Applied Biochemistry and Biotechnology
https://doi.org/10.1007/s12010-022-04304-w

ORIGINAL ARTICLE

The Lichen Flavin-Dependent Halogenase, *Dn*Hal: Identification, Heterologous Expression and Functional Characterization

Nurain Shahera Hasan[1,2] · Jonathan Guyang Ling[1] · Mohd. Faizal Abu Bakar[3] · Wan Mohd Khairulikhsan Wan Seman[3] · Abdul Munir Abdul Murad[1] · Farah Diba Abu Bakar[1] · Rozida Mohd. Khalid[2]

Accepted: 16 December 2022

Abstract
Enzymatic halogenation captures scientific interest considering its feasibility in modifying compounds for chemical diversity. Currently, majority of flavin-dependent halogenases (F-Hals) were reported from bacterial origin, and as far as we know, none from lichenized fungi. Fungi are well-known producers of halogenated compounds, so using available transcriptomic dataset of *Dirinaria* sp., we mined for putative gene encoding for F-Hal. Phylogenetic-based classification of the F-Hal family suggested a non-tryptophan F-Hals, similar to other fungal F-Hals, which mainly act on aromatic compounds. However, after the putative halogenase gene from *Dirinaria* sp., *dnhal* was codon-optimized,cloned, and expressed in *Pichia pastoris*, the ~63 kDa purified enzymeshowed biocatalytic activity towards tryptophan and an aromatic compound methylhaematommate, which gave the tell-tale isotopic pattern of a chlorinated product atm/z 239.0565 and 241.0552; and m/z 243.0074 and 245.0025, respectively. This study is the start of understanding thecomplexities of lichenized fungal F-hals and its ability to halogenate tryptophan andother aromatic. compounds which can be used as green alternatives forbiocatalysis of halogenated compounds.

Keywords Biocatalysis · Chlorination · Halogenating enzymes · Lichen · Recombinant proteins · Phenolic compounds

✉ Rozida Mohd. Khalid
rozidakhalid@ukm.edu.my

[1] Department of Biological Sciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

[2] Department of Chemical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

[3] Malaysia Genome & Vaccine Institute, National Institutes of Biotechnology Malaysia, Jalan Bangi, 43000 Kajang, Selangor, Malaysia

Published online: 13 March 2023

Abbreviations

F-Hal	Flavin-dependent halogenase
FAD	Flavin adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide hydrogen
RP-TLC	Reversed-phase thin-layer chromatography
LC-MS	Liquid chromatography-mass spectrometry

Introduction

Enzymatic halogenation has emerged as a tool to advance synthesis applications. The high-regioselective reaction governed by the Flavin-dependent halogenases (F-Hals) offer high substrate specificity under benign condition [1, 2], therefore paving the way to improvise the synthetic processes. Investigations on the bacterial F-Hals that exhibit high substrate specificity towards tryptophan provide significant contributions towards enzymatic modification of tryptophan. By taking advantage of the chemistry of tryptophan, the indole moiety that is prone to electrophilic aromatic substitution allows it to undergo other biochemical transformations. The structural reactivity permits it to be the substrate of several enzymes catalyzing halogenation, oxidization or methylation to generate functional materials for chemical diversity [3]. Often, enzymatic halogenation regulates C-H activations; however, its activity on tryptophan confers low stability and reactivity, rendering its applications as a whole. Several improvements have been made to produce the halogenated products in preparative scales. The enzymatic reaction, including its cofactor regeneration system, was immobilized with a solid biocatalyst (combiCLEAs) to enhance the production of the halo-compounds [4]. This 57 improvement advances the synthetic processes, allowing a feasible reaction of integrating the halo-tryptophan in late-stage derivatization and cyclization [5, 6].

Interestingly, the fungal F-Hals appear as biocatalysts that mediate a wide range of substrates scope, therefore are involved in functional diversity of aromatic substrates. The fungal Rdc2 enzyme is involved in tailoring monocillin in the biosynthetic production of radicicol. Subsequent evaluation on other similar lactone structures also led to the generation of chlorinated products, thus signifying the potential of fungal F-Hals [7]. The isoenzyme RadH also displays similar behaviour, able to modify its natural substrates and other phenolic compounds. The promiscuity exhibited by the RadH enzyme make way for further exploration of F-Hal enzymes from other eukaryotic resources. Also, attempts in integrating the RadH enzyme in the engineered pathway producing coumarin [8] adding in the applicability of the F-Hal enzymes. Either by integrating the F-Hal enzymes in a catalysis or integrating it in the natural products producing pathways, it is without a doubt that their existence had significant impacts in chemical diversity, and without further due, expanding the research area, especially in identifying new and functional enzymes, could benefit the bio- and chemosynthesis community.

It is undeniable that numbers of F-Hal enzymes were available in the literature, yet we noticed that most of them were of bacteria origin, and only a few were from fungal or higher organisms. Most bacterial enzymes were only selective towards tryptophan [9–11]. Meanwhile, the fungal F-Hal enzymes are speculated to be inactive against tryptophan; instead, they prefer to catalyse the respective natural substrates or structurally similar compounds in the range of simple to complex phenolic substrates [7, 13, 14]. However, the *dnhal* gene from *Dirinaria* sp. which was heterologously expressed, and tested with tryptophan and a mono-aromatic compound methyl haematommate showed chlorination in both products. It is thus postulated that the lichen enzyme can halogenate tryptophan and the aromatic compounds, and this effort expands the available F-Hal enzymes for aromatic

modifications. It should be noted that methyl haematommate is a precursor for the production of atranorin, a depside commonly isolated from *Dirinaria* sp. To summarize, this study aims to scout a potential gene encoding for F-Hal from the lichen *Dirinaria* sp. and investigate its function. This study provides the basis for future studies targeting alterations of other phenolic substrates to generate halogenated active compounds or chemical building blocks for subsequent synthetic processes using green alternatives.

Materials and method

Microbial Strains, Cloning and Expression Vector

Pichia pastoris X-33, *Escherichia coli* DH5α and BL21 STAR (DE3) were obtained from Invitrogen (Thermo Fisher Scientific, USA). The cloning vector pMAT_*Dnhal* harbouring the codon-optimized *dnhal* was also obtained from Invitrogen (Thermo Fisher Scientific, USA). Meanwhile, the expression vector pET28B_prnF_BL915 harbouring the *Pseudomonas fluorescens* flavin reductase was a kind gift from Prof. Dr. Karl-Heinz van Pèe (TU Dresden, Germany).

Chemicals and Solvents

Tryptone, Yeast Nitrogen Base (YNB) with ammonium sulphate without amino acids and Sorbitol were obtained from Amresco, USA. Yeast extract and Bacteriological agar were obtained from Oxoid, Thermo Fisher Scientific, UK. Peptone from Pronadisa, Condalab, Spain. Glucose, isopropyl β-D-1-thiogalactopyranoside (IPTG), β-nicotinamide adenine dinucleotide (NADH), flavin adenine dinucleotide (FAD) and LC-grade chemicals (water, methanol, ethyl acetate and trifluoroacetic acid), including all other chemicals, were obtained from Merck, USA. L-Tryptophan was purchased from Merck, USA. Methyl haematommate was extracted and purified by Anis and colleagues [15].

Identification and Bioinformatic Analyses of Putative *Dn*Hal Encoding Genes

The transcriptome dataset of *Dirinaria* sp. was previously pre-processed, reconstructed and assembled by I. Bharudin and colleagues [16]. We used the raw transcript data as a reference to mine for the putative genes that encode the F-Hal. The raw FASTQ files analysed during the current study are available in the NCBI SRA database with the accession number SRP138994 (https://www.ncbi.nlm.nih.gov/sra/?term=SRP138994). The Trinity platform version 2.0.2 (https://trinityrnaseq.github.io) was used to generate a reference assembly by combining all the read datasets into a single target [17]. Then, the original RNA-seq reads were aligned against the Trinity transcripts. A database of known halogenases was created in the Trinity platform. Gene identification of putative gene encoding for halogenase was performed by using tBLASTn search of the translated transcripts against halogenase database in the Trinity platform. For further confirmation, the candidate of genes was BLASTx against these following databases: NCBI non-redundant protein sequences (Nr, https://blast.ncbi.nlm.nih.gov/), UniProt KnowledBase (UniProtKB, https://www.uniprot.org/) and RSCB Protein DataBank (https://www.rcsb.org/). The Pfam

database (https://pfam.xfam.org/) and InterPro scanning accessible at https://www.ebi.ac.uk/interpro/ were used to predict the conserved domains and important sites of the *Dn*Hal enzyme [18, 19].

Multiple Sequence Alignment

Multiple sequence alignment analysis of putative *Dn*Hal and the characterized F-Hal sequences was conducted in T-Coffee simple MSA web-based (http://tcoffee.crg.cat/) server [20]. The protein sequences annotated as tryptophan halogenases in the NCBI databases were also included in the analysis. The flavin-dependent oxygenase, *p*-hydroxybenzoate hydroxylase PHBH (PDB ID: 1PHH), was used as a reference to differentiate and characterize the function of the F-Hal enzymes. The generated alignment file was visualized using the ESPript tool available at https://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi [21]. GenBank accession numbers of all protein sequences are listed in Table S1.

Phylogenetic Analyses

The previous alignment data was checked manually and exported into a MEGA file using the software MEGA-X [22]. For the phylogenetic construction of F-Hals amino acid sequences, the neighbour-joining method was implemented and is estimated using the *p*-distance model and a gamma-shaped rate variation with a proportion of invariable sites. Internal branches' reliability was assessed using 1000 bootstrap replicates. The phylogenetic tree was edited and visualized via iTOL tool accessible at https://itol.embl.de/ [23]. The flavin-dependent oxygenase, *p*-hydroxybenzoate hydroxylase PHBH (PDB ID: 1PHH), was used as an outgroup in the analysis. GenBank accession numbers for all F-Hal sequences can be found in Table S1.

Construction of Expression Plasmid and Transformation of *Pichia pastoris* X-33

The *dnhal* gene was codon-optimized based on codon usage in *P. pastoris* and chemically synthesized with extended restriction sites: *Kpn*I at the N-terminal and *Xba*I at the C-terminal end. A Kozak consensus sequence (CGAAACG) was integrated between the *Kpn*I restriction site and the start codon. The cloning plasmid pMA-T_*dnhal* harbouring the synthesized *dnhal* gene was digested with restriction enzymes *Kpn*I and *Xba*I and the excised product was ligated into the pre-digested expression plasmid pPICZB using T4 ligase (Promega, USA), incubated at 4 °C for 16 h, and the ligation reaction was subsequently used to transform *E. coli* strain DH5α competent cells. The resulting expression construct, pPICZB_*dnhal*, was confirmed via restriction enzyme digestion and PCR amplification using the primer pair F_*dnhal* 5′-GGTACCCGAAACGATGTCCATTCC-3′ and R_*dnhal* 5′-TCTAGAATAGTCTTGATAGCAGTTGGCAATGG-3′ and validated via Sanger sequencing.

To increase transformation efficiency, pPICZB_*dnhal* was linearized using the restriction enzyme *Bst*XI, concentrated via ethanol precipitation and electroporated into *P. pastoris* X-33 competent cells. A total of 10–15 μg linearized plasmid was mixed with 80

μL competent cells in 0.4-cm-gap electroporation cuvettes and pulsed for 4.5 ms at a field strength of 7500 V/cm using a Bio-Rad Gene Pulser. The transformants were selected on YPDSZ plates (1% yeast extract, 2% peptone, 2% glucose, 1 M sorbitol, 1.5% agar, 100 μg/mL Zeocin), and the resultant strain was denoted as X33-pPICZB_*dnhal*. Using the same method, the vector pPICZB was transformed into *P. pastoris* X-33 to generate X33-pPICZB strain, and this was used as the control strain during the experiments. Colony PCR analysis was carried out to verify the integration of expression construct in the *AOX*1 locus of the *P. pastoris* X-33 genome. A pair of primers, *AOX*1 forward, 5′-GACTGGTTC CAATTGACAAGC-3′, and *AOX*1 reverse, 5′-GCAAATGGCATTCTGACATCC-3′, was synthesized and used for PCR amplification of the *AOX*1 locus containing the *dnhal* coding sequence.

