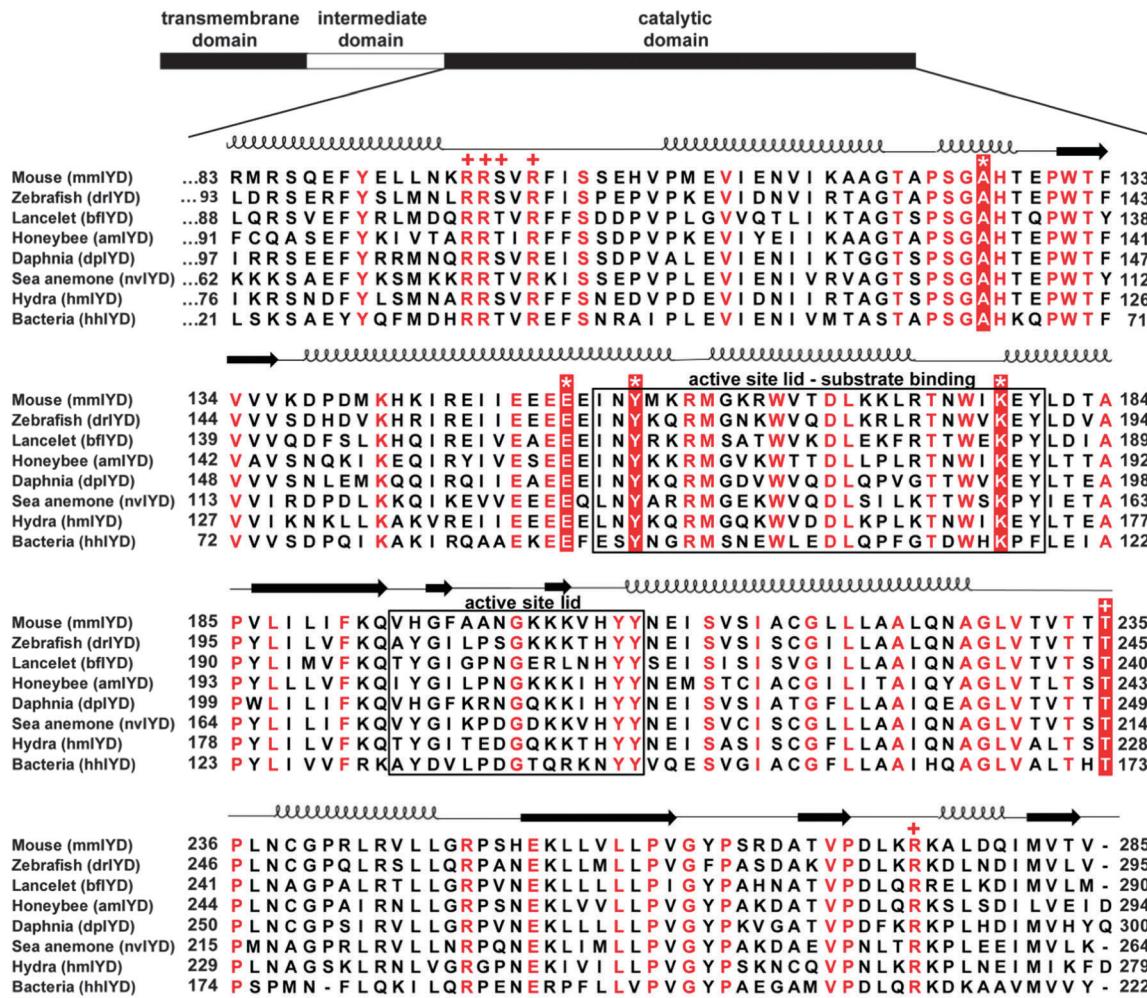
SUMO-IYD fusions were purified by Ni<sup>2+</sup> affinity chromatography using HisPur Ni-NTA resin. Supernatant was loaded and washed with 5 column volumes of lysis buffer containing increasing concentrations of imidazole (20 mM, 60 mM and 100 mM). SUMO-IYD fusions were eluted with 250 mM imidazole and treated with Ulp1 (1:200 w/w). Proteins from zebrafish (drIYD), lancelet (bfiIYD), honeybee (amiIYD) and *H. hydrossis* (hhIYD) were further purified using a 1 ml HiTrap Ni<sup>2+</sup> column. This was washed successively with 20 mM imidazole (10 ml), 60 mM imidazole (10 ml) and 100 mM imidazole (5 ml) in lysis buffer. The proteins were eluted with 250 mM imidazole in lysis buffer. Homologs from sea anemone (nvIYD), daphnia (dpIYD) and hydra (hmIYD) were dialyzed with 300 mM NaCl, 1 mM DTT, 10% glycerol and 50 mM sodium phosphate pH 8.0 and further purified by size exclusion chromatography using Sephadryl S-200 HR resin. A third affinity step using a HiTrap Ni<sup>2+</sup> column as described above was necessary for dpiIYD.

