

SCIENTIFIC REPORTS



OPEN

Enzyme discovery beyond homology: a unique hydroxynitrile lyase in the Bet v1 superfamily

Received: 21 November 2016

Accepted: 27 March 2017

Published: 03 May 2017

Elisa Lanfranchi^{1,†}, Tea Pavkov-Keller^{1,2}, Eva-Maria Koehler¹, Matthias Diepold², Kerstin Steiner¹, Barbara Darnhofer^{1,3,4}, Jürgen Hartler^{4,5}, Tom Van Den Bergh⁶, Henk-Jan Joosten⁶, Mandana Gruber-Khadjawi^{1,‡}, Gerhard G. Thallinger^{1,4,5}, Ruth Birner-Gruenberger^{1,3,4}, Karl Gruber^{1,2}, Margit Winkler^{1,5} & Anton Glieder^{1,5}

Homology and similarity based approaches are most widely used for the identification of new enzymes for biocatalysis. However, they are not suitable to find truly novel scaffolds with a desired function and this averts options and diversity. Hydroxynitrile lyases (HNLs) are an example of non-homologous isofunctional enzymes for the synthesis of chiral cyanohydrins. Due to their convergent evolution, finding new representatives is challenging. Here we show the discovery of unique HNL enzymes from the fern *Davallia tyermannii* by coalescence of transcriptomics, proteomics and enzymatic screening. It is the first protein with a Bet v1-like protein fold exhibiting HNL activity, and has a new catalytic center, as shown by protein crystallography. Biochemical properties of *D. tyermannii* HNLs open perspectives for the development of a complementary class of biocatalysts for the stereoselective synthesis of cyanohydrins. This work shows that systematic integration of -omics data facilitates discovery of enzymes with unpredictable sequences and helps to extend our knowledge about enzyme diversity.

Hydroxynitrile lyases (HNLs) catalyze the cleavage of cyanohydrins into the corresponding aldehyde or ketone and hydrogen cyanide^{1–3}. HNLs have been mostly found in plants, but examples exist from bacteria^{4,5} and arthropods⁶ as well. HNLs from the three kingdoms show no homology, and in plants three non-homologous HNL families are currently known. HNLs are involved in the cyanogenic pathway, a widespread defense mechanism against herbivores and pathogens. Plants produce cyanogenic glycosides e.g. prunasin or amygdalin, store these compounds in different tissues, and enzymatically degrade them upon tissue disruption, in order to finally release toxic hydrogen cyanide². Similarly, arthropods produce HCN from mandelonitrile as a chemical for their defense⁶. The role of HNLs in bacteria is not yet explored.

HNLs are well-established tools for selected biocatalytic industrial processes. They can perform the reverse condensation reaction (carbon-carbon bond formation), and synthesize α -cyanohydrins in a stereospecific manner. Cyanohydrins are building blocks for a number of follow-up reactions, and the products find application in agrochemical, pharmaceutical and cosmetic industries^{7–10}. Although (R)- and (S)-selective HNLs are exploited in biocatalysis, new enzymes with outstanding characteristics are constantly sought for. Limitations of currently known HNLs are low activity¹¹, low stability¹², low expression levels⁶, or low flexibility in the choice of the expression host¹³. Approximately 3,000 plant species utilize cyanogenic defense, and HNL activity has been shown in a wide variety of families^{14–17}, but only few amino acid sequences of confirmed HNLs are found in common gene and protein databases. However, knowledge of an enzyme's amino acid sequence is essential for its recombinant expression and engineering. This – in turn – is prerequisite for applications on scale.

¹Austrian Centre of Industrial Biotechnology, Petersgasse 14, 8010 Graz, Austria. ²University of Graz, Institute of Molecular Biosciences, NAWI Graz, Humboldtstrasse 50, 8010 Graz, Austria. ³Medical University of Graz, Institute of Pathology, Stiftingtalstrasse 24, 8010 Graz, Austria. ⁴Omics Center Graz, BioTechMed Graz, Stiftingtalstrasse 24, 8010 Graz, Austria. ⁵Graz University of Technology, Institute of Molecular Biotechnology, NAWI Graz, Petersgasse 14, 8010 Graz, Austria. ⁶Bio-product BV, Nieuwe Marktstraat, 54E 6511 AA, Nijmegen, The Netherlands. [†]Present address: University of Groningen, Department of Biochemistry, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), Nijenborgh 4, 9726 AG Groningen, The Netherlands. [‡]Present address: Fresenius Kabi Austria GmbH, Hafnerstraße 36, A-8055 Graz, Austria. Correspondence and requests for materials should be addressed to M.W. (email: margit.winkler@acib.at)