Growth Curve Generation of X-33_pPICZB_*dnhal* in Growth Medium

Positive individual colonies harbouring *dnhal* gene were grown in a 10 mL BMGY medium (100 mM potassium phosphate, pH 6.0, 1.34% YNB, 1% glycerol) in a rotary shaker cultivated at 28 °C and shaken at 250 rpm as a pre-culture. Approximately 1×10^6 cells/mL were transferred into 200 mL of BMGY cultured in a 1 L conical flask, shaken at 250 rpm and continuously cultured at 28 °C for 4 days. The cells were harvested every 4 h, and the optical density (OD_{600}) with the number of cells was calculated using a hemacytometer. A graph was plotted to observe the growth curve of the positive transformant in the selected growth medium.

Selection of Methanol Concentration and Incubation Period for the Optimization of *Dn*Hal Production in *P. pastoris*

The transformant with highest resistance towards zeocin was chosen and cultured for 24 h in 5 mL of BMGY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate buffer, pH 6.0, 1.34% (w/v) YNB, 4×10^{-5}% (w/v) biotin and 1% (v/v) glycerol) as a pre-culture. An ~ 1×10^6 cells/mL were transferred into 10 mL of BMGY cultured in a 50-mL conical flask, shaken at 250 rpm and cultivated at 28 °C for 44 h until OD600 reached ~ 30. The cells were harvested using a centrifugation system (at 2000 g for 10 min) before resuspending in 10 mL BMMY (100 mM potassium phosphate, pH 6.0, 1.34% YNB, 1% methanol). The concentration of methanol (100%) was manipulated to 0.5%, 1.0% and 2.0% in the growing cultures. The cells were harvested every 24 h and centrifuged at 10,000 x g for 10 min. The cell pellet was lysed with YeastBuster™ Protein Extraction Reagent, and the supernatant was used as a crude enzyme solution for total protein calculation (calculated using the Bradford assay method [24]).

Large-Scale Expression of *Dn*Hal in *P. pastoris* X-33

Individual colonies of X-33_pPICZB_*dnhal* were grown in 5 mL BMGY medium (100 mM potassium phosphate, pH 6.0, 1.34% YNB, 1% glycerol) and cultivated at 28 °C with 250 rpm shaken in a rotary shaker. An inoculum of 1×10^6 cells/mL was added into 800 mL of BMGY and cultivated at 30 °C with shaking at 250 rpm for 48 h

until the cell density achieved OD_{600} of 33 ~ 34. Cells were harvested via centrifugation at 3100 × g for 5 min at 4 °C and the pellet was resuspended with 400 mL of MM medium (1.34% (w/v) YNB, 4×10^{-5}% (w/v) biotin and 0.5% (v/v) methanol). Methanol was added every 24 h to a final concentration of 1.0% to maintain induction of gene expression during the 72-h incubation period at 30 °C with shaking at 250 rpm. The cells were pelleted via centrifugation at 3100 × g for 5 min at 4 °C and stored at − 80 °C until further analyses. The cell lysate was purified to confirm the expression of *Dn*Hal (view the "Protein Purification" section), and the purified enzyme was separated in the SDS-PAGE analysis. Later, the band corresponding to ~ 63 kDa was excised, and was enzymatically cut into small peptides before being sent to the Mass Spectrometry Technology Section (MSTS), Malaysia Genome And Vaccine Institute (MGVI). Identification of *Dn*Hal was analysed using the reversed-phase nanoLC coupled with Orbitrap Fusion. Later, the data analysis software is used to search the acquired MS/MS spectra to identify the proteins. The software will confirm the identified protein and the post-translational modifications if present (the MSTS, MGVI run the analysis and process the data). We manually mapped the identified peptides to our reference sequence and presented the data in the "Results" section.

Cloning and Recombinant Expression of Flavin Reductase prnF in *E. coli*

The plasmid pET28B_prnF-BL915 harbouring the flavin reductase encoding gene was cloned in *E. coli* strain DH5α competent cells. The plasmid was isolated and verified through restriction enzyme digestion before being used to transform *E. coli* strain BL21 STAR (DE3) competent cells. An overnight pre-culture cultivated at 37 °C with vigorous shaking at 250 rpm was used to inoculate a culture in LB medium (1% tryptone, 0.5% yeast extract and 1% NaCl) containing the appropriate antibiotic, 50 µg/mL kanamycin. This culture was grown at 37 °C with continuous shaking at 250 rpm up to an OD_{600} of 0.6. Recombinant protein expression was then induced with isopropyl-*β*-D-thiogalactopyranoside (IPTG) to a final concentration of 1.0 mM, and the cells were subsequently incubated at 20 °C with shaking at 250 rpm for 22 h. Afterwards, the cells were pelleted by centrifugation at 3100 × g for 10 min at 4 °C and stored at − 80 °C for subsequent analyses.

Protein Purification

The prnF cell pellet was resuspended in 10 mL of lysis buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 20 mM βME) per gram (wet weight) of pellet. The cell suspensions were disrupted on ice with a 15 min sonication composed of 10 s pulse with 20 s rest at amplitude of 40% power using an ultrasonicator. The cell lysate was then centrifuged at 10,000 × g for 30 min at 4 °C. The clarified lysate was then filtered through a syringe filter (0.2 µm pore size) to produce cell-free extract. The *Dn*Hal cell pellet was resuspended in 10 mL of lysis buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 20 mM βME) per gram (wet weight) of pellet. The clarified cell-free extract was applied onto a HisTrapTM HP 5 mL column (GE Healthcare), pre-equilibrated with binding buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 20 mM βME, 30 mM Imidazole) at a flow rate of 5 mL/min. After washing with 20 column volumes (CV) of binding buffer, the protein was eluted with elution buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 20 mM βME, 500 mM Imidazole) over a 10 CV linear gradient. The eluted protein was collected in 1

mL fractions. Eluted protein fractions were then pooled and concentrated using an Amicon ultracentrifugal device (10 kDa MWCO; Merck, U.S.A.) at 4 °C. All purifications were performed using an AKTA purifier (GE Healthcare).

Functional Characterization

The crude *Dn*Hal and prnF enzymes were used to screen the enzymatic potential of *Dn*Hal, and the reaction products were analyzed using RP-TLC analysis. Meanwhile, the purified enzymes were included in an *in-vitro* assay to validate the enzymatic activity of *Dn*Hal towards tryptophan, methyl haematommate, and the reaction products were separated and identified in the LC-MS analysis. The enzymatic assay of *Dn*Hal was carried out following the Zeng and Zhan with slight modifications (reaction mixture consisting of 100 μM of FAD, 10 mM of NADH, 10 mM of KCl, 1.0 mM of substrates, 32 μM of prnF, 32 μM of *Dn*Hal in 100 μL of 0.1 M phosphate buffer (pH 7.0)) was prepared and incubated at 28 ◦C with vigorous shaking for two hours [7]. A reaction assay without the substrate was used as a control reaction.

After two hours, the reaction assay was quenched with a 1:1 ratio of ethyl acetate, and the organic layer was used in reverse-phase thin layer chromatography (RP-TLC) analysis. A modified R-18 silica gel coated with fluorescent indicator F254s was used as the stationary phase. Meanwhile, a varied mobile phase concentration of 10%, 30%, 50% and 70% acetonitrile was optimized to separate the halogenation products. For the LC-MS analysis, the reaction assay was quenched with a 1:1 ratio of ethyl acetate, and the organic layer was filtered and sent to i-CRIM Centralized Lab, Universiti Kebangsaan Malaysia (UKM). The samples were analyzed on an Agilent Single Quad LC-MS with a Zorbax SB-C18 reversed-phase analytical column (particle size of 5 mm with a diameter of 150 mm × 4.6 mm) at 310 nm, with a flow rate of 1 mL/min. A gradient of the MeCN/H2O system (10 – 90 %) containing 0.1 % trifluoroacetic acid was run over 30 minutes. The raw data was analyzed using the MZmine version 3. The raw data was filtered, and the exact mass detector method was implemented to scan the presence of the halogenated product from the mass spectra. We utilized the isotope pattern checking tool to reconfirm the detection of the halogenated product. The molecular ion mass adducts of tryptophan and methyl haematommate with their respective halogenated products were predicted by manually scanning to the data provided by Fiehn Laboratory [25].

Results

Identification and Characterization of the Putative F-Hal Encoding Genes

In this work, we screened for protein sequences that exhibited sequence similarity with classes of halogenating enzymes. A total of 23 candidate genes that putatively encode for halogenating enzymes were discovered from the transcriptome mining analysis, out of which 19 genes were predicted to encode for F-Hals, with the four remaining genes encoding for tryptophan halogenases and non-heme-dependent halogenases (Table 1a). The identified NH-dependent halogenase encoding genes (DN65145_c0_g1_i1, DN49640_c1_g1_i1) were matched to the recent report on NH-dependent halogenase from the lichen *C. uncialis* [26] with nearly 55.8% of sequence similarity. As most of the predicted protein

Table 1 Identification of the putative gene encoding for the F-Hal enzyme. (a) The tBLASTn searching leads to the identification of 23 candidates of genes encoding for halogenating enzymes. (b) The blast searching analysis of the targeted genes encoding for the F-Hal enzymes

Part a

Gene identity	Description (NCBI accession number)	*E*-value
DN90121_c1_g1_i2	radH F-Hal *Lasallia pustulata* (KAA6409706)	2.20E-204
DN90121_c1_g1_i1	radH F-Hal *Lasallia pustulata* (KAA6409706)	6.30E-204
DN90121_c0_g1_i2	radH F-Hal *Lasallia pustulata* (KAA6409706)	6.40E-204
DN90121_c0_g1_i1	radH F-Hal *Lasallia pustulata* (KAA6409706)	1.20E-198
DN90121_c1_g1_i3	radH F-Hal *Lasallia pustulata* (KAA6409706)	9.10E-189
DN90121_c0_g1_i3	radH F-Hal *Lasallia pustulata* (KAA6409706)	8.50E-188
DN65145_c1_g1_i1	aclH F-Hal *Lachnellula suecica* (TVY85213)	6.30E-37
DN124391_c0_g1_i1	aclH F-Hal *Lachnellula suecica* (TVY85213)	1.80E-34
DN65145_c0_g1_i1	putatif αK-DH *Cladonia uncialis* (ANM27731)	6.60E-28
DN49640_c1_g1_i1	putatif αK-DH *Cladonia uncialis* (ANM27731)	1.60E-25
DN52864_c1_g1_i1	aclH F-Hal *Lachnellula suecica* (TVY84588)	1.10E-20
DN52864_c0_g1_i1	aclH F-Hal *Lachnellula suecica* (TVY84588)	4.40E-17
DN39858_c0_g1_i1	radH F-Hal *Aspergillus costaricaensis* (XP_025545186)	3.70E-15
DN39858_c0_g2_i1	radH F-Hal *Aspergillus costaricaensis* (XP_025545186)	3.70E-15
DN16924_c1_g1_i1	Triptofan halogenase *Penicillium* sp. (PCH01679)	3.30E-14
DN151026_c0_g1_i1	aclH F-Hal *Lachnellula suecica* (TVY85213)	4.10E-11
DN112384_c0_g1_i1	aclH F-Hal *Lachnellula suecica* (TVY85213)	4.70E-11
DN49640_c0_g1_i1	aclH F-Hal *Lachnellula suecica* (TVY85213)	1.90E-10
DN129818_c0_g1_i1	radH F-Hal *Aspergillus costaricaensis* (XP_025545186)	2.60E-09
DN132883_c0_g1_i1	aclH F-Hal *Lachnellula suecica* (TVY84588)	1.50E-08
DN140916_c0_g1_i1	aclH F-Hal *Lachnellula suecica* (TVY84588)	1.20E-05
DN46536_c0_g1_i1	radH F-Hal *Metarhizium robertsii* (XP_007817603)	5.10E-05
DN16924_c0_g1_i1	Triptofan halogenase *Metarhizium rileyi* (OAA35535)	3.20E-04

Part b

Dirinaria sp. genes	*E*-value	Sequence identity (%)	Matched protein sequences

Table 1 (continued)