The amino acid sequence of the active dpiIYD fragment was identified by N-terminal sequencing (JHMI sequencing and synthesis facility). The extinction coefficients ( $\epsilon_{280}$ ) for the IYD homologs were determined from their amino acid sequence using ExPASy ProtParam.<sup>19</sup> The concentration of FMN bound to enzyme was determined from  $A_{450}$  ( $\epsilon_{450}$  12 500 M<sup>-1</sup> cm<sup>-1</sup>),<sup>20</sup> and the concentration of enzyme was determined from its  $A_{280}$  after correcting for FMN absorbance using  $A_{280}/A_{450}$  of 1.57.

### Estimation of enzyme concentration for hmIYD

Western blotting was used to confirm the presence of hmIYD in isolates exhibiting deiodinase activity (ESI,† Fig. S2). Total protein concentration was determined by the BCA assay per manufacturer's protocol (Pierce), and the percentage of hmIYD was determined by densitometric analysis of a Coomassie-stained SDS-PAGE gel using ImageQuant TL. The concentration of hmIYD was estimated from its average MW (33.5 kDa) observed in ESI,† Fig. S1B and S2B.





**Fig. 3** Multiple sequence alignment of the catalytic domains of IYD homologs. Alignment of homologs chosen for protein expression was generated using MUSCLE.<sup>45</sup> Numbering of the amino acids for each protein is indicated on the left and right of the alignment. Residues in red (or white) are fully conserved, and the sequences forming the active site lids are indicated with a box. Key residues coordinating to substrate and FMN are indicated with an (\*) and (+), respectively. Secondary structure elements are derived from the crystal structure of mmIYD bound to DIT (PDB ID 3GH8).<sup>11</sup>

### Assay for catalytic activity

The rate of [<sup>125</sup>I]-iodide release from [<sup>125</sup>I]-DIT was determined by a discontinuous assay reported previously.<sup>7,21</sup> Values of  $k_{cat}$  were calculated based on protein concentration as opposed to FMN occupancy since excess FMN is present in the activity assay. BlubB (gift of Dr M. E. Taga, UC Berkeley) was tested for native activity by a FMN depletion assay<sup>22</sup> in the presence of FMN reductase (SsuE, gift of Dr H. R. Ellis, Auburn University).

## Results

### Selection and sequence analysis of IYD homologs

Using the sequence of mmIYD as reference, BLAST<sup>23</sup> identified homologous proteins within the nitro-FMN reductase superfamily from a wide range of organisms. Most sequences within the animal kingdom were annotated as deiodinases, but some from invertebrates lacked annotation, and most from bacteria were annotated as nitroreductases or simply as oxidoreductases. To test the predictive powers of the IYD determinants described

above, genes from organisms representing distinct phyla and classes within Metazoa as well as one from Eubacteria were chosen for expression in *E. coli* (Fig. 4). The specific examples were selected based on a variety of criteria including an absence of domains superfluous to IYD activity and a minimal number of cysteines to avoid potential problems with protein folding. Organisms that have become representative of a phylum or class were also chosen when possible. Until very recently, few potential genes for IYD had been identified in bacteria, but a candidate from *H. hydrossis* seemed intriguing since this species is commonly found in waste water after sewage treatment.<sup>24</sup> The selected homologs all contained the conserved residues of the active site lid region that distinguishes the IYD subgroup within the superfamily and those critical for substrate and FMN coordination (Fig. 3).<sup>11,16</sup>

### Expression and purification of IYD homologs

Candidate proteins were expressed as fusions with SUMO to enhance their solubility. The final yields of drIYD, amIYD,