The determination of truly new sequences is the weak point for this enzyme class and the reason is related to protein evolution. HNLs are an example of non-homologous isofunctional enzymes (NISE), a group of unrelated proteins that catalyze the same chemical reaction as a result of convergent evolution¹⁸. Therefore, the typical homology based discovery approaches are unsuitable to elucidate novel HNL sequences. To date, five different protein folds have been associated to HNL activity and they do not share any conserved motif. Crystal structures of HNLs belonging to cupin¹¹, GMC oxidoreductase¹⁹, α/β -hydrolase^{20–22} and peptidase S10²³ families have been determined. Moreover, by sequence homology, an HNL has been described as a member of the zinc-binding dehydrogenase family²⁴. The first HNL sequence from invasive millipedes was discovered by a five step purification of the protein from kilogram quantities of millipede animals and subsequent Edman degradation for elucidation of the amino acid sequence. Its sequence indicates that it does not belong to any of the above mentioned protein folds⁶. Finally there is a number of characterized HNLs with yet unpublished amino acid sequences and protein folds, for example, *PatHNL* (*Prunus amygdalus turcomanica*)²⁵, *ParsHNL* (*Prunus armeniaca* L.)²⁶, *PeHNL* (*Passiflora edulis*)²⁷ (sequence information for PeHNL is public since December 2016, see Ref 28), and the fern HNL from *Phlebodium aureum* (*PhaHNL*)²⁸. For the latter enzyme, outstanding specific activity was reported, and also indications that it is distinct to known HNLs²⁹. Biochemical characterizations and experiments towards cyanohydrin synthesis were performed with proteins isolated from the natural sources.

A classical HNL identification workflow would comprise of several steps of protein purification starting from large amounts of the natural source to obtain highly pure enzyme. Sequence information is then typically obtained by Edman degradation or tryptic digestion and mass spectrometry. The full sequence can then be identified by PCR using degenerate primers. Altogether, this is a very laborious and time consuming process. Here, we describe an alternative approach for the discovery of a novel and unique HNL from the white rabbit's foot fern *Davallia tyermannii* (alternative names: *Humata tyermannii*; *Humata tyermannii* T. Moore; *Davallia tyermannii* (T. Moore) Backer) by the combination of different *-omics* data and enzymatic screening. To our knowledge, it is the first time where *-omics* techniques have been combined for the determination of an HNL sequence from scratch. In addition to enzyme characterization and synthesis of cyanohydrins, we determined the tertiary structure of *Davallia tyermannii* HNL (*DtHNL*) and proposed its catalytic mechanism. Finally, we investigated the presence of HNLs in different fern families.

Results

From the enzymatic activity to the sequence. Based on the first report of a highly active HNL in the fern *Phlebodium aureum*²⁹, we first screened a selection of ferns and identified cyanogenic activity of *Davallia tyermannii* leaves by olfactory screening of disrupted plant tissue. Cyanogenic phenotype does not always correlate with the presence of a hydroxynitrile lyase (HNL). Therefore, HNL activity was confirmed in the cyanogenic fern *Davallia tyermannii* by the addition of (*RS*)-2-hydroxy-2-phenylacetonitrile (racemic mandelonitrile) to a protein preparation as described in the Methods section and subsequent detection of the release of hydrogen cyanide³⁰ (Supplementary Result 1). Mandelonitrile was chosen, as this is the natural cyanohydrin identified in the genus *Davallia* and different other fern genera^{28,31,32}.

The sequence discovery workflow combined several techniques. We collected the information about all expressed genes, which is well described by a sequenced normalized transcriptome. For this purpose, high quality mRNA from *D. tyermannii* leaves and croziers was isolated and the normalized cDNA library was sequenced.

Transcriptome data is sufficient for enzyme discovery, only when information about the primary sequence, conserved residues, domains or motifs is available. As homology based approaches produced no hits in the transcriptome, retrieval of information on the protein level was necessary. To pinpoint HNL fragments, we subjected proteins from active tissues to anion exchange chromatography and subsequently to an BN PAGE coupled HCN detection assay³³. Several elution fractions showed enzymatic activity, which correlated to blue spots at approximately 20 kDa bands (Fig. 1a). These were excised and the protein mixture was subjected to tryptic digestion and LC-MS/MS analysis. LC-MS/MS peptide data were matched by searching the translated nucleotide transcriptome database from *D. tyermannii*. Thirty-six identified protein sequences were further ranked by an exclusion process based on predicted protein size, signal peptide and similarity with known protein sequences (Supplementary Dataset 1). The sequence with the highest probability to represent the protein with HNL activity was isotig02643. Nevertheless, the top six candidates were recombinantly expressed in *E. coli* and tested for HNL activity (Fig. 1b; Supplementary Table 3).

The protein encoded by the open reading frame (ORF) of isotig02643 showed HNL activity, when racemic mandelonitrile was added to cell free lysate from *E. coli* (Fig. 1b). The other five candidate proteins did not show HNL activity and were not investigated further. The peptides retrieved from mass spectrometry cover 72% of the translated ORF of isotig02643 (Fig. 1c). Translated nucleotide transcriptome database revealed three additional highly similar sequences to the confirmed HNL: translated ORFs of isotig02641, isotig07602 and contig00751 show at least 93.5% identity with isotig02643 (Supplementary Fig. 5a). The respective three genes were amplified from *D. tyermannii* gDNA by PCR, and their nucleotide sequences were confirmed by Sanger sequencing (Supplementary Table 4). After recombinant expression in *E. coli*, the HNL activity assay confirmed cyanogenic activity of all three isoenzymes (Supplementary Fig. 5b). The four proteins coded by the ORF of isotig02643, 02641, 07602 and contig00751 were named *DtHNL1*, *DtHNL2*, *DtHNL3* and *DtHNL4*, respectively.