DN90121_c1_g1_i2	3.00E-154	49.8	GsfI
DN90121_c1_g1_i1	2.00E-153	50.0	GsfI
DN90121_c0_g1_i2	3.00E-153	50.0	GsfI
DN90121_c0_g1_i3	9.00E-27	28.3	Mpy16
DN90121_c1_g1_i3	5.00E-23	29.4	Flavin-dependent chlorinase

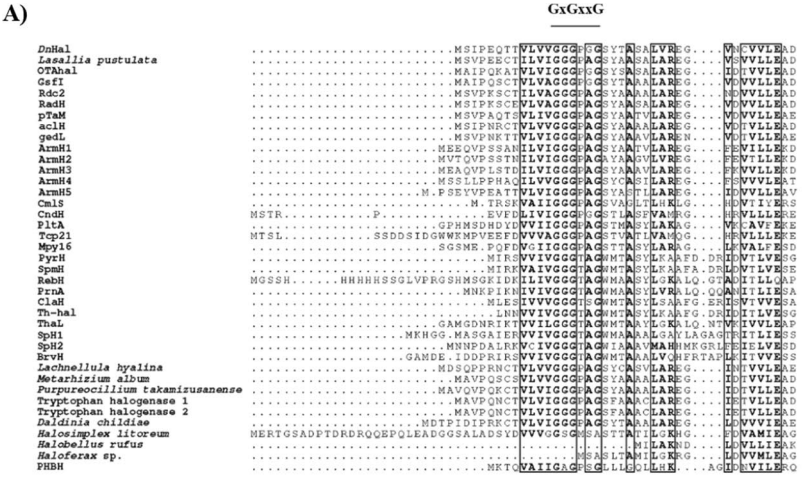
Fig. 1 A consensus motifs and residue of *Dn*Hal. The peptide sequence of DnHal was aligned with the characterized F-Hal sequences. (**A**) The conserved GxGxxG motif is marked above the sequences. (**B**) The conserved WxWxIP motif is marked above the sequences. (**C**) The conserved lysine residue is marked with black reversed triangle

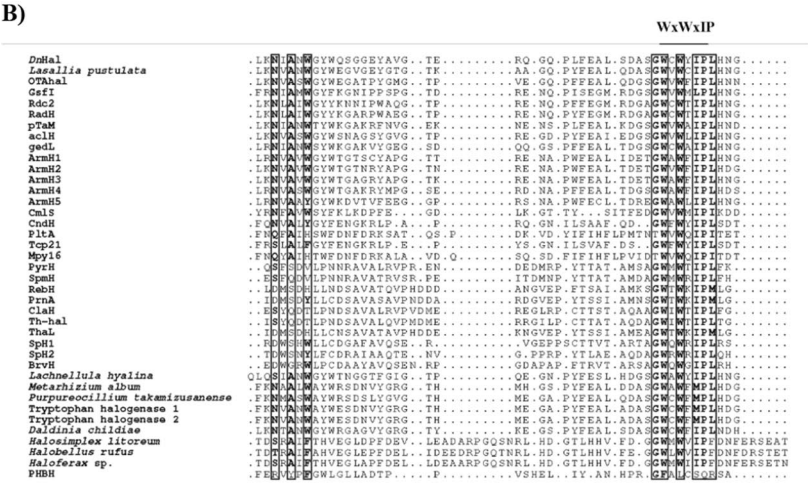
sequences show high similarity to F-Hal protein sequences, we directed this study towards identifying novel gene(s) encoding for F-Hal enzymes. Subsequent tBLASTn analysis leads to the identification of five potential candidates of F-Hal genes (Table 1b). However, only one gene, designated as *dnhal* (DN90121_c1_g1_i2), possessed a complete open reading frame with a start and a stop codon. The *dnhal* gene has a full length of 1689 bp, encoding for a protein of 563 amino acids with a deduced molecular weight of ~ 61 kDa and a theoretical isoelectric point (pI) of 6.11. We noticed that this gene shared almost 65.0% sequence identity with the identified eukaryotic F-Hal enzyme, RadH from the lichen *Lasallia pustulata* (NCBI accession ID: KAA6409706, unpublished). When comparing *Dn*Hal with the lichen *C. uncialis* non-heme halogenase CU-PKS-4(4) in BLAST searching analysis, we observed that both enzymes shared nearly 58.7% sequence similarity.

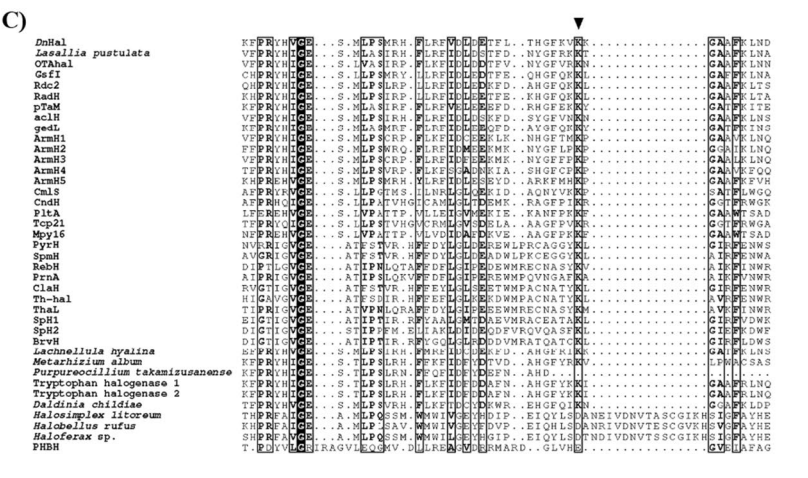
Regardless, conserved domain searching analyses showed the presence of a conserved domain of the tryptophan halogenase family (PF04820), which places *Dn*Hal into the F-Hal enzyme family (Interpro ID: IPR006905), and therefore supported our preliminary BLAST results. We also noticed the presence of the overlapped FAD/NAD(P)-binding domain (IPR036188) of the homologous superfamilies, suggesting close relation to the monooxygenases family [27]. Multiple sequence alignment analysis of the *Dn*Hal encoding genes with the identified and biochemically characterized F-Hal enzymes showed a conserved sequence motif GxGxxG at the amino terminal of the sequence. The consensus motif at the FAD-binding domain highlights the vital feature of *Dn*Hal as an FAD-dependent enzyme responsible for the enzymatic halogenation of aromatic substrates. Another conserved sequence motif, WxWxIP, was also observed at the midpoint of the *Dn*Hal sequences (Figure 1). The presence of the WxWxIP motif differentiates the catalytic function of the F-Hal enzyme from the monooxygenase enzyme family [28]. We observed that the *Dn*Hal sequence possessed the K74 residue that is well conserved with the reference F-Hal enzymes; thus, we assume that lysine-chloramine is likely to be involved in the halogenation mechanism of the putative *Dn*Hal [11, 29, 30]. All things considered, we suggest that *Dn*Hal is mechanistically similar to extant characterized F-Hal enzymes [2, 29–33] (Figure 1).

Construction of the Phylogenetic Tree

The overall layout in Fig. 2 portrayed two clades of enzymes, substantially categorizing the F-Hal enzymes into two major families. A thorough observation of the function indicates that the monophyletic group 2 was involved in the selective halogenation of tryptophan, thereof was commonly categorized as tryptophan F-Hal enzymes (marked with a dark blue line). We noticed that the enzymes SpH1, BrvH and SpH2 that build paraphyletic group 2B (marked with a light purple dashed line) were active towards indole derivatives and also capable of installing bromine halogen into their respective substrates [33, 34]. Meanwhile, the F-Hal enzymes grouped into paraphyletic group 2B possess more stringent substrate specificity, in which the enzymes were only catalysing the C5, C6 and C7 regio-selective halogenation of tryptophan [9, 10, 12, 32, 35–37]. In this study, the putative tryptophan F-Hal enzymes originating from archaea were also included, and we found that the three respective enzymes were clustered as a paraphyletic group 2C (marked with a turquoise line).


A)
GxGxxG
DnHal
Lasallia pustulata
OTahal
GsfI
Rdc2
RadH
pTaM
aclH
gedL
ArmH1
ArmH2
ArmH3
ArmH4
ArmH5
CmlS
CndH
PltA
Tcp21
Mpy16
PyrH
SpmH
RebH
PrnA
ClaH
Th-hal
ThaL
SpH1
SpH2
BrvH
Lachnellula hyalina
Metarhizium album
Purpureocillium takamizusanense
Tryptophan halogenase 1
Tryptophan halogenase 2
Daldinia childiae
Halosimplex litoreum
Halobellus rufus
Haloferax sp.
PHBH


B)
WxWxIP
DnHal
Lasallia pustulata
OTahal
GsfI
Rdc2
RadH
pTaM
aclH
gedL
ArmH1
ArmH2
ArmH3
ArmH4
ArmH5
CmlS
CndH
PltA
Tcp21
Mpy16
PyrH
SpmH
RebH
PrnA
ClaH
Th-hal
ThaL
SpH1
SpH2
BrvH
Lachnellula hyalina
Metarhizium album
Purpureocillium takamizusanense
Tryptophan halogenase 1
Tryptophan halogenase 2
Daldinia childiae
Halosimplex litoreum
Halobellus rufus
Haloferax sp.
PHBH


C)
DnHal
Lasallia pustulata
OTahal
GsfI
Rdc2
RadH
pTaM
aclH
gedL
ArmH1
ArmH2
ArmH3
ArmH4
ArmH5
CmlS
CndH
PltA
Tcp21
Mpy16
PyrH
SpmH
RebH
PrnA
ClaH
Th-hal
ThaL
SpH1
SpH2
BrvH
Lachnellula hyalina
Metarhizium album
Purpureocillium takamizusanense
Tryptophan halogenase 1
Tryptophan halogenase 2
Daldinia childiae
Halosimplex litoreum
Halobellus rufus
Haloferax sp.
PHBH

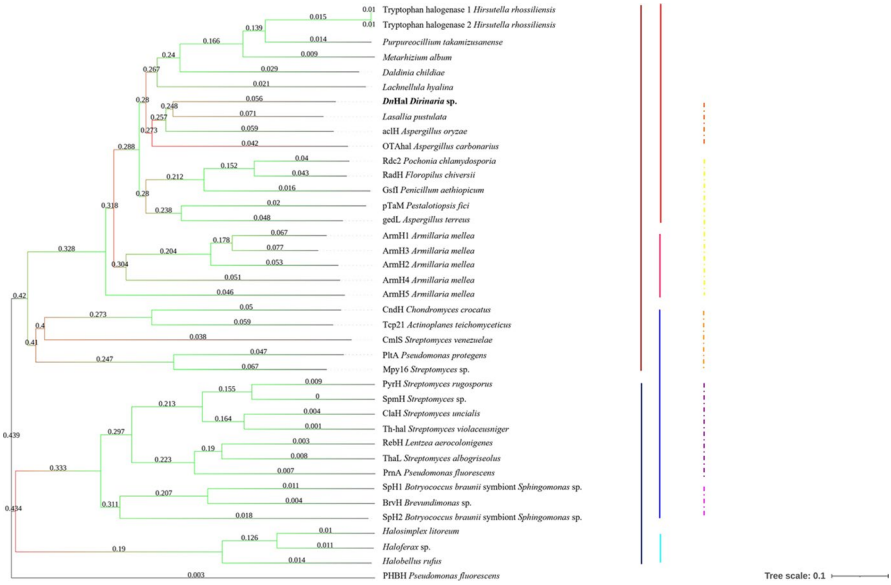


Fig. 2 Phylogenetic analysis of *Dn*Hal. Neighbour-joining analysis was based on an alignment of the *Dn*Hal with 38 characterized F-Hal sequences. The monooxygenase PHBH was used as the outgroup

In contrast, the enzymes clustered in the monophyletic group 1 were not active towards tryptophan. Instead, the enzymes clustered into paraphyletic group 1A prefer to halogenate the free-standing phenolic substrates (marked with a dark red line). On the other hand, those in the paraphyletic group 1B (marked with a brown dashed line) would instead catalyse the halogenation of substrates tethered to carrier proteins [12, 38, 39, 42, 43]. We noticed that group 1A mainly consisted of F-Hal enzymes from phylum Ascomycota (marked with a pink line) and Basidiomycota (marked with a red line). Even though the ArmH5 is homologous to the ArmH2 to ArmH4, we noticed that it was excluded from its node [8, 13, 15, 44–48].