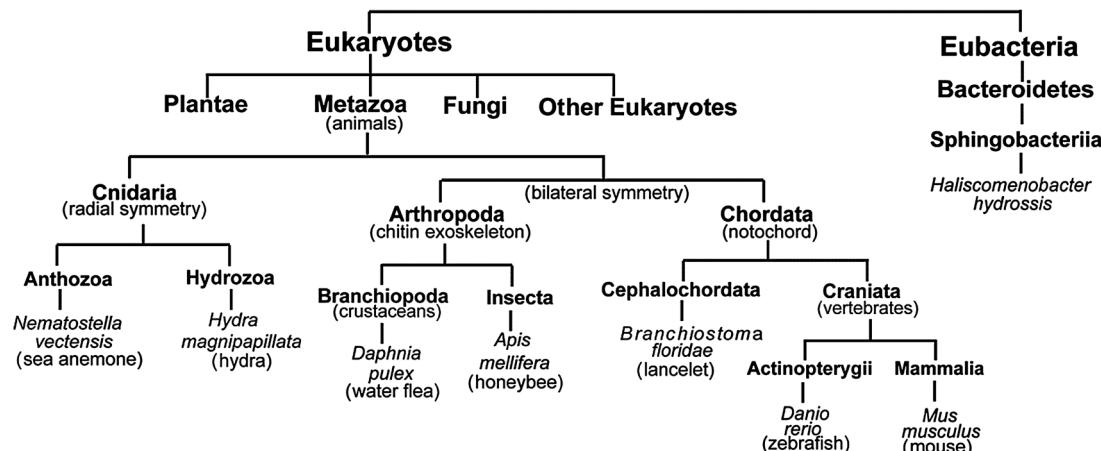


Fig. 4 The tree of life. A tree indicating the diversity of organisms in which IYD homologs were identified and chosen for expression. Phylum and class for each organism were obtained from the NCBI taxonomy browser.<sup>46</sup> Branches are not drawn to scale.

nvIYD and hhIYD after purification ranged from 3 to 5 mg L<sup>-1</sup> of growth media. Expression of bfiIYD, dpIYD and hmIYD primarily formed inclusion bodies even under optimized conditions (ESI,† Table S2), and yields of active enzyme were consequently less than 1 mg L<sup>-1</sup> of growth media. The active fraction isolated for dpIYD was found to be truncated *in situ* by 55 amino acids from the N-terminus. Multiple proteins were evident after separation of the soluble fraction of SUMO-hmIYD by Ni<sup>2+</sup> affinity chromatography. As evident from densitometry, only ~55% of the soluble protein was susceptible to Ulp1, the protease used to remove the SUMO fusion. This is consistent with misfolding of the expressed protein. Further purification using size exclusion chromatography yielded an active fraction containing multiple proteins including those derived from SUMO-hmIYD and host proteins as identified by Western blot and Coomassie staining (ESI,† Fig. S2). A truncated SUMO-hmIYD comprising *ca.* 15% of the protein was detected in the active fraction by scanning densitometry. Difficulties in hmIYD expression likely arose from use of sequence information that has since been revised in the NCBI database. The updated version differs in the first 42 amino acids. All other IYD homologs were purified to homogeneity (ESI,† Fig. S1), and most contained FMN at 95% occupancy. Only nvIYD and hmIYD had lower FMN occupancies of 55% and 59% respectively.

### Catalytic activity

All homologs showed deiodinase activity with DIT as a substrate. Their  $k_{\text{cat}}/K_m$  values were clustered between *ca.* 0.3 min<sup>-1</sup> μM<sup>-1</sup> to 3.5 min<sup>-1</sup> μM<sup>-1</sup> (Table 1). Their  $K_m$  values ranged from 4 μM to 105 μM and  $k_{\text{cat}}$  values ranged from 4.1 min<sup>-1</sup> to 32 min<sup>-1</sup>. The sea anemone enzyme (nvIYD) generated the highest  $K_m$  and  $k_{\text{cat}}$  values, but the daphnia and hydra enzymes (dpIYD and hmIYD, respectively) generated the highest  $k_{\text{cat}}/K_m$  values. In contrast, the closest structural homolog of IYD (BluB) expressed no deiodinase activity above the limit of detection at  $K_m \geq 133 \mu\text{M}$ ,  $k_{\text{cat}} \leq 0.06 \text{ min}^{-1}$  and a  $k_{\text{cat}}/K_m \leq 5 \times 10^{-4} \text{ min}^{-1} \mu\text{M}^{-1}$ . Likewise, no BluB-like activity was previously observed for mmIYD after its reduced form was exposed to molecular