The novelty of *DtHNL* was investigated by BLAST³⁴ of known HNLs against the obtained transcriptome. Low identity and sequence coverage in the output confirmed that *DtHNL* is distinct from known HNL classes (Supplementary Results 3).

Characterization of *DtHNL* isoenzymes. Biochemical features of all four isoenzymes were determined after heterologous expression and purification of the His-tagged proteins by affinity chromatography (Supplementary Results 6). The influence of pH and temperature on the catalytic activity was investigated by

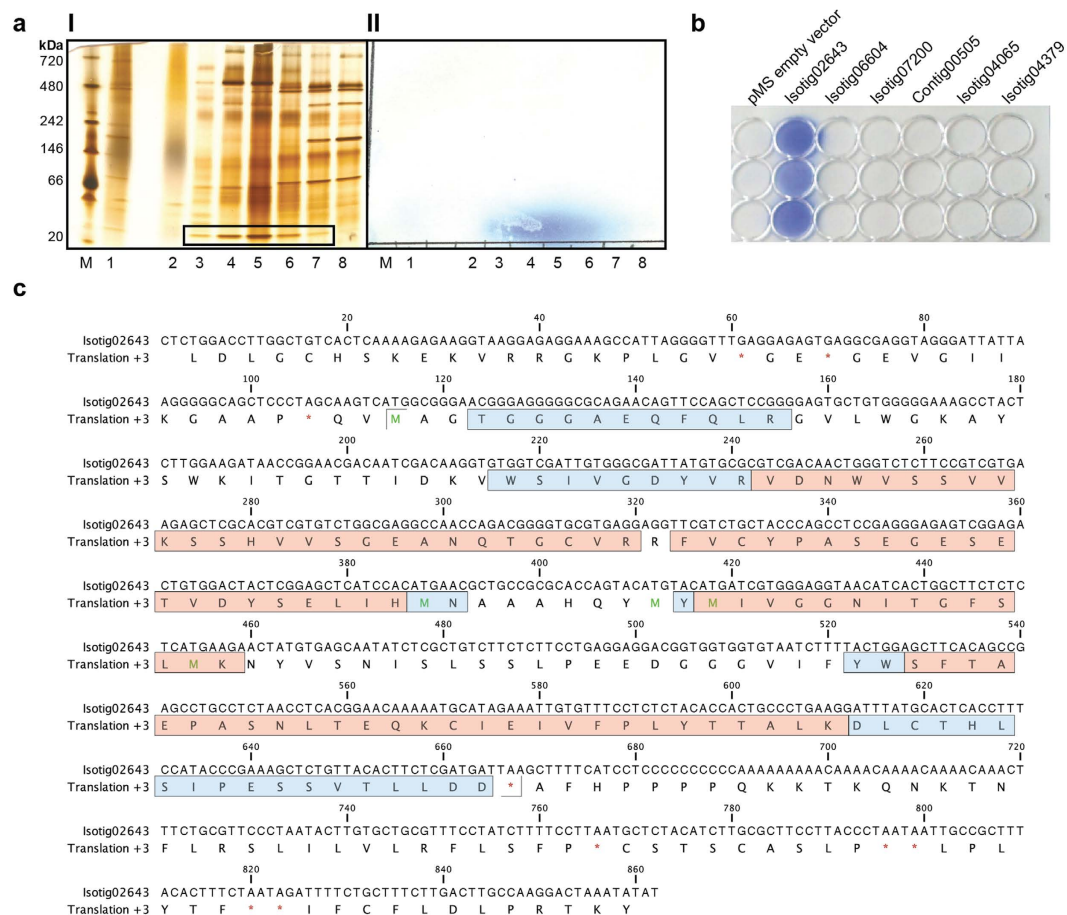


Figure 1. Identification of the *DthNL* sequence. A: Blue Native PAGE followed by HNL activity assay. (a) I: Concentrated fractions of anion exchange purification were applied separately on BN PAGE. NativeMark™ Unstained Protein Standard (Thermo Fisher Scientific); total protein extract from *D. tyermannii* leaves (1); flow through (2); elution fractions (3–8); active bands at 20 kDa are highlighted in the box. (a) II: HNL activity is depicted by the blurred blue spot corresponding to purification fractions 2–7. The total protein extract shows a weak signal (1). Assay conditions: 100 mM citrate buffer pH 4.0; substrate: racemic mandelonitrile. Incubation time 8 min. (b) Screening of putative HNL sequences. Cell free lysate of *E. coli* TOP 10F' strains expressing six putative HNL proteins. Each sample was tested in triplicate: 30 μ L (1), 20 μ L (2), 10 μ L (3) of cell free lysate, respectively. *E. coli* TOP 10F' transformed with pMS vector was used as negative control. Assay conditions: 100 mM citrate buffer pH 4.5; substrate: racemic mandelonitrile 13 mM. The intensive blue spots developed after few seconds of incubation. Proteins with unknown function were named as the relative isotig or contig number found in the transcriptome. (c) Nucleotide and amino acid sequence of isotig 02643 (*DthNL1*). Fragments detected by mass spectrometry are labeled. Indicated results were obtained from the protein bands excised from lanes 4 (red) and 5 (red and blue). The peptides identified by mass spectrometry cover 72% of the open reading frame.