The annotated RadH from *L. pustulata* and six representative of the putative F-Hals identified from fungi were included in the phylogenetic study. We observed that the six candidates of fungal F-Hal enzymes were grouped into a node that deviated them from other fungal F-Hal enzymes. On the other hand, *Dn*Hal was clustered into a node consisting of the *L. pustulata* RadH and the *Aspergillus* aclH and OTAhal enzymes (marked with an orange dashed line). This finding gave insight into substrate preference of *Dn*Hal by delving further into mechanistic halogenation of aclH and OTAhal. Both enzymes catalyse late-stage halogenation on their respective free-standing phenolic substrates [45, 47]. Therefore, we postulate that *Dn*Hal is also prone to halogenate the structurally similar substrates and that it is not active towards tryptophan.

Generation of the Growth Curve for X-33_pPICZB_dnhal Cells

The growth curve of the X-33_pPICZB_*dnhal* cells demonstrated a short lag phase, lasting for 12 h, followed by exponential growth for 48 h before reaching the late exponential phase and beginning the stationary phase for an additional 36 h (Fig. 3). This finding was

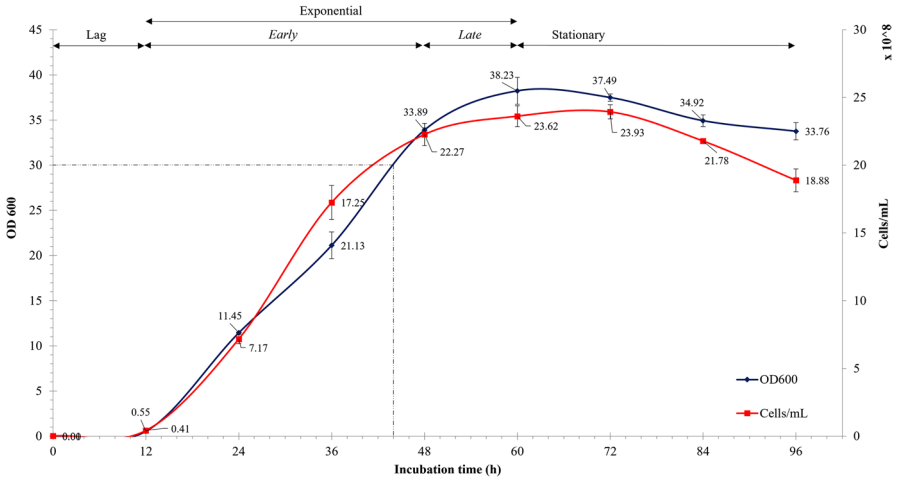


Fig. 3 Generation of the growth curve for X-33_pPICZB_dnhal cells. The cells were grown in a BMGY medium, and the growth was observed by calculating the number of cells and visualizing the optical density (OD600) every 4 h until 96 h of cultivation. The plot indicating the number of cells was highlighted in red, while the plot for optical density values was coloured in blue

comparable to the growth of yeast cells observed by Asaduzzaman, who stated that the lag, exponential and stationary phases were involved during the growth of the yeast cells [48].

We observed that the number of cells remained constant during the lag phase, and the initial number of cells ~ 1×10^6 cells/mL was sufficient to induce rapid growth of our yeast cells cultivated in a standard complex BMGY medium. It is widely accepted that the exponential phase in which the cells rapidly grow is almost the best period to harvest cells for most of the subsequent experimental studies, and it is important to note that the cells grow with a defined generation time during this period. The generation time, i.e. the period in which the number of the X-33_pPICZB_*dnhal* cells is doubled, is therefore estimated and computed by using the equation described below:

$$\text{Generation time (g)} = 0.301/\text{m} \tag{1}$$

where *m* is the slope of the straight-line function obtained in a plot of exponential growth in which the respective value is converted to a natural logarithm function.

By implying Eq. (1), the generation time for our transformant cells was calculated. We observed that the generation time for X-33_pPICZB_*dnhal* cells was between ~ 167 (visualized on the OD_{600} plot) and ~ 169 min (visualized on the cells count plot) (Figure S2). It is slightly slower compared to the generation time of the wild-type host reported by Farsiani and colleagues, in which its generation time is between 60 and 120 min [49]. Further observation showed that the growth of the cells begins to slow during the late exponential phase. During this stage, the sugar is almost depleted, slowing down the growth of the cells. Because high biomass of cells is required to enhance the production of recombinant proteins [50], we proposed that the 44 h of incubation, in which OD_{600} of the cells is at ~ 30, is the suitable period to harvest the cells and that the biomass of cells is sufficient for subsequent heterologous expression processes (Fig. 3).

Intracellular Expression of DnHal in P. pastoris

Primarily, the *dnhal* gene was designed to include the conserved Kozak sequence, and later, the sequence was optimized based on the codon preference of *P. pastoris*, and cloned to the cloning vector (designed, synthesized and cloned by Invitrogen, USA). Then, the *dnhal* gene was excised by the restriction enzymes and ligated to generate the expression vector pPICZB_*dnhal* (Figure S3). The generated expression vector was linearized and transformed into the genome of *P. pastoris* (Figure S3) to generate the X-33_pPICZB_*dnhal* cells that are ready for the intracellular expression process. Based on profiling in Fig. 3, we observed that the ~63 kDa of *Dn*Hal was visible in the SDS-PAGE. Optimization of methanol concentration provides no significant difference in the production of the *Dn*Hal enzyme. Only a thin band was visible on the SDS-PAGE, suggesting that other factors might be affecting the overall recombinant expression processes, and further experimental study is required to improve the production of our *Dn*Hal enzyme.

Regardless, protein identification via LC–MS demonstrated that the expressed peptides were highly similar to the protein sequences of *Dn*Hal (Fig. S4). We noted that 23 peptides of the expressed *Dn*Hal were detected from the SEQUEST database. The total number of identified peptide sequences (peptide spectrum matches, PSM) for the *Dn*Hal was attached in Table 2. At least two peptides mapped to the database were required to confirm protein identity [51]. We also sequenced the peptides of the host *P. pastoris* X-33 and discovered that none of the expressed peptides matched the peptide sequence of *Dn*Hal (Table 2). Thus, it is confirmed that the expressed protein was identified as *Dn*Hal, the putative F-Hal enzyme with a molecular weight of 63.733 kDa. Therefore, it is suggested that the *Dn*Hal is expressed under optimal conditions and that the expression system of *P. pastoris* is suitable for expressing the recombinant enzyme.

Enzymatic halogenation of *Dn*Hal

To test the function of the recombinant enzyme, we used tryptophan and a mono-aromatic compound, methyl haematommate as substrates. Tryptophan was tested as substrate to confirm the phylogenetic analysis that suggested *Dn*Hal is typical non-tryptophan fungal F-Hals, while methyl haematommate is a precursor of the depside atranorin, a very common metabolite from *Dirinaria* sp. The products were analysed via LC-MS.

The *Dn*Hal and prnF (the flavin reductase catalyzing reduction of the co-factor FAD) were expressed to obtain a sufficient amount of enzymes for the halogenation assay. The expression profile of prnF is available in Figure S5. It can be observed that there was an active reaction in the halogenation assay of the crude *Dn*Hal enzyme with tryptophan. As shown in Fig. S5, four putative spots were visible on the RP-TLC plate, suggesting the presence of reaction products. We observed that the halogenation product is better separated in the mobile phase consisting of 30% acetonitrile (Figure S6). Concurrently, the same mobile phase is used to separate the control reaction in Fig. S6. A control reaction consisting of the single components of FAD, NADH and tryptophan was separated on the RP-TLC plate.

We found that compounds denoted as NADH (compound ii) separated faster, followed by FAD (compound iii) than tryptophan (compound iii). Compound i that emitted a yellow spot on the chromatogram is detected as flavin in considering the same colour was also observed from separation analysis conducted by Takahashi and colleagues [52]. A new spot at Rf 0.76 (compound iv) was detected from the RP-TLC plate in Fig. S5. Considering our theoretical hypothesis that *Dn*Hal can mediate enzymatic halogenation of tryptophan, we deduced the spot as a

Table 2

Confidence	PSM Ambiguity	Annotated Sequence	Protein Accessions	MH+ [Da]	RT [min]	Percolator q-Value	Percolator PEP
A							
High	Unambiguous	[K].LSQEEVSR.[T]	*Dn*HaI	947.4795398	9.3281693	0	0.1959
High	Unambiguous	[R].REDTLNIANFTNDAINGYVPVVQR.[G]	*Dn*HaI	2719.385349	43.7476283	0	0.1052
High	Unambiguous	[R].FVDLDETFLTHGFKVK.[K]	*Dn*HaI	1895.991598	41.85278393	0	0.0566
High	Unambiguous	[K].QAAVLKKmETAmLVEnVAAPADLAK.[T]	*Dn*HaI	2645.380758	41.01146207	0	0.194
High	Unambiguous	[K].NIANWGYWQSGGEYAVGTER.[Q]	*Dn*HaI	2258.024218	37.80150682	0	0.06502
High	Unambiguous	[K].NIANWGYWQSGGEYAVGTER.[Q]	*Dn*HaI	2258.024218	37.79974257	0	0.04484
High	Unambiguous	[K].METAMLVENVAAPADLAK.[T]	*Dn*HaI	1873.941088	36.8099236	0	0.03829
High	Unambiguous	[K].METAMLVENVAAPADLAK.[T]	*Dn*HaI	1873.941088	36.80800432	0	0.006857
High	Unambiguous	[K].mETAMLVENVAAPADLAK.[T]	*Dn*HaI	1889.936571	34.28906221	0	0.07315
High	Unambiguous	[K].METAmLVENVAAPADLAK.[T]	*Dn*HaI	1889.936571	34.28906221	0	0.05521
High	Unambiguous	[R].LGEHNIDVLEGNEQFINAAK.[I]	P00761	2211.102587	33.91689306	0	0.06304
High	Unambiguous	[R].LGEHNIDVLEGNEQFINAAK.[I]	P00761	2211.102587	33.91515578	0	0.153
High	Unambiguous	[R].LGEHNIDVLEGNEQFINAAK.[I]	P00761	2211.10709	33.84775446	0	0.002097876
High	Unambiguous	[K].KmETAmLVENVAAPADLAK.[T]	*Dn*HaI	2034.024875	29.24587828	0	0.03541
High	Unambiguous	[K].AMVDPEGLSEDEMR.[I]	*Dn*HaI	1578.677904	28.33588051	0	0.03784
High	Unambiguous	[K].AMVDPEGLSEDEMR.[I]	*Dn*HaI	1578.677904	28.33394892	0	0.009701
High	Unambiguous	[K].AmVDPEGLSEDEMR.[I]	*Dn*HaI	1594.675218	24.85086755	0	0.08903
High	Unambiguous	[R].VATVSLPR.[S]	P00761	842.5098132	20.90633053	0	0.09483
High	Unambiguous	[K].LSQEEVSR.[T]	*Dn*HaI	947.4795398	9.329454972	0	0.05087
High	Unambiguous	[R].ILNTIR.[A]	*Dn*HaI	729.4613513	14.81710152	0	0.1499
High	Unambiguous	[R].ILNTIRAR.[Q]	*Dn*HaI	956.5991552	14.84533552	0	0.08581
High	Unambiguous	[R].ILNTIR.[A]	*Dn*HaI	729.4612292	15.2419809	0	0.1501
High	Unambiguous	[R].ILNTIR.[A]	*Dn*HaI	729.4612292	15.24334753	0	0.0928
High	Unambiguous	[K].LSSPATLNSR.[V]	P00761	1045.564989	15.41150922	0	0.08484
High	Unambiguous	[R].REDTLNIANFTNDAINGYVPVVQR.[G]	*Dn*HaI	2719.385349	43.74923827	0	0.0642

Table 2 (continued)