Table 1 Iodotyrosine deiodinase activity of homologs

Organism	$K_m$ <sup>c</sup> (μM)	$k_{\text{cat}}$ <sup>c</sup> (min <sup>-1</sup> )	$k_{\text{cat}}/K_m$ <sup>c</sup> (min <sup>-1</sup> μM <sup>-1</sup> )
Mouse (mmIYD) <sup>a</sup>	19 ± 3	6.9 ± 1.3	0.36 ± 0.09
Zebrafish (drIYD)	8 ± 1	4.1 ± 0.2	0.51 ± 0.07
Lancelet (bfiIYD)	6 ± 3	7 ± 1	1.2 ± 0.6
Honeybee (amiIYD)	29 ± 7	8 ± 1	0.28 ± 0.07
Daphnia (dpIYD)	7 ± 1	17.5 ± 0.8	2.5 ± 0.4
Sea anemone (nvIYD)	105 ± 26	32 ± 4	0.30 ± 0.08
Hydra (hmIYD) <sup>b</sup>	4 ± 1	14 ± 1	3.5 ± 0.9
Bacteria (hhIYD)	6.6 ± 0.9	5.4 ± 0.3	0.8 ± 0.1

<sup>a</sup> Parameters for mmIYD were previously reported.<sup>16</sup> <sup>b</sup> Parameters  $k_{\text{cat}}$  and  $k_{\text{cat}}/K_m$  calculated from estimated enzyme concentration (see experimental procedures). <sup>c</sup> Significant figures were determined from the magnitude of associated uncertainties.

oxygen,<sup>25</sup> and physiological concentrations of reduced nicotinamides were not capable of driving efficient turnover of mmIYD as expected if it could express a flavin or nitroreductase activity.<sup>26</sup>

### Discussion

All of the tested homologs expressed iodotyrosine deiodinase activity with  $k_{\text{cat}}/K_m$  values narrowly distributed over a single order of magnitude, and thus the key residues selected from the crystal structure of mmIYD (Fig. 2) are excellent criteria for predicting deiodinase activity within the nitro-FMN reductase superfamily. Phylogenetic analysis of the catalytic domains of the deiodinases, nitro-FMN reductases and BluB (Fig. 5) suggests a divergent evolution from a common prokaryotic ancestor. BluB, the closest structural relative of IYD, did not exhibit deiodinase activity although BluB and IYD have active site lids that derive from the same region of primary sequence<sup>11,12</sup> and flavin chemistry that is poised for sequential one electron transfer processes.<sup>4</sup> The same lid regions are not evident in NOX and FRP and instead other regions of their primary sequence form their active site lids. Such differences serve to distinguish the subclasses of enzymes within the superfamily.<sup>11</sup> NOX and FRP are also distinct from the IYD and BluB subclasses