following the cleavage reaction of racemic mandelonitrile. Maximum activity was observed at pH 5.0. The activity linearly decreased in more acidic environment and was zero at pH 2.0. Remarkably, all isoforms were active at pH 2.5, and retained about 80% of their activity at pH 4.0 (Fig. 2a). Enzymatic activity at low pH is an essential asset for HNLs in terms of their application, since cyanohydrins easily degrade at pH \geq 5.0 as indicated by the background reaction line (Fig. 2a). Enzymatic syntheses of cyanohydrins are preferably performed at pH 4.0 or lower and robust biocatalysts in acidic environment are highly desired. Therefore, we investigated *DthNL* stability at pH 2.5 and 4.0 by incubation at 8 °C for 72 hours. Residual activity was more than 50% at both pH 2.5 and pH 4.0. (Fig. 2c,d) after 24 hours. Isoform 1 appeared to be less stable at pH 4.0 than *DthNL*2, 3 and 4 (Fig. 2d). Additionally, enzymatic stability was tested at pH 5.0, and residual activity was more than 65% at 72 hours of incubation (Supplementary Table 5).

Figure 2b shows *DthNL* activity at different temperatures. The optimum was observed at 35 °C for *DthNL*2 and 4 and 40 °C for *DthNL*1 and 3, respectively. Mandelonitrile spontaneously degrades above pH 5.5 and elevated temperature also contributes to degradation, resulting in increased apparent reaction rates, as indicated by the background reaction curve (Fig. 2a,b). pH 5.0 and 25 °C were used as a good compromise between optimal enzymatic activity and repressed substrate decomposition.

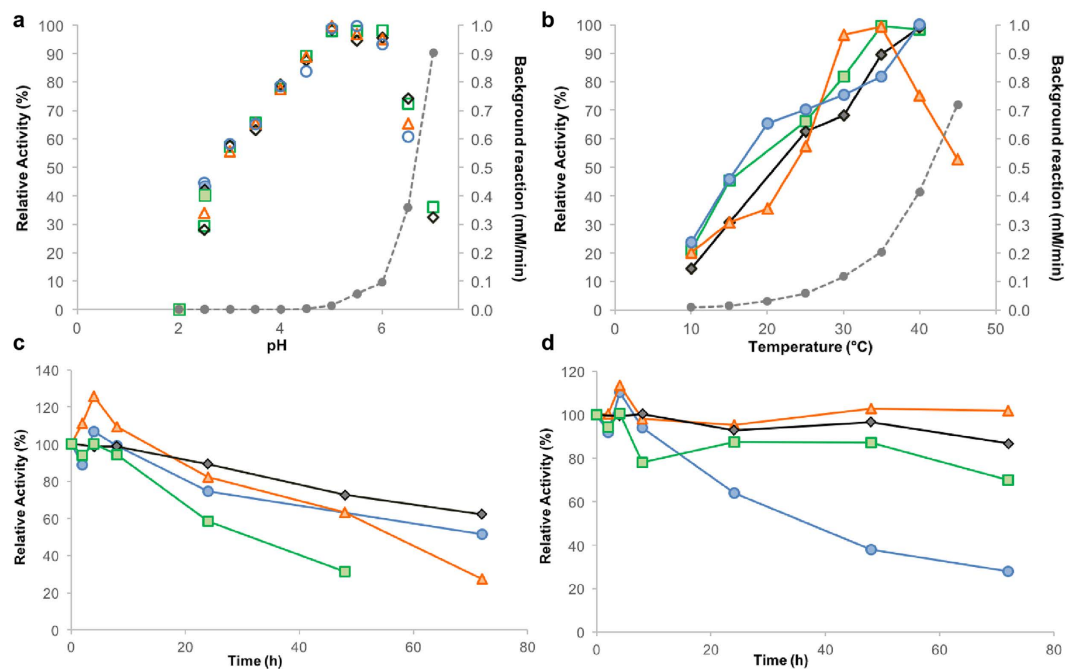


Figure 2. Biochemical properties of *DtHNL* isoenzymes. *DtHNL1* ●; *DtHNL2* ▲; *DtHNL3* ◆; *DtHNL4* ■. Grey dashed lines indicate the spontaneous degradation of racemic mandelonitrile in a negative control reaction without enzyme addition (background reaction). Standard enzymatic assay was performed by monitoring benzaldehyde formation at 280 nm. Values were obtained from the average of two and a maximum of three independent samples, each of which is the average of two or three technical replicates. Standard deviations are within the 20% threshold (or 25% for temperature profile). For clarity, error bars have been omitted. (a) pH profile. Relative activity of *DtHNL* isoenzymes at different pH values from 2.0 to 7.0. The assay was performed in HCl-potassium chloride buffer (filled symbols), or sodium citrate-phosphate buffer (empty symbols). Activity of *DtHNL1* and 2 at pH 7.0 is not depicted due to high standard deviations. (b) Temperature profile. Relative activity of *DtHNL* isoenzymes at different temperatures from 10 to 50 °C. The assay was performed at pH 5.0. Omitted points are due to high standard deviations. Enzyme stability at pH 2.5 (c) and at pH 4.0 (d). Activity after incubation of *DtHNL* isoenzymes at pH 2.5 or 4.0, respectively, and 8 °C. Relative activity is based on the activity before incubation.