Confidence	PSM Ambiguity	Annotated Sequence	Protein Accessions	MH+ [Da]	RT [min]	Percolator q-Value	Percolator PEP
High	Unambiguous	[K].LSSPATLNSR.[V]	P00761	1045.56389	15.42356014	0	0.1526
High	Unambiguous	[K].LSSPATLNSR.[V]	P00761	1045.564379	15.80155669	0	0.1028
High	Unambiguous	[R].GTLGLVK.[I]	*Dn*Hal	687.4390735	17.51289474	0	0.1959
High	Unambiguous	[R].VATVSLPR.[S]	P00761	842.5099963	19.77855468	0	0.1272
High	Unambiguous	[R].VATVSLPR.[S]	P00761	842.5085925	20.15119059	0	0.1469
High	Unambiguous	[R].VATVSLPR.[S]	P00761	842.5091418	20.49989502	0	0.1959
High	Unambiguous	[R].VATVSLPR.[S]	P00761	842.5091418	20.50056064	0	0.09931
High	Unambiguous	[K].LSSPATLNSR.[V]	P00761	1045.564379	15.80023584	0	0.04713
High	Unambiguous	[K].AMVDPEGLSEDEMRILNTIR.[A]	*Dn*Hal	2289.122654	45.37146101	0	0.1876

Protein FDR Confidence	Accession	Description	*Dn*Hal	Exp. q-value	Sum PEP Score	Coverage	#AAs	MW [kDa]	Score Sequest HT	#Peptides Sequest HT
B										
High	P04842	Alcohol oxidase 1	FALSE	0	391.5070203	73.454	663	73.851	1342.036795	52
High	C4R702	Alcohol oxidase 2	FALSE	0	277.8278489	52.79035	663	73.905	1002.400932	39
High	O74192	Dihydroxyacetone kinase	FALSE	0	130.0889787	41.28289	608	65.272	414.2974367	31
High	P53623	Heat shock protein 70 2	FALSE	0	126.2095839	28.77138	643	70.03	420.8407903	23
High	P53421	Heat shock protein 70 1	FALSE	0	123.8546889	22.7907	645	70.095	422.7354431	17
High	Q9C124	Isocitrate lyase	FALSE	0	25.44793779	21.63636	550	61.705	56.61623883	10
High	P00761	Trypsin	FALSE	0	12.72390281	16.45022	231	24.394	33.5931778	3
High	Q92263	Glyceraldehyde-3-phosphate dehydrogenase	FALSE	0	9.337924925	15.01502	333	35.595	21.83858788	3
High	P30263	Peroxisomal catalase	FALSE	0	7.979641425	3.944773	507	57.812	25.33568239	3
High	C4R2P3	Methionine aminopeptidase 2	FALSE	0	4.28009063	10.04464	448	50.173	10.00062346	3
High	P45353	Histidine biosynthesis trifunctional protein	FALSE	0	3.521617246	2.731591	842	91.884	8.588482141	2
High	P06834	Dihydroxyacetone synthase	FALSE	0	3.271606606	2.676056	710	78.68	8.153642893	2

Table 2 (continued)

Confidence	PSM Ambiguity	Annotated Sequence			Protein Accessions	MH+ [Da]	RT [min]		Percolator q-Value	Percolator PEP
High	Q9HG01	78 kDa glucose-regulated protein homolog	FALSE	0	2.9476909	2.406015	665	73.177	4.740409136	1
High	Q874B9	Elongation factor 2	FALSE	0	2.801893001	1.068884	842	93.41	5.038891077	1
High	A3GH91	ATP-dependent RNA helicase	FALSE	0	2.587203571	3.2	500	55.675	6.252739668	1
High	Q9P4D1	DBP5 Actin	FALSE	0	2.336110701	4.255319	376	41.665	4.091575623	1
High	A3LQ01	ATP-dependent RNA helicase	FALSE	0	2.237321436	2.318393	647	68.857	2.720205545	1
High	P34736	DED1 Transketolase	FALSE	0	1.7070797	1.477105	677	72.76	6.950209856	1
High	Q5H7C3	Sorting nexin-4	FALSE	0	1.524619407	3.782148	661	76.862	0	1
High	Q7ZA46	Phosphoglycerate kinase	FALSE	0	1.452717692	4.326923	416	44.09	0	1
High	C4QXN2	MICOS complex subunit	FALSE	0	1.450016389	1.768173	509	57.095	4.117641211	1
Low	O93939	Glucan 1,3-beta-glucosidase 1	FALSE	0.08	1.251967106	1.204819	498	58.027	7.178422451	1
Protein	Accession	Description	*Dn*Hal	Exp.	Sum	Coverage	# AAs	MW	Score Sequest	# Peptides
FDR Confidence				q-value	PEP Score			[kDa]	HT	Sequest HT
C										
High	P04842	Alcohol oxidase 1 (*Komagataella phaffii* (strain GS115 / ATCC 20864))	FALSE	0	406.4327456	76.31975867	663	73.851	1419.833611	59
High	C4R702	Alcohol oxidase 2 (*Komagataella phaffii* (strain GS115 / ATCC 20864))	FALSE	0	262.8842063	62.14177979	663	73.905	961.9134561	47
High	O74192	Dihydroxyacetone kinase (*Komagataella pastoris*)	FALSE	0	151.8021475	49.67105263	608	65.272	488.7799518	35
High	P53421	Heat shock protein 70 1 (*Pichia angusta*)	FALSE	0	104.8499373	22.79069767	645	70.095	368.6018039	18
High	P53623	Heat shock protein 70 2 (*Pichia angusta*)	FALSE	0	101.4570394	28.77138414	643	70.03	350.9738773	22
High	P30263	Peroxisomal catalase (*Pichia angusta*)	FALSE	0	13.31867931	5.917159763	507	57.812	47.12409115	4

Table 2 (continued)

Confidence	PSM Ambiguity	Annotated Sequence			Protein Accessions	MH+ [Da]	RT [min]		Percolator q-Value	Percolator PEP
High	*Dn*Hal	*Dn*Hal	TRUE	0	18.6526514	18.63247863	585	63.733	47.00790763	8
High	*Dn*Hal	*Dn*Hal	TRUE	0	18.6526514	19.39501779	562	61.014	47.00790763	8
High	Q9C124	Isocitrate lyase (*Komagataella pastoris*)	FALSE	0	13.96684081	12	550	61.705	35.38488269	8
High	P45353	Histidine biosynthesis trifunctional protein (*Komagataella pastoris*)	FALSE	0	7.250202929	7.363420428	842	91.884	24.98744082	4
High	P00761	Trypsin	FALSE	0	10.07062645	25.10822511	231	24.394	23.63299155	4
High	Q874B9	Elongation factor 2 (*Komagataella pastoris*)	FALSE	0	9.555592921	7.482185273	842	93.41	17.0720799	5
High	P06834	Dihydroxyacetone synthase (*Pichia angusta*)	FALSE	0	4.818732484	2.957746479	710	78.68	14.06938839	3
High	Q92263	Glyceraldehyde-3-phosphate dehydrogenase (*Komagataella pastoris*)	FALSE	0	4.793116402	8.708708709	333	35.595	12.56598032	2
High	O93939	Glucan 1,3-beta-glucosidase 1 (*Wickerhamomyces anomalus*)	FALSE	0	1.360613131	1.204819277	498	58.027	10.54587758	1
High	C4R2P3	Methionine aminopeptidase 2 (*Komagataella phaffii* (strain GS115 / ATCC 20864))	FALSE	0	3.476817963	3.571428571	448	50.173	7.454492092	2
High	A3GH91	ATP-dependent RNA helicase DBP5 (*Scheffersomyces stipitis* (strain ATCC 58785 / CBS 6054 / NBRC 10063 / NRRL Y-11545))	FALSE	0	2.234033575	3.2	500	55.675	6.096784353	1
High	P33677	Formate dehydrogenase (*Pichia angusta*)	FALSE	0	2.20697834	3.867403315	362	39.886	5.400453329	1
High	Q9HG01	78 kDa glucose-regulated protein homolog (*Pichia angusta*)	FALSE	0	1.594141601	2.406015038	665	73.177	2.933979988	1

Table 2 (continued)

Confidence	PSM Ambiguity	Annotated Sequence			Protein Accessions	MH+ [Da]	RT [min]		Percolator q-Value	Percolator PEP
High	A3LUF7	Lon protease homolog 2, peroxisomal (*Scheffersomyces stipitis* (strain ATCC 58785 / CBS 6054 / NBRC 10063 / NRRL Y-11545))	FALSE	0	1.587371479	2.118644068	1180	131.971	2.578657627	1

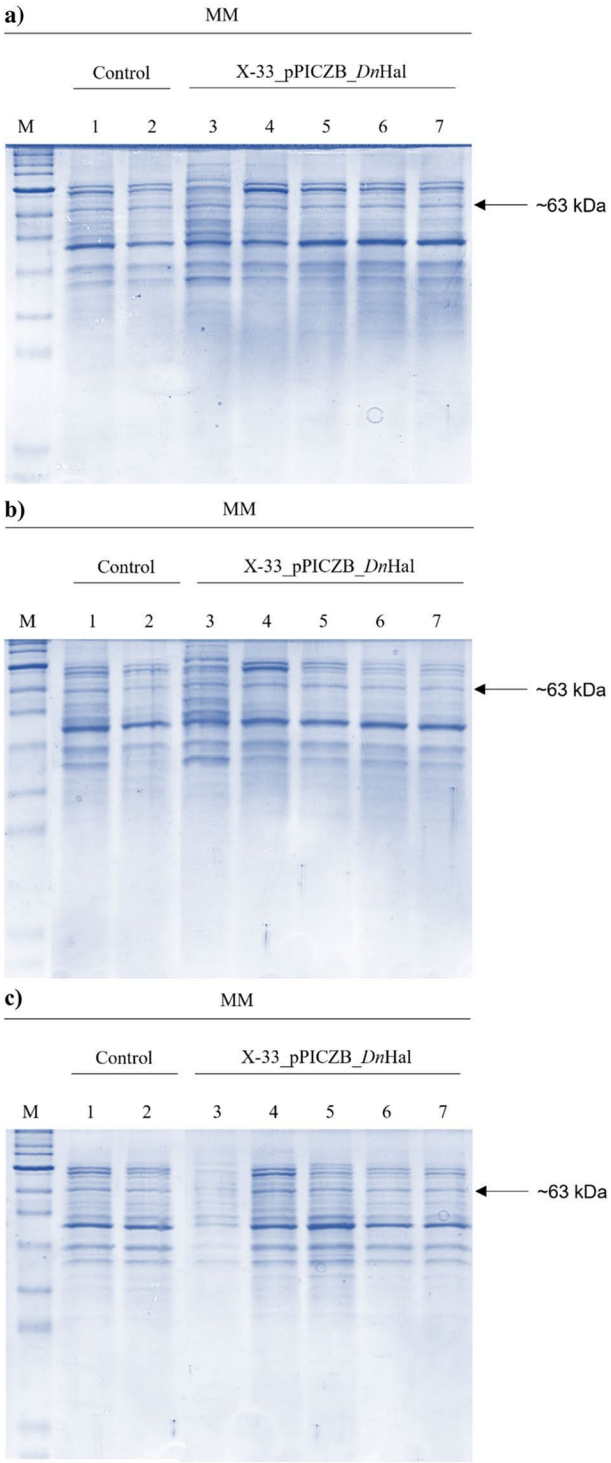


Fig. 4 Expression analysis of recombinant *Dn*Hal in P. pastoris X-33. **a** Expression profiling for the crude enzyme. M: prestained protein ladder, broad range (10–230 kDa) (NEB), 1–4: the X-33_pPICZB_*dnhal* cells that were induced with 1.0% methanol for every 24 h, and incubated until 4 days. **b** SDS-PAGE for the purified *Dn*Hal. M: prestained protein ladder, broad range (10–230 kDa) (NEB), 1: the purified *Dn*Hal that was induced with 1.0% methanol for every 24 h and incubated for 2 days