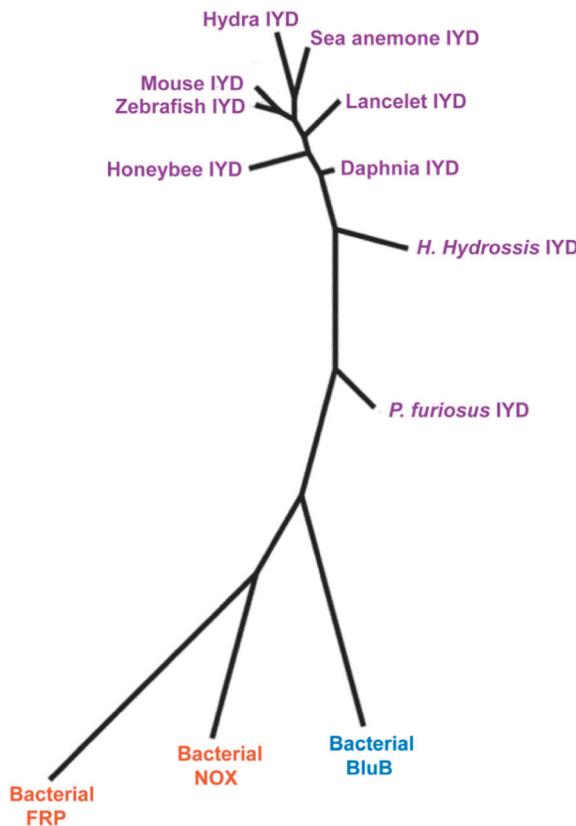


Fig. 5 Phylogenetic analysis focusing on the IYD branch of the nitro-FMN reductase superfamily. A phylogenetic tree was generated by phylogeny.fr<sup>47</sup> integrating MUSCLE<sup>45</sup> PhyML<sup>48</sup> and TreeDyn<sup>49</sup> for multiple sequence alignment, phylogenetic analysis and tree rendering, respectively. Catalytic domains of IYD homologs and aligned regions of BluB (PDB ID 2ISJ), NOX (PDB ID 1NOX) and FRP (PDB ID 2BKJ) were used as input sequences. The branch lengths are proportional to the number of amino acid substitutions per aligned residue.

by promoting substrate reduction through an obligate two electron process.<sup>10</sup>

The presence of IYD in Craniata (Fig. 4) is easily explained by the presence of a thyroid gland, its associated TH and the concomitant need to salvage the scarce nutrient iodide.<sup>27</sup> IYD and iodide salvage is also expected in the remaining branches of Chordata lacking the thyroid. For example, the genome of lancelet contains homologs of most genes required for generating TH and an endostyle that may represent a proto-thyroid.<sup>28</sup> The presence of deiodinase activity in Arthropoda is very intriguing since there is no confirmed evidence for endogenous synthesis of TH<sup>29</sup> although a limited number of reports suggest that TH can be obtained exogenously<sup>29,30</sup> and affect insect metabolism.<sup>31-33</sup> The requirements for iodide recycling and the effects of iodinated amino acids on insect development has yet to be determined.<sup>33</sup> IYD might even have an alternative role in catabolism of scleroproteins that often contain halogenated tyrosines.<sup>27</sup>

The physiological role of IYD in the cnidarians hydra and sea anemone is currently unknown, and no reports on their sensitivity to iodide, iodotyrosine or TH could be found.

However, these materials are all capable of initiating strobilation (metamorphosis from polyp to medusa) in another cnidarian, the jellyfish *Aurelia aurita*.<sup>34</sup> DIT was most effective in these studies and is consistent with a proposal that iodotyrosines could be evolutionary precursors of THs.<sup>35</sup> This presents a possibility that IYD may assume a regulatory role by its ability to degrade this primitive iodinated signal.

An equivalent function cannot be extrapolated to bacteria (e.g., *H. Hydrossis*) even though certain strains have been shown to accumulate iodine.<sup>36,37</sup> More likely, bacterial IYD serves a catabolic or detoxification role enabling the bacteria to adapt to environments containing iodinated compounds such as those created by algae blooms. Chlorinated and brominated derivatives may also be consumed by bacterial IYD since at least mmIYD is known to dehalogenate bromo- and chlorotyrosine.<sup>25</sup>

A recent BLAST search using stringencies lower than those of our original survey (including hits with expect values  $>10^{-59}$ ) revealed over 200 bacteria contain proteins within the nitro-FMN reductase superfamily that should now be classified as deiodinases according to the criteria established in this study. Some species even contain genes from multiple subclasses of the nitro-FMN reductase superfamily. For example, IYD and a FMN reductase are present in *Vibrio harveyi*, a model bacterium for bioluminescence and quorum sensing studies.<sup>38,39</sup> Similarly, IYD and a nitroreductase are present in *Deinococcus peraridilitoris*, a recent addition to the *Deinococcus* genus of bacteria resistant to ionizing radiation.<sup>40</sup>

A few candidate deiodinases are also evident in Archaea (e.g. *Pyrococcus furiosus*) although these consistently contain a Met at the site corresponding to A126 in mmIYD. Their presence may originate from lateral gene transfer between thermophilic bacteria that also contain the Ala to Met substitution.<sup>41</sup> In contrast, no deiodinase gene has yet been discovered in Fungi. Even *Neurospora crassa* that relies significantly on flavoproteins for metabolism<sup>42</sup> does not contain an IYD gene. The genes most related to IYD in fungi had very weak homologies as indicated by e values of greater than  $9 \times 10^{-6}$ . Greater homologies were evident between FRP and its homologues in fungi ( $e \geq 10^{-39}$ ). In contrast, no protein sequences from plants registered an e value of less than 0.2 when compared to IYD, BluB, NOX or FRP.