Finally, the Michaelis constants K_m and turnover numbers k_{cat} were determined on basis of the Michaelis-Menten model for the cleavage of (*R*)-mandelonitrile. All *DtHNL*s have high affinity for (*R*)-mandelonitrile as indicated by K_m values of 0.30 ± 0.03 mM *DtHNL1*, 0.45 ± 0.05 mM *DtHNL2*, 0.75 ± 0.08 mM *DtHNL3* and 0.63 ± 0.06 mM *DtHNL4*, respectively. These results are consistent with K_m values of reported (*R*)-HNL enzymes². The turnover numbers were 144 ± 2 s⁻¹ for *DtHNL1*, 156 ± 3 s⁻¹ for *DtHNL2*, 356 ± 8 s⁻¹ for *DtHNL3* and 272 ± 5 s⁻¹ for *DtHNL4*, respectively.

Kinetic parameters for *DtHNL1* mediated synthesis of (*R*)-mandelonitrile were determined in a biphasic system (1:2 aqueous:organic phase) at pH 4.0 and 10 °C. Hydrocyanic acid was kept in saturating concentration and the initial rate was measured at different concentrations of benzaldehyde. Kinetic parameters were calculated using the Michaelis-Menten model as described for the reaction in the cleavage direction. The majority of (*R*)-mandelonitrile is in the organic layer and the minor amount in the aqueous buffer was omitted for the quantification. Under these conditions, v_{max} of *DtHNL1* was 377 ± 47 μmol min⁻¹ mg⁻¹, and K_m for benzaldehyde was 14 ± 2 mM. k_{cat} was 70 ± 7 s⁻¹ and the enzymatic efficiency k_{cat}/K_m was 5 ± 1 s⁻¹ mM⁻¹, assuming *DtHNL1* activity in the aqueous phase only.

***DtHNL* structure and reaction mechanism.** We determined the crystal structure of *DtHNL* using selenomethionine single-wavelength-anomalous-dispersion (SeMet-SAD) to a resolution of 1.85 Å. The enzyme is a dimer and exhibits a Bet v1-like fold. The Bet v1 superfamily is composed of sequences related to the major Birch (*Betula verrucosa*) pollen allergen Bet v1. The fold is composed of an anti-parallel β-sheet, which is wrapped around a long C-terminal α-helix (Fig. 3a). The ligand binding cavity is situated between the β-sheet and the helix. In proteins of the Bet v1 superfamily, this cavity plays important roles in the binding and metabolism of large, hydrophobic compounds such as lipids, hormones, and antibiotics³⁵. We also determined the structures of complexes of *DtHNL* with 4-hydroxy benzaldehyde, benzoic acid and (*R*)-mandelonitrile/benzaldehyde using soaking techniques (Supplementary Table 6). Clear electron density was observed for those ligands bound in the canonical binding cleft of the Bet v1 fold (Supplementary Fig. 7a).

In all three structures, the aromatic moiety of the different ligands is bound in a hydrophobic pocket formed by the side chains of Val44, Val48, Trp47, Val51, Val52, Phe71, Cys73, Ile108, Phe111, Trp138, Leu160 and Ala164.