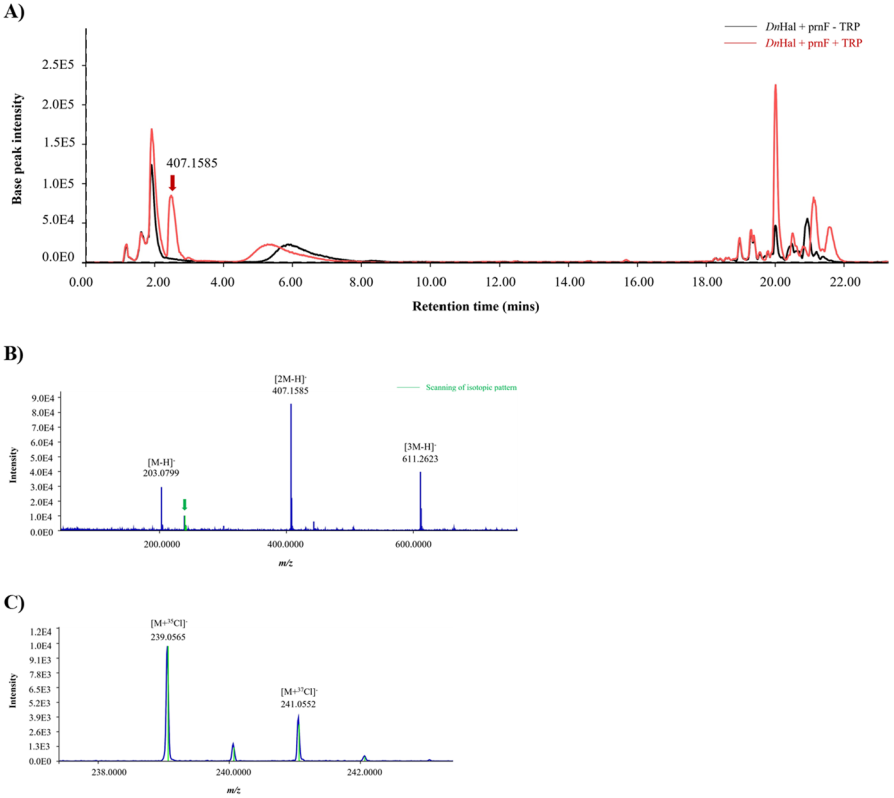


Fig. 5 Enzymatic halogenation of tryptophan validated by LC-MS analysis. (**A**) Overlayed chromatogram of the halogenation assay with a control reaction containing all components without tryptophan. The black line signifies the separation of the control reaction, while the red line represents the separation of halogenation products. The peak containing the halogenated products was marked with a red arrow. (**B**) Mass spectrum of the scanned isotopic pattern. The spectra showing an isotopic pattern for chlorinated tryptophan were highlighted in green and visualized in a larger image in panel (**C**)

representative of the chlorinated product, chloro-tryptophan. In terms of chromatographic separation, we found that the tryptophan retained longer in the stationary phase and that the installation of chlorine changed the polarity of the predicted halogenated product, making it more polar than the tryptophan and eluted faster in the chromatographic separation.

The purified enzymes (Figure S6) was used in the halogenation assay of tryptophan which was then analyzed via LC-MS analysis. The full scanning of LC-MS negative ion mass spectra of chloro-tryptophan displays a putative product peak with a base peak of 407.1585 m/z eluted at 2.48 minutes (Fig. 5). Two molecular ion adducts [MH]- and [3M-H]- at m/z 203.0799 and 611.2623 were found, respectively. These results were supported by the isotopic pattern for the chlorinated product (highlighted with a green line) in Fig. 5. A typical 3:1 isotopic ratio was visible at m/z 239.0565 and 241.0552, which correspond to the molecular ions of [M+35Cl]- and [M+37Cl]- proposing that chloro-tryptophan was generated from the halogenation reaction (Fig. 5). It is worth noting that these mass spectra data were comparable to the exact monoisotopic masses of molecular ion adducts often observed in the ESI mass spectra [25].

Halogenation assay with methyl haematommate as substate gave the presence of putative peak at 14.83 min which showed the presence of doublet molecular ion [C10H10O5-2H+Cl]- with 1.99705 Da difference at *m/z* 243.0074 and 245.0025 and was consistent with isotopic pattern for chlorinated product (Fig. 6). It is thus suggesting that *Dn*Hal is also active towards monoaromatic phenolic substrate and generates the respective chlorinated product, chloro-methyl haematommate. An additional control experiment was run with the absence of substrate in the reaction mixture. The LC-MS scanning showed no traces of chlorinated products, confirming that *Dn*Hal is responsible for the halogenation of tryptophan and methyl haematommate.

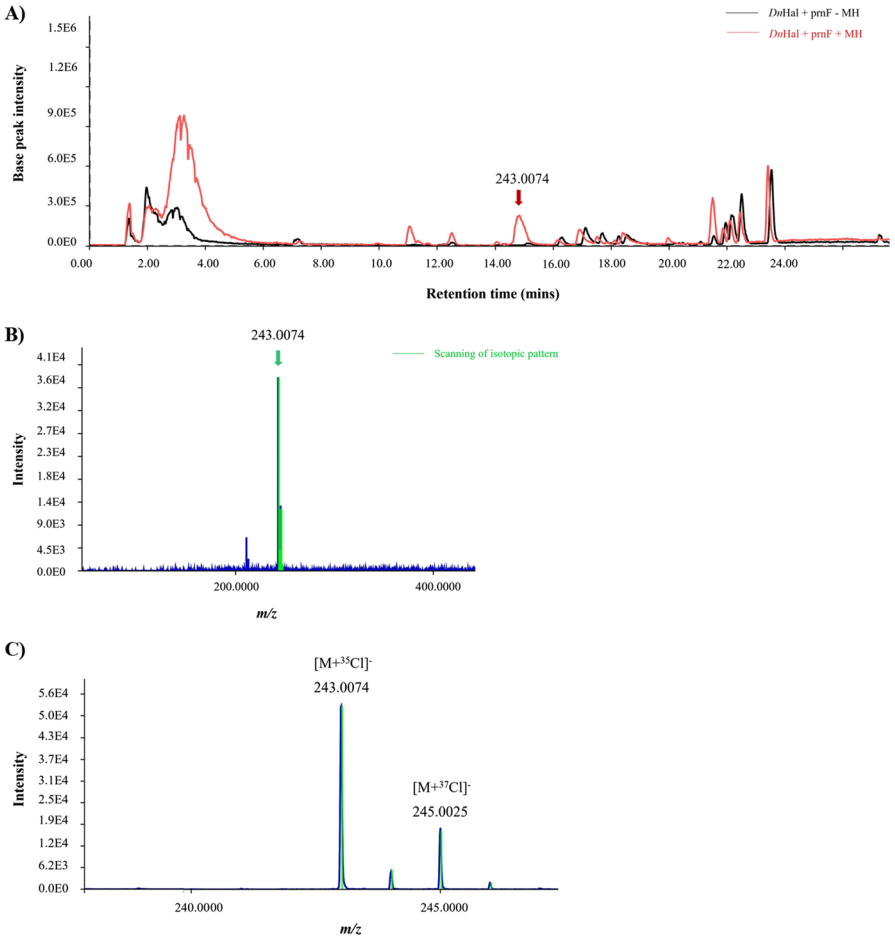


Fig. 6 Enzymatic halogenation of methyl haematommate validated by LC-MS analysis. (**A**) Overlayed chromatogram of the halogenation assay with a control reaction containing all components without methyl haematommate. The black line signifies the separation of the control reaction, while the red line represents the separation of halogenation products. The peak containing the halogenated products was marked with a red arrow. (**B**) Mass spectrum of the scanned isotopic pattern. The spectra showing an isotopic pattern for chloro-methyl haematommate were highlighted in green and visualized in a larger image in panel (**C**)

Discussion

Characterization of DnHal and Its Involvement in Enzymatic Halogenation of Tryptophan and Mono Aromatics

These findings provide a basis for the exploitation of omics data to identify novel genes from unculturable or difficult-to-culture fungal species such as lichens. Although halogenating activity has previously been reported in other lichen species, no experimental data has hitherto been reported for F-Hal enzymes from lichens or lichen-forming fungi, therefore limiting our understanding of the halogenation mechanism employed by these eukaryotic sources. Furthermore, the current F-Hal enzyme family is saturated with bacterial tryptophan halogenase, whereas the function of eukaryotic F-Hals has only been explored limitedly. Our findings show that the *dnhal* gene exhibits novelty in sequence and its functional role as well. The identification of *dnhal* from *Dirinaria* transcripts and its putative function that is closely related to the tryptophan halogenases shows that it is likely involved in the enzymatic modification of tryptophan. Because there is as of yet no report on the biochemical characterization of the lichen halogenases, except the putatively annotated CU-PKS-4(4) gene, hence, the present study provides for the first functional characterization of a lichen F-Hal enzyme.

Due to the consensus and conservation sequences of *Dn*Hal, we note here that *Dn*Hal is mechanistically conserved with other F-Hal enzymes. We propose that once FAD forms contact with the GxGxxG motif sequence in the FAD-binding domain, *Dn*Hal oxidizes the reduced cofactor, and in the presence of chloride ions, the HOCl will be generated and later guided alongside the 10Å tunnel to interact with the conserved lysine residue. Herein, the conserved lysine residue will position HOCl close to the active site, or be covalently bound with HOCl to generate a stable halogenating species, lysine chloramine. Either way, the halogenating species is proposed to interact with phenolic substrates to generate the halogenated products [2]. The presence of the WxWxIP motif distinguishes *Dn*Hal from other monooxygenase enzymes. It was proposed that these conserved sequences operate as a barrier to block the binding of substrate closer to *Dn*Hal [54]. In conjunction with the investigation of the crystal structure of tryptophan F-Hals that displays a longer distance connecting the isoalloxazine ring of the flavin intermediate with the substrate [56], this motif is suggested hampering direct modification of the substrate.

The phylogenetic study demonstrated that the *Dn*Hal enzyme shares an ancestor with other characterized fungal F-Hals, and we observed a clear divergence between tryptophan and non-tryptophan F-Hal enzymes. It is noticeable that the free-standing phenolic compounds are more susceptible to the enzymatic halogenation of *Dn*Hal.

Nevertheless, the inability to bridge the *Dn*Hal enzyme with its natural substrate has been one of the main obstructions in our attempt to functional characterize *Dn*Hal as a potential candidate for halogenation biocatalysts. In light of the presence of the tryptophan halogenase domain in the *Dn*Hal sequence, and by taking advantage of chemical features of tryptophan that possess electron-rich indole moiety susceptible to electrophilic halogenation reaction, enzymatic halogenation of *Dn*Hal is assumed active against tryptophan. Accordingly, the functional study proceeds with tryptophan as a substrate for *Dn*Hal. The LC–MS analyses displayed a presence of the chlorinated product, therefore supporting our preliminary hypothesis suggesting that *Dn*Hal is responsible for introducing halogen into the indole moiety of tryptophan. Significantly, this study presents the first report of lichen F-Hal enzyme showing activity in the halogenation of tryptophan. Often, the tryptophan

F-Hals display high regioselectivity. The enzymatic chlorination of tryptophan mediated by RebH enzyme demonstrated high selectivity at the C7 of the indole moiety [57]. We notice that genetic modification of determinant residues in the tryptophan 6 halogenase SatH shifts the regioselectivity in installing the halogen into C5 of the indole moiety [58]. Considering the high regioselectivity of the tryptophan F-Hals, we propose that enzymatic halogenation of *Dn*Hal presumably occurred in a selective manner. This finding lays the foundation to investigate the regioselectivity of DnHal and other lichen F-Hal enzymes.

The in-vitro halogenation assays of DnHal with methyl haematommate proved the function of *Dn*Hal as the phenolic F-Hal, which is prone to install chlorine onto the backbone structure of the orsellinic derivatives. Even though there is still unclear information on the natural substrate of *Dn*Hal, the phylogenetic classification with phenolic F-Hals that are active in the late-stage enzymatic halogenation suggests that methyl haematommate is the probable substrate for *Dn*Hal. Also, by considering that the methyl haematommate is the product of transesterification of atranorin, a well documented depsides from many lichens [59], we deduce that *Dn*Hal is also active against compounds similar to its natural substrate. A similar finding was also observed in the behaviour of the fungal Rdc2/RadH that is also able to incorporate halogen into moieties that are similar to their natural substrates [7, 8]. Hence, our findings provide a guidance to expand the substrate scope significant in improving the biosynthetic halogenation processes.