## Conclusion

The successful identification of a specialized enzyme such as IYD in diverse organisms demonstrates the importance of experimental data to support the functional assignment of a flavoprotein.<sup>42,43</sup> IYD revealed an origin that is much more ancient than would be anticipated by its iodide salvage activity in mammals and may indicate requirements for iodide or iodinated molecules in most multicellular organisms. The enzyme is now ready to join the ranks of other TH-associated proteins in describing the evolution of TH signaling.<sup>28,44</sup>



## Acknowledgements

We thank Drs Ellis, Lima and Taga for providing essential materials for this study, Dr Stephen Mount for many helpful discussions and support from NIH grant RO1 DK084186 to SER.

## References

- W. W. Mohn and J. M. Tiedje, *Microbiol. Rev.*, 1992, **56**, 482–507.
- J. Pandey, H. J. Heipieper, A. Chauhan, P. K. Arora, D. Prakash, M. Takeo and R. K. Jain, *Appl. Microbiol. Biotechnol.*, 2011, **92**, 597–607.
- D. L. St. Germain, V. A. Galton and A. Hernandez, *Endocrinology*, 2009, **150**, 1097–1107.
- S. E. Rokita, J. M. Adler, P. M. McTamney and J. A. Watson Jr., *Biochimie*, 2010, **92**, 1227–1235.
- I. Callebaut, C. Curcio-Morelli, J. P. Mornon, B. Gereben, C. Buettner, S. Huang, B. Castro, T. L. Fonseca, J. W. Harney, P. R. Larsen and A. C. Bianco, *J. Biol. Chem.*, 2003, **278**, 36887–36896.
- M. J. Berry, J. D. Kieffer, J. W. Harney and P. R. Larsen, *J. Biol. Chem.*, 1991, **266**, 14155–14158.
- J. E. Friedman, J. A. Watson, D. W. H. Lam and S. E. Rokita, *J. Biol. Chem.*, 2006, **281**, 2812–2819.
- A. Goswami and I. N. Rosenberg, *J. Biol. Chem.*, 1979, **254**, 12326–12330.
- S. E. Rokita, in *Handbook of Flavoproteins*, ed. R. Hille, S. M. Miller and B. Palfey, DeGruyter, Berlin, 2013, vol. 1, pp. 337–350.
- M. D. Roldán, E. Pérez-Reinado, F. Castillo and C. Moreno-Vivián, *FEMS Microbiol. Rev.*, 2008, **32**, 474–500.
- S. R. Thomas, P. M. McTamney, J. M. Adler, N. LaRonde-LeBlanc and S. E. Rokita, *J. Biol. Chem.*, 2009, **284**, 19659–19667.
- M. E. Taga, N. A. Larsen, A. R. Howard-Jones, C. T. Walsh and G. C. Walker, *Nature*, 2007, **446**, 449–453.
- H. J. Hecht, H. Erdmann, H. J. Park, M. Sprinzel and R. D. Schmid, *Nat. Struct. Biol.*, 1995, **2**, 1109–1114.
- J. J. Tanner, B. Lei, S. C. Tu and K. L. Krause, *Biochemistry*, 1996, **35**, 13531–13539.
- A. M. Schnoes, S. D. Brown, I. Dodevski and P. C. Babbitt, *PLoS Comput. Biol.*, 2009, **5**, e1000605.
- J. M. Buss, P. M. McTamney and S. E. Rokita, *Protein Sci.*, 2012, **21**, 351–361.
- A. Krogh, B. Larsson, G. von Heijne and E. L. L. Sonnhammer, *J. Mol. Biol.*, 2001, **303**, 567–580.
- E. Mossessova and C. D. Lima, *Mol. Cell*, 2000, **5**, 865–876.
- E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel and A. Bairoch, in *The Proteomics Protocols Handbook*, ed. J. M. Walker, Humana Press, Totowa, NJ, 2005, pp. 571–607.
- K. Jacek, *Methods Enzymol.*, 1971, **18**, 253–285.
- I. N. Rosenberg and A. Goswami, *Methods Enzymol.*, 1984, **107**, 488–500.
- T.-Y. Yu, K. C. Mok, K. J. Kennedy, J. Valton, K. S. Anderson, G. C. Walker and M. E. Taga, *Protein Sci.*, 2012, **21**, 839–849.