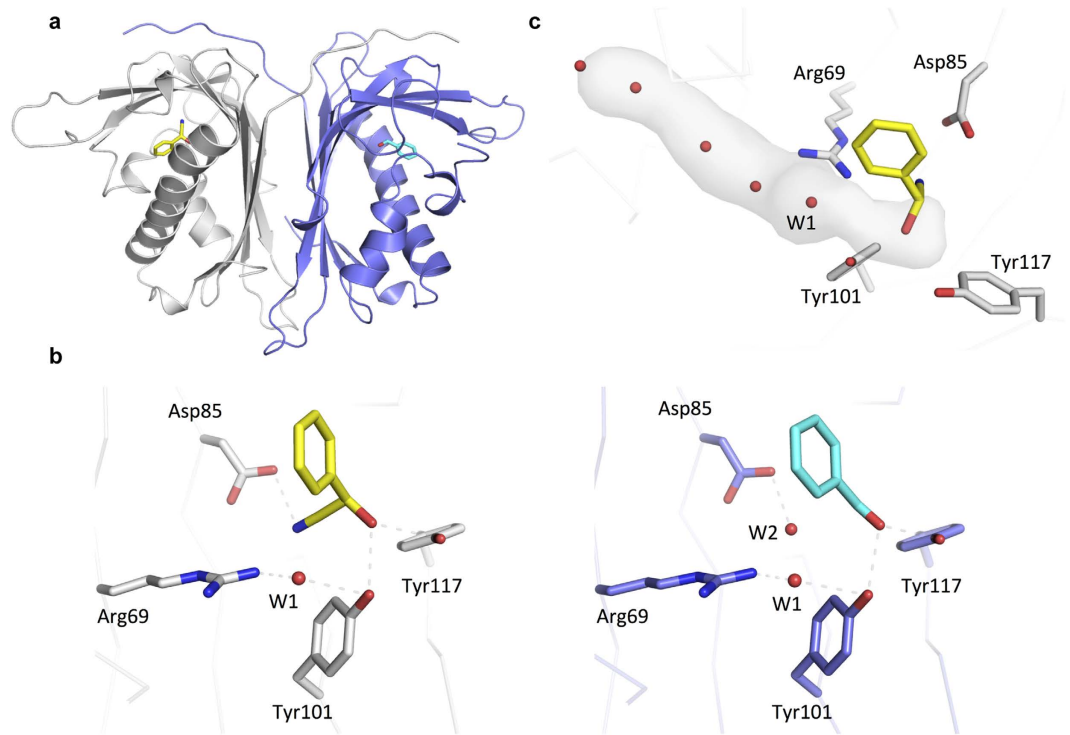


Figure 3. Crystal structure *DtHNL1*. (a) Overall structure of the *DtHNL1* dimer. Individual protomers are shown in a grey or blue ribbon representation. Bound ligands are shown in a stick representation (mandelonitrile in yellow and benzaldehyde in cyan). (b) Close-up view of the active site. Residues potentially important for the enzymatic activity are shown in a sticks representation and water molecules are depicted as red spheres. The bridging water is labeled W1. Dashed lines indicate plausible hydrogen bonding interactions. Bound benzaldehyde (cyan, right) and (*R*)-mandelonitrile (yellow, left) as observed in a crystal soaked with (*R*)-mandelonitrile are drawn in a sticks representation. (c) Access to the active site. Water channel extending from the active site to the surface of the protein as calculated using the program CAVER⁴⁹.

Especially the valine residues at positions 44, 48, 51 and 52 together with Phe71 and Leu160 seem to be important to shape the cavity. A part of the phenyl ring is also pointing towards the solvent, therefore, larger substrate may be bound in a way that substituents are located in the active site access tunnel.

The polar parts of the ligands (the OH-, carbonyl- or carboxyl-group) are hydrogen bonded to Tyr101 and Tyr117 (Fig. 3b and Supplementary Fig. 7b). In the complex with (*R*)-mandelonitrile, the cyano group interacts with the guanidinium group of Arg69 (distance 3.5 Å) and the carboxylic acid group of Asp85 (2.9 Å). The latter interaction requires Asp85 to be protonated in this complex structure. In the other complexes, a water molecule occupies this position. The active site cavity is clearly asymmetric which provides a reasonable, qualitative explanation for the stereospecificity of *DtHNL*.

The observed polar interactions between (*R*)-mandelonitrile and *DtHNL* suggest that the OH-group of the substrate is deprotonated by Tyr101, which is facilitated by the additional hydrogen bond from Tyr117. A tyrosine residue alone, however, is not a typical base, especially at lower pH-values (pK_a of tyrosine ~ 10) and there is no other amino acid residue in vicinity (such as a histidine), which could activate the phenol. A water molecule bridges the OH-group of Tyr101 and the guanidinium group of Arg69 (Fig. 3b). There are several possibilities, how two protons can be distributed between these groups: a) Tyr-OH, OH⁻, Arg⁺, b) Tyr-OH, H₂O, Arg⁰ and c) Tyr-O⁻, H₂O, Arg⁺. Based on the x-ray crystal structures, it is not possible to decide which of the three configurations is the correct one, although a positively charged arginine residue appears to be more plausible (Fig. 4). After cyanohydrin cleavage only one configuration (Tyr-OH, H₂O, Arg⁺) is consistent with the structural data. The negative charge emerging at the cyano group upon C-C bond cleavage is very likely stabilized by the positive charge of Arg69 and a hydrogen bond from the protonated Asp85.

The bridging water is the first of five water molecules, which are nicely aligned in a channel that runs from the active site to the surface of the protein (Fig. 3c). This channel is approximately orthogonal to the main entrance tunnel to the active site and could serve as an access/exit pathway for HCN or as a proton relay to the bulk solvent.

Based on this mechanistic proposal, a number of amino acids were replaced and the enzymatic activity of these variants was determined. Exchange of Tyr101 by phenylalanine led to a complete loss of activity. The replacement of Asp85 and Ser87 by alanine or of Tyr117 and Tyr161 to phenylalanine decreased the activity by at least 90%. Other amino acid exchanges (especially of Arg69) produced insoluble protein and prevented activity measurements (Supplementary Table 7).