Considering theoretical and experimental findings, we deduced here that the *Dn*Hal possesses promiscuous enzyme activity, either it is active towards tryptophan or non-tryptophan substrates, mainly the free-standing phenolic substrates. Thus, it is the best option as a biocatalyst involved in the chemical modification of tryptophan, producing halogenated building blocks for chemical diversity. The application of halogenated tryptophan as a building block in the chemosynthetic application [5] highlighted the importance of the F-Hal enzymes for synthetic applications. Therefore, the ability of *Dn*Hal as a biocatalyst could aid in expanding the available biocatalytic toolbox for the generation of halogenated intermediates that can be utilized for chemical cyclization or derivatization. Regardless, more experimental data are needed, especially on the substrate tolerance, preference and selectivity of *Dn*Hal towards indole derivatives and a broader range of phenolic or more complex substrates.

Production of *Dn*Hal in *P. pastoris*

The methylotrophic yeast *P. pastoris* is selected as a host to produce the recombinant enzyme *Dn*Hal. Based on our preliminary sequence characterization, four generations of disulphide bond (personal communication) were predicted, suggesting the need for post-translational modifications to produce a functional *Dn*Hal enzyme. The *Pichia* expression system provides appropriate folding in the endoplasmic reticulum [60]; therefore, *P. pastoris* is the best host to produce a soluble *Dn*Hal enzyme. Comparable to the ~ 63 kDa band observed from the SDS-PAGE analysis (Fig. 4), we report that the soluble eukaryotic enzyme is successfully expressed in the *P. pastoris* expression system.

Given the characteristics of the methylotrophic *P. pastoris*, we use the methanol utilization pathway to induce the expression of *Dn*Hal. Under the regulation of a tight and inducible *AOX*1 promoter, the presence of methanol induces the *AOX*1 gene and promotes a high amount of AOX enzyme, substantially producing a high amount of recombinant enzymes. Hence, an expression vector containing the *AOX*1 promoter sequence is selected as a vector to express the *Dn*Hal enzyme. We also predicted *Dn*Hal as an intracellular enzyme

(personal communication); hence, we propose that intracellular expression is the best strategy to produce the recombinant enzyme. Accordingly, the pPICZB expression was selected as a vector to express the *Dn*Hal enzyme.

In preparing the expression construct, we also integrated the consensus Kozak sequence between the restriction site *Kpn*I and the start codon of the *dnhal* gene. We prepared the gene to include the consensus sequence mainly because including the Kozak consensus provides efficient translation initiation for most intracellular expression processes [61]. It is acceptable that modifying the gene to the preferred codon usage of *P. pastoris* necessitates high levels of heterologous expression [62, 63]. Nevertheless, our expression profiling contradicts the statement, presenting a low level of the expressed recombinant *Dn*Hal enzyme.

Among the significant feature present in the pPICZB vector is designed to facilitate the construction of expression vectors with multiple copies of the expression cassette, and additional unique restriction sites that permit linearization of the expression vector at the *AOX*1 locus for efficient integration into the *Pichia* genome. The presence of a band with a theoretical mass of 5035 bp on the agarose gel indicates that the 1707 bp *dnhal* gene was successfully cloned to the pPICZB expression vector (Figure S3). Later, the pPICZB_*dnhal* was linearized with *Bst*XI and was integrated into the *AOX*1 locus of the *Pichia* genome (Figure S3). It is important to note that integrating the expression cassette into its genome is crucial to generating stable transformants [64]. We report that our expression cassette was successfully transformed into the yeast, and the generated X-33_pPICZB_*dnhal* cells are stable for expression purposes.

After the generation of the transformant is accomplished, the cells are cultured in a complex medium containing glycerol to enhance the high yield production of cell biomass. The growth curve in Fig. 3 suggests that cultivating the cells for 44 h in the BMGY medium is sufficient to accumulate maximum biomass for the induction process. Similar to the previous finding, the *Pichia* growth rate is higher when grown in a complex medium [65]. Literature findings suggest that the *AOX*1 promoter is strongly repressed when *P. pastoris* is grown in a medium containing a glycerol carbon source. The promoter is de-repressed upon depletion of the carbon source and will be fully induced in the presence of methanol [40]. The inductive media MM was chosen to induce the *Dn*Hal enzyme. We observed a low level of total intracellular proteins from the MM culture, suggesting that carbon starvation might affect the expression level. Even though *Pichia*'s protein expression system depends on methanol consumption [60], increasing the amount of methanol does not affect the expression level of the *Dn*Hal enzyme. We observe a similar finding from the literature; low expression of tumour necrosis factor-alpha is detected, typically less than 1% of the total intracellular proteins [53]. It is thus suggested that other factors might take a role in manipulating the expression level of our *Dn*Hal enzyme. Therefore, several strategies can be implemented in the future study, focusing on the optimization to produce high amount of the *Dn*Hal enzyme; by opting to utilize the constitutive promoter, or by changing into an extracellular expression system, or turning into scalable production in fermenter cultures.

Conclusion

The experimental study provides the first biochemical characterization of a lichens F-Hal enzyme. Significantly, the basis information paves the way towards exploiting the potential of *Dn*Hal and the F-Hal enzyme class as a whole, as potential biocatalysts for bio- and chemosynthetic applications. Besides, the successful expression of the *Dn*Hal enzyme has

rendered the methylotrophic yeast *P. pastoris* one of the most suitable and robust protein production host systems. It is thus suggested that the *P. pastoris* serves as an emerging platform producing functional F-Hal enzymes from eukaryotic sources, and this study provides basis for subsequent improvement producing high-volume production of the eukaryotic F-Hal enzymes.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12010-022-04304-w.

Acknowledgements We thank Prof. Dr. Karl-Heinz van Pèe (TU Dresden, Germany) for providing the plasmid pET28B_prnF-BL915. We would like to thank the Malaysia Genome & Vaccine Institute for providing computational facilities.

Author Contribution Nurain Shahera Hasan, Jonathan Guyang Ling, Farah Diba Abu Bakar and Rozida Mohd Khalid conceived, conceptualized and designed the project and interpreted the data. Nurain Shahera Hasan planned and performed all experiments and analyses. Jonathan Guyang Ling, Mohd Faizal Abu Bakar and Wan Muhammad Khairulikhsan Wan Seman supported the investigation by performing parts of the experiments and analysis. Abdul Munir Abdul Murad and Farah Diba Abu Bakar contributed reagents/materials/analysis tools and contributed ideas that helped shape the design and implementation of the research. Nurain Shahera Hasan and Rozida Mohd Khalid wrote the original manuscript. Nurain Shahera Hasan, Jonathan Guyang Ling and Rozida Mohd Khalid reviewed and edited the manuscript. Rozida Mohd Khalid and Farah Diba Abu Bakar oversaw and supervised the project. All authors discussed the results and reviewed the final manuscript.

Funding This work was supported by Fundamental Research Grant Scheme (FRGS) from the Ministry of Higher Education Malaysia (FRGS/1/2018/STG01/UKM/02/7).

Data Availability It is thus suggested that the *P. pastoris* serves as an emerging platform producing functional F-Hal enzymes from eukaryotic sources, even though several improvements are required to raise the expression level and this study provides basis for subsequent improvement producing high-volume production of the eukaryotic F-Hal enzymes.

Declarations

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The authors declare no competing interests.

References

1. Baker Dockery, S. A., & Narayan, A. R. H. (2019). Flavin-dependent biocatalysts in synthesis. *Tetrahedron, 75*(9), 1115–1121.
2. Andorfer, M. C., & Lewis, J. C. (2018). Understanding and improving the activity of Flavin-dependent halogenases via random and targeted mutagenesis. *Annual Review of Biochemistry, 5*(10), 1285–1297.
3. Alkhalaf, L. M., & Ryan, K. S. (2015). Biosynthetic manipulation of tryptophan in bacteria: Pathways and mechanisms. *Chemistry and Biology, 22*(3), 317–328.
4. Frese, M., & Sewald, N. (2015). Enzymatic halogenation of tryptophan on a gram scale. *Angewandte Chemie - International Edition, 54*(1), 298–301.
5. Gruß, H., & Sewald, N. (2020). Late-stage diversification of tryptophan-derived biomolecules. *Chemistry - A European Journal, 26*(24), 5328–5340.

6. Romero, E., Jones, B. S., Hogg, B. N., Rué Casamajo, A., Hayes, M. A., Flitsch, S. L., Turner, N. J., & Schnepel, C. (2021). Enzymatic late-stage modifications: Better late than never. *Angewandte Chemie - International Edition, 60*(31), 16824–16855.
7. Zeng, J., & Zhan, J. (2010). A novel fungal flavin-dependent halogenase for natural product biosynthesis. *ChemBioChem, 11*(15), 2119–2123.
8. Menon, B. R. K., Brandenburger, E., Sharif, H. H., Klemstein, U., Shepherd, S. A., Greaney, M. F., & Micklefield, J. (2017). RadH: A versatile halogenase for integration into synthetic pathways. *Angewandte Chemie - International Edition, 56*(39), 11841–11845.
9. Dong, C., Flecks, S., Unversucht, S., Haupt, C., van Pée, K.-H., & Naismith, J. H. (2005). Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science, 309*(5744), 2216–2219.
10. Zehner, S., Kotzsch, A., Bister, B., Süssmuth, R. D., Méndez, C., Salas, J. A., & Van Pée, K.-H. (2005). A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chemistry and Biology, 12*(4), 445–452.
11. Yeh, E., Blasiak, L. C., Koglin, A., Drennan, C. L., & Walsh, C. T. (2007). Chlorination by a long-lived intermediate in the mechanism of flavin-dependent halogenases. *Biochemistry, 46*(5), 1284–1292.
12. Buedenbender, S., Rachid, S., Muller, R., & Schulz, G. E. (2009). Structure and action of the myxobacterial chondrochloren halogenase CndH: A new variant of Fad-dependent halogenases. *Journal of Molecular Biology, 385*(2), 520–530.
13. Zeng, J., Lytle, A. K., Gage, D., Johnson, S. J., & Zhan, J. (2013). Specific chlorination of isoquinolines by a fungal flavin-dependent halogenase. *Bioorganic and Medicinal Chemistry Letters, 23*(4), 1001–1003.
14. Cochereau, B., Meslet-Cladière, L., Pouchus, Y. F., Grovel, O., & Roullier, C. (2022). Halogenation in fungi: What do we know and what remains to be discovered? *Molecules, 27*(10), 3157.
15. Azmi, A. A., Nadiah, N., Rosandy, A. R., Alwi, A., Kamal, N., Mohd. Khalid, R., & Abu Bakar, M. (2021). Antimicrobial activity and LC-MS data comparison from Lichen *Parmotrema praesorediosum* in Bangi, Selangor, Malaysia. *Sains Malaysiana, 50*(2), 383–393.
16. Bharudin, I., Abdul Rahim, S. N., Abu Bakar, M. F., Ibrahim, S. N., Kamaruddin, S., Latif, M. T., Samsudin, M. W., Abdul Murad, A. M., & Abu Bakar, F. D. (2018). De novo transcriptome resources of the lichens, *Dirinaria* sp. UKM-J1 and UKM-K1 collected from Jerantut and Klang. *Malaysia. Data in Brief, 19*, 2416–2419.
17. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology, 29*(7), 644–652.
18. Jaina, M., Sara, C., Lowri, W., Matloob, Q., Gustavo, A. S., Erik, L. L. S., Silvio, C. E. T., Lisanna, P., Shriya, R., Lorna, J. R., Robert, D. F., & Alex, B. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research, 49*(D1), D412–D419. https://doi.org/10.1093/nar/gkaa913
19. Philip, J., David, B., Hsin-Yu, C., Matthew, F., Weizhong, L., Craig, M., Hamish, M., John, M., Alex, M., Gift, N., Sebastien, P., Antony, F. Q., Amaia, S.-V., Maxim, S., Siew-Yit, Y., Rodrigo, L., & Sarah, H. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics, 30*(9), 1236–1240.
20. Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology, 302*(1), 205–217.
21. Xavier, R., & Patrice, G. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research, 42*(W1), W320–W324. https://doi.org/10.1093/nar/gku316
22. Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution, 35*(6), 1547–1549.
23. Ivica, L., & Peer, B. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research, 49*(W1), W293–W296. https://doi.org/10.1093/nar/gkab301
24. Hammond, J. B. & Kruger, N. J. (1988). in The Bradford method for protein quantitation, Chapter 2 (Walker J. M.,ed.), Humana Press, 10(I), pp. 25–32.
25. High, E., Huang, N., Siegel, M. M., Kruppa, G. H., & Laukien, F. H. (1999). Automation of a Fourier transform ion cyclotron resonance mass spectrometer for acquisition, analysis, and e-mailing of high-resolution exact-mass electrospray ionization mass spectral data. *Journal of the American Society for Mass Spectrometry, 10*(99), 1166–1173.
26. Abdel-Hameed, M., Bertrand, R. L., Piercey-Normore, M. D., & Sorensen, J. L. (2016). Identification of 6-hydroxymellein synthase and accessory genes in the lichen *Cladonia uncialis*. *Journal of Natural Products, 79*(6), 1645–1650.
27. Stefania, M., & Van, B. W. J. H. (2013). The flavin monooxygenases in Handbook of Flavoproteins, vol. 2. In R. M., Hille, M., Susan, & B., Palfey (Eds.), *Complex flavoproteins, dehydrogenases and physical methods* (pp. 51–72). De Gruyter.