- S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucleic Acids Res.*, 1997, **25**, 3389–3402.
- S. M. Kotay, T. Datta, J. Choi and R. Goel, *Water Res.*, 2011, **45**, 694–704.
- P. M. McTamney and S. E. Rokita, *J. Am. Chem. Soc.*, 2009, **131**, 14212–14213.
- J. A. Watson Jr., P. M. McTamney, J. M. Adler and S. E. Rokita, *ChemBioChem*, 2008, **9**, 504–506.
- C. Valverde-R., A. Orozco, A. Becerra, M. C. Jeziorski, P. Villalobos and J. C. Solis-S., *Int. Rev. Cytol.*, 2004, **234**, 143–199.
- M. Paris, F. Brunet, G. V. Markov, M. Schubert and V. Laudet, *Dev. Genes Evol.*, 2008, **218**, 667–680.
- J. G. Eales, *Proc. Soc. Exp. Biol. Med.*, 1997, **214**, 302–317.
- A. Heyland and L. L. Moroz, *J. Exp. Biol.*, 2005, **208**, 4355–4361.
- B. S. Thyagaraja, E. P. Masler, T. J. Kelly and A. B. Borkovec, *Comp. Biochem. Physiol.*, 1993, **104A**, 247–253.
- K. G. Davey, *Insect Biochem. Mol. Biol.*, 2000, **30**, 877–884.
- T. Flatt, L. L. Moroz, M. Tatar and A. Heyland, *Integr. Comp. Biol.*, 2006, **46**, 777–794.
- M. Silverstone, V. A. Galton and S. H. Ingbar, *Gen. Comp. Endocrinol.*, 1978, **34**, 132–140.
- S. J. Crockford, *Integr. Comp. Biol.*, 2009, **49**, 155–166.
- S. Amachi, Y. Mishima, H. Shinoyama, Y. Muramatsu and T. Fujii, *Appl. Environ. Microbiol.*, 2005, **71**, 741–745.
- H.-P. Li, R. Brinkmeyer, W. L. Jones, S. Zhang, C. Xu, K. A. Schwehr, P. H. Santschi, D. I. Kaplan and C. M. Yeager, *Appl. Environ. Microbiol.*, 2011, **77**, 2153–2160.
- B. L. Bassler, M. Wright, R. E. Showalter and M. R. Silverman, *Mol. Microbiol.*, 1993, **9**, 773–786.
- M. G. Surette, M. B. Miller and B. L. Bassler, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 1639–1644.
- F. A. Rainey, M. Ferreira, M. F. Nobre, K. Ray, D. Bagaley, A. M. Earl, J. R. Battista, B. Gómez-Silva, C. P. McKay and M. S. da Costa, *Int. J. Syst. Evol. Microbiol.*, 2007, **57**, 1408–1412.
- K. E. Nelson, R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, O. White, S. L. Salzberg, H. O. Smith, J. C. Venter and C. M. Fraser, *Nature*, 1999, **399**, 323–329.
- P. Macheroux, B. Kappes and S. E. Ealick, *FEBS J.*, 2011, **278**, 2625–2634.
- J. Kim and S. D. Copley, *Biochemistry*, 2013, **52**, 2905–2913.
- A. Heyland, J. Hodin and A. M. Reitzel, *Bioessays*, 2005, **27**, 64–75.
- R. C. Edgar, *Nucleic Acids Res.*, 2004, **32**, 1792–1797.



46 E. W. Sayers, T. Barrett, D. A. Benson, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. DiCuccio, R. Edgar, S. Federhen, M. Feolo, L. Y. Geer, W. Helmberg, Y. Kapustin, D. Landsman, D. J. Lipman, T. L. Madden, D. R. Maglott, V. Miller, I. Mizrachi, J. Ostell, K. D. Pruitt, G. D. Schuler, E. Sequeira, S. T. Sherry, M. Shumway, K. Sirotnik, A. Souvorov, G. Starchenko, T. A. Tatusova, L. Wagner, E. Yaschenko and J. Ye, *Nucleic Acids Res.*, 2009, **37**, D5–D15.

47 A. Dereeper, V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J. F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J. M. Claverie and O. Gascuel, *Nucleic Acids Res.*, 2008, **36**, W465–W469.

48 S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk and O. Gascuel, *Syst. Biol.*, 2010, **59**, 307–321.

49 F. Chevenet, C. Brun, A.-L. Banuls, B. Jacq and R. Christen, *BMC Bioinf.*, 2006, **7**, 439.