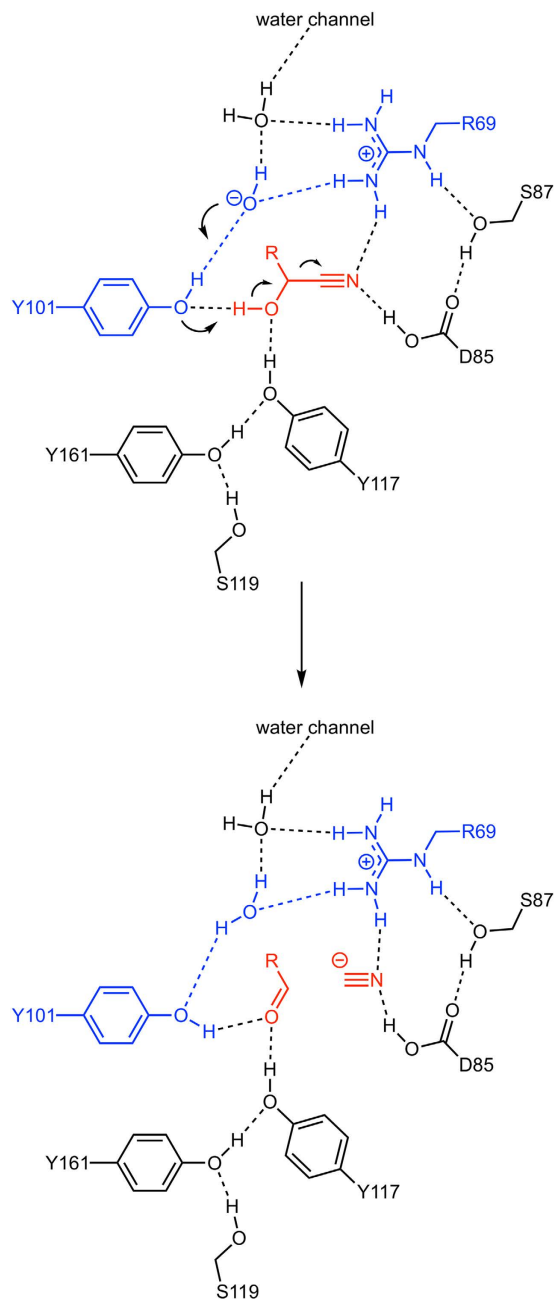


Figure 4. Proposed catalytic mechanism for the *DthHNL1*. Cyanohydrin cleavage of (*R*)-mandelonitrile based on complex-crystal structures and mutagenesis experiments.

Different HNL classes from ferns. Ferns are a group of sparsely studied vascular plants, which include several classes and consequently thousands of species. Ferns can be considered as an outgroup of the plant kingdom as they maintained the ancestral condition and show different characteristics compared to seed plants. For example, typically they have much higher chromosome numbers and larger genomes³⁶. The sequence of *DthHNL* is the first HNL sequence identified from ferns, although several fern species adopt cyanogenesis as chemical defense strategy. We aimed to investigate whether cyanogenic ferns express homologous HNLs or developed them independently, as often happens in seed plants.

The bracken fern *Pteridium aquilinum* (L.) Kuhn is widespread throughout the northern hemisphere and Africa and belongs to the *Dennstaedtiaceae* family³⁶. It is another well-known example of cyanogenic fern in addition to *D. tyermannii* and *P. aureum*, and was easily accessible for our studies from an Austrian forest. First, high quality mRNA was isolated from enzymatically active leaves and croziers. The transcriptome was obtained by sequencing the normalized cDNA library (the quality of the transcriptome and assembly results are reported in Supplementary Results 2). The transcriptome did not show proteins with significant similarity to any known HNL sequences upon TBLASTN search (Supplementary Results 3). When *DthHNL1* was used as query for TBLASTN search in the *P. aquilinum* transcriptome database, we obtained 17 hits (Supplementary Table 8). Two sequences