28. Paul, C. E., Guarneri, A. & Berkel, W. J. H. (2021). Flavoprotein monooxygenases and halogenases. *Flavin-Based Catalysis*, 169–199.
29. Shepherd, S. A., Karthikeyan, C., Latham, J., Struck, A. W., Thompson, M. L., Menon, B. R. K., Styles, M. Q., Levy, C., Leys, D., & Micklefield, J. (2015). Extending the biocatalytic scope of regiocomplementary flavin-dependent halogenase enzymes. *Chemical Science, 6*(6), 3454–3460.
30. Van Pée, K.-H., & Patallo, E. P. (2006). Flavin-dependent halogenases involved in secondary metabolism in bacteria. *Applied Microbiology and Biotechnology, 70*(6), 631–641.
31. Zeng, J., & Zhan, J. (2011). Characterization of a tryptophan 6-halogenase from *Streptomyces toxytricini*. *Biotechnology Letters, 33*(8), 1607–1613.
32. Menon, B. R. K., Latham, J., Dunstan, M. S., Brandenburger, E., Klemstein, U., Leys, D., Karthikeyan, C., Greaney, M. F., Shepherd, S. A., & Micklefield, J. (2016). Structure and biocatalytic scope of thermophilic flavin-dependent halogenase and flavin reductase enzymes. *Organic and Biomolecular Chemistry, 14*(39), 9354–9361.
33. Neubauer, P. R., Widmann, C., Wibberg, D., Schröder, L., Frese, M., Kottke, T., Kalinowski, J., Niemann, H. H., & Sewald, N. (2018). A flavin-dependent halogenase from metagenomic analysis prefers bromination over chlorination. *PLoS ONE, 13*(5), e0196797. https://doi.org/10.1371/journal.pone.0196797
34. Neubauer, P. R., Blifernez-Klassen, O., Pfaff, L., Ismail, M., Kruse, O. & Sewald, N. (2021). Two novel, flavin-dependent halogenases from the bacterial consortia of *Botryococcus braunii* catalyze mono-and dibromination. *Catalysts*, 11(4). https://doi.org/10.3390/catal11040485
35. Seibold, C., Schnerr, H., Rumpf, J., Kunzendorf, A., Hatscher, C., Wage, T., Ernyei, A. J., Dong, C., Naismith, J. H., & van Pée, K.-H. (2006). A flavin-dependent tryptophan 6-halogenase and its use in modification of pyrrolnitrin biosynthesis. *Biocatalysis and Biotransformation, 24*(6), 401–408.
36. Ryan, K. S. (2011). Biosynthetic gene cluster for the cladoniamides, bis-indoles with a rearranged scaffold. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0023694
37. Liu, Z., Ma, L., Zhang, L., Zhang, W., Zhu, Y., Chen, Y., Zhang, W., & Zhang, C. (2019). Functional characterization of the halogenase SpmH and discovery of new deschloro-tryptophan dimers. *Organic and Biomolecular Chemistry*. https://doi.org/10.1039/c8ob02775g
38. Dorrestein, P. C., Yeh, E., Garneau-Tsodikova, S., Kelleher, N. L., & Walsh, C. T. (2005). Dichlorination of a pyrrolyl-S-carrier protein by FADH2- dependent halogenase PltA during pyoluteorin biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America, 102*(39), 13843–13848.
39. Podzelinska, K., Latimer, R., Bhattacharya, A., Vining, L. C., Zechel, D. L., & Jia, Z. (2010). Chloramphenicol biosynthesis: The structure of CmlS, a flavin-dependent halogenase showing a covalent flavin-aspartate Bond. *Journal of Molecular Biology, 397*(1), 316–331.
40. Inan, M., & Meagher, M. M. (2001). Non-repressing carbon sources for alcohol oxidase (*AOX*1). *Journal of Bioscience and Bioengineering, 92*(6), 585–589.
41. El Gamal, A., Agarwal, V., Diethelm, S., Rahman, I., Schorn, M. A., Sneed, J. M., Louie, G. V., Whalen, K. E., Mincer, T. J., Noel, J. P., Paul, V. J., & Moore, B. S. (2016). Biosynthesis of coral settlement cue tetrabromopyrrole in marine bacteria by a uniquely adapted brominase-thioesterase enzyme pair. *Proceedings of the National Academy of Sciences of the United States of America, 113*(14), 3797–3802.
42. Kittilä, T., Kittel, C., Tailhades, J., Butz, D., Schoppet, M., Büttner, A., Goode, R. J. A., Schittenhelm, R. B., Van Pee, K.-H., Süssmuth, R. D., Wohlleben, W., Cryle, M. J., & Stegmann, E. (2017). Halogenation of glycopeptide antibiotics occurs at the amino acid level during non-ribosomal peptide synthesis. *Chemical Science, 8*(9), 5992–6004.
43. Nielsen, M. T., Nielsen, J. B., Anyaogu, D. C., Holm, D. K., Nielsen, K. F., Larsen, T. O., & Mortensen, U. H. (2013). Heterologous reconstitution of the intact geodin gene cluster in *Aspergillus nidulans* through a simple and versatile PCR based approach. *PLoS ONE, 8*(8), e72871. https://doi.org/10.1371/journal.pone.0072871
44. Chankhamjon, P., Boettger-Schmidt, D., Scherlach, K., Urbansky, B., Lackner, G., Kalb, D., Dahse, H. M., Hoffmeister, D., & Hertweck, C. (2014). Biosynthesis of the halogenated mycotoxin aspirochlorine in koji mold involves a cryptic amino acid conversion. *Angewandte Chemie - International Edition, 53*(49), 13409–13413.
45. Xu, X., Liu, L., Zhang, F., Wang, W., Li, J., Guo, L., Che, Y., & Liu, G. (2014). Identification of the first diphenyl ether gene cluster for pestheic acid biosynthesis in plant endophyte *Pestalotiopsis fici*. *ChemBioChem, 15*(2), 284–292.
46. Ferrara, M., Perrone, G., Gambacorta, L., Epifani, F., Solfrizzo, M., & Gallo, A. (2016). Identification of a halogenase involved in the biosynthesis of ochratoxin A in *Aspergillus carbonarius*. *Applied and Environmental Microbiology, 82*(18), 5631–5641.

47. Wick, J., Heine, D., Lackner, G., Misiek, M., Tauber, J., Jagusch, H., Hertweck, C., & Hoffmeister, D. (2016). A fivefold parallelized biosynthetic process secures chlorination of *Armillaria mellea* (honey mushroom) toxins. *Applied and Environmental Microbiology, 82*(4), 1196–1204.
48. Asaduzzaman, M. (2007). Standardization of yeast growth curves from several curves with different initial sizes. Master thesis, Chalmers University of Technology, Gothenburg, Sweden.
49. Karbalaei, M., Rezaee, S. A., & Farsiani, H. (2020). *Pichia pastoris*: A highly successful expression system for optimal synthesis of heterologous proteins. *Journal of Cellular Physiology, 235*(9), 5867–5881.
50. Cos, O., Ramón, R., Montesinos, J. L., & Valero, F. (2006). Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast *Pichia pastoris* under different promoters: A review. *Microbial Cell Factories, 5*, 1–20.
51. Carr, S., Aebersold, R., Baldwin, M., Burlingame, A., Clauser, K., & Nesvizhskii, A. (2004). The need for guidelines in publication of peptide and protein identification data: Working group on publication guidelines for peptide and protein identification data. *Molecular and Cellular Proteomics, 3*(6), 531–533.
52. Takahashi, S., Furukawara, M., Omae, K., Tadokoro, N., Saito, Y., Abe, K., & Kera, Y. (2014). A highly stable D-amino acid oxidase of the thermophilic bacterium *Rubrobacter xylanophilus*. *Applied and Environmental Microbiology, 80*(23), 7219–7229.
53. Rees, G. S., Gee, C. K., Ward, H. L., Ball, C., Tarrant, G. M., Poole, S., & Bristow, A. F. (1999). Rat tumour necrosis factor-alpha: Expression in recombinant *Pichia pastoris*, purification, characterization and development of a novel ELISA. *European Cytokine Network, 10*(3), 383–392.
54. Zhu, X., De Laurentis, W., Leang, K., Herrmann, J., Ihlefeld, K., van Pée, K. H., & Naismith, J. H. (2009). Structural insights into regioselectivity in the enzymatic chlorination of tryptophan. *Journal of Molecular Biology, 391*(1), 74–85.
55. Hiserodt, R. D., Swijter, D. F. H., & Mussinan, C. J. (2000). Identification of atranorin and related potential allergens in oakmoss absolute by high-performance liquid chromatography-tandem mass spectrometry using negative ion atmospheric pressure chemical ionization. *Journal of Chromatography. A, 888*(1–2), 103–111.
56. Bitto, E., Huang, Y., Bingman, C. A., Singh, S., Thorson, J. S., & Phillips, G. N., Jr. (2007). The structure of flavin-dependent tryptophan 7-halogenase RebH. *Proteins, 70*(1), 289–293.
57. Karabencheva-Christova, T. G., Torras, J., Mulholland, A. J., Lodola, A., & Christov, C. Z. (2017). Mechanistic insights into the reaction of chlorination of tryptophan catalyzed by tryptophan 7-halogenase. *Scientific Reports, 7*, 17395.
58. Lee, J., Kim, J., Kim, H., Kim, E.-J., Jeong, H.-J., Choi, K.-Y., & Kim, B.-G. (2020). Characterization of a tryptophan 6-halogenase from Streptomyces albus and its regioselectivity determinants. *ChemBioChem, 21*(10), 1446–1452.
59. Vos, C., Mckinney, P., Pearson, C., Heiny, E., Gunawardena, G., & Holt, E. (2018). The optimal extraction and stability of atranorin from lichens, in relation to solvent and pH. *The Lichenologist, 50*(4), 499–512.
60. Juturu, V., & Wu, J. C. (2018). Heterologous protein expression in *Pichia pastoris*: Latest research progress and applications. *ChemBioChem, 19*(1), 7–21.
61. Ahmad, M., Hirz, M., Pichler, H., & Schwab, H. (2014). Protein expression in *Pichia pastoris*: Recent achievements and perspectives for heterologous protein production. *Applied Microbiology and Biotechnology, 98*(12), 5301–5317.
62. Daly, R., & Hearn, M. T. W. (2005). Expression of heterologous proteins in *Pichia pastoris*: A useful experimental tool in protein engineenring and production. *Journal of Molecular Recognition, 18*(2), 119–138.
63. Xu, Y., Liu, K., Han, Y., Xing, Y., Zhang, Y., Yang, Q., & Zhou, M. (2021). Codon usage bias regulates gene expression and protein conformation in yeast expression system *P. pastoris*. *Microbial Cell Factories, 20*(1), 1–15.
64. Cregg, J. M., & Higgins, D. R. (1995). Production of foreign proteins in the yeast *Pichia pastoris*. *Canadian Journal of Botany, 73*(S1), 891–897.
65. Matthews, C. B., Kuo, A., Love, K. R., & Love, J. C. (2018). Development of a general defined medium for *Pichia pastoris*. *Biotechnology and Bioengineering, 115*(1), 103–113.