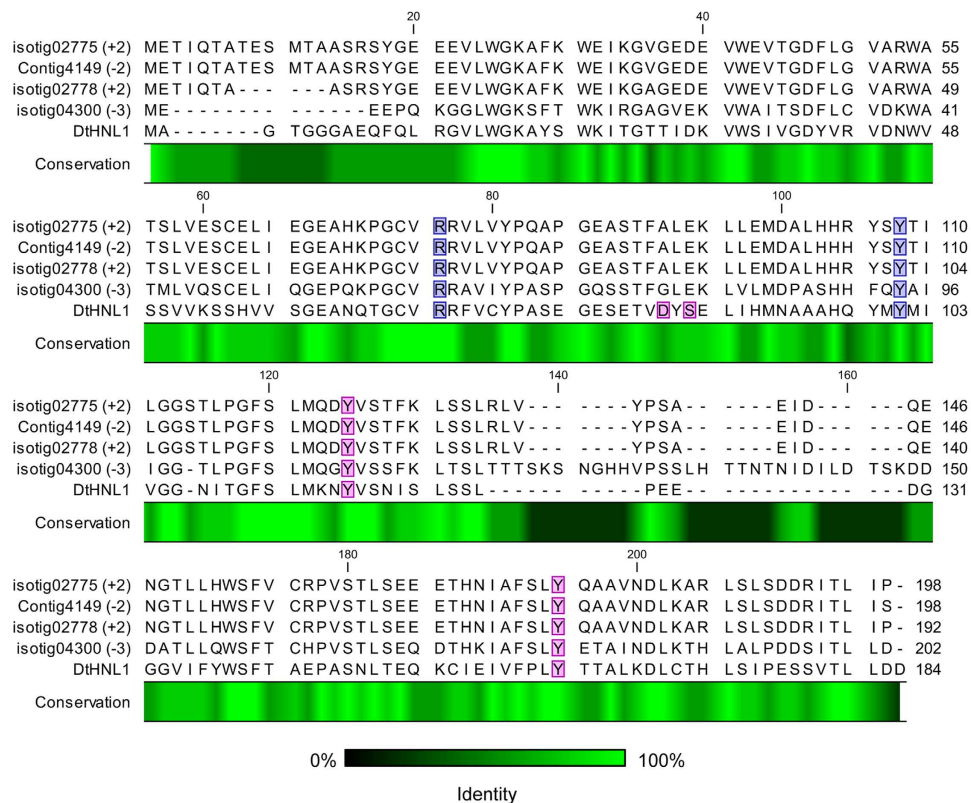


Figure 5. Multiple sequence alignment. Isotig02775 and isotig02778 are proteins with similarity to *DtHNL1* from the transcriptome of *P. aquilinum* leaves. Contig4149 was found in the transcriptome obtained from a *P. aquilinum* gametophyte³⁶. Isotig04300 is the sequence most similar to isotig02775 and isotig02778, which was found in *D. tyermannii*. Isotig translation frame is indicated in brackets. Conservation % is indicated by a color code. Residues involved in substrate binding and catalysis are highlighted. Alignment was built with CLC Main Workbench 7.6.2 (QIAGEN). Proteins with unknown function were named as the relative isotig or contig number found in the transcriptome.

with identity to *DtHNL1* above 35%: isotig02775 and isotig02778 were chosen. They belong to the same isogroup and their ORFs encode for two 96% identical proteins (Supplementary Table 8). The identity to *DtHNL1* is 41% and the catalytic residues Arg69 and Tyr101 as well as Tyr117 and Tyr161 in the active site are conserved (Fig. 5). *DtHNL1* was subjected to a second TBLASTN search in a published translated transcriptome from *P. aquilinum* obtained during its gametophyte life stage³⁶. The protein coded by the ORF of contig4149 is 99% identical to isotig02775, corroborating our hit (Fig. 5). However, when we subjected the three amino acid sequences obtained by the translation of isotig02775, isotig02778 and contig4149 to a TBLASTN search in the *D. tyermannii* transcriptome database, isotig04300 was obtained as the best hit instead of expected *DtHNL1* or its isoenzymes.

Isotig04300 codes for a protein which belongs to the Bet v1 protein superfamily and it is 39% identical to *DtHNL1* (55% similarity). However, it was not identified by mass spectrometry in fractions of *D. tyermannii* with HNL activity (Supplementary Dataset 1). This led to the hypothesis that similar proteins from *P. aquilinum* (Isotig02775, isotig02778 and contig4149) do not catalyze cyanohydrin cleavage and formation. To test this assumption, isotig02775 was expressed as soluble protein in *E. coli* and purified by affinity chromatography. The cyanogenic activity of isotig02775 was zero for racemic mandelonitrile, confirming our hypothesis. The protein was also inactive after mutation of Ala92 for Ser (the position corresponding to Asp85 in *DtHNL1*). A double mutation of isotig02775 Ala92Asp and Glu94Ser, simulating *DtHNL1* at this site, lead to quantitative inclusion body formation, preventing the determination of the mutant's activity. The substitution of Ala92 into aspartate seems to be a destabilizing mutation. This might be due to its strong polar character and bulkier structure of the side chain of aspartate.

***DtHNL* catalyzes the synthesis of cyanohydrins.** To explore the potential of *DtHNL*, we studied activity and stereoselectivity for the synthesis of cyanohydrins. In HNL-catalyzed reactions, high yields of enantiomerically pure product are compromised by spontaneous non-enzymatic formation of racemic cyanohydrin and racemization of the product due to equilibration of the reaction. Therefore, it is particularly important to suppress the chemical condensation and racemization of cyanohydrins and opt for conditions which allow the enzyme to outperform the non-enzymatic transformations. The decrease of the water content by the use of a biphasic systems and low pH are two solutions extensively reported in literature³⁷. The choice of pH and the organic solvent/

Entry	Substrate		0.5 hours		2 hours		6 hours		24 hours	
	→		conv	ee	conv	ee	conv	ee	conv	ee
	Product		%	%	%	%	%	%	%	%
1	benzaldehyde → (2R)-2-hydroxy-2-phenylacetone nitrile	<i>DtHNL1</i>	95.1	99.0	97.4	94.5	97.9	85.9	99.5	61.4
		Control	-4.9	-	11.7	0.8	18.5	0.3	39.7	0.5
2	2-chlorobenzaldehyde → (2R)-2-(2-chlorophenyl)-2-hydroxyacetone nitrile	<i>DtHNL1</i>	38.6	89.4	75.6	92.1	96.6	88.4	99.6	71.3
		Control	11.3	-14.0	19.6	-5.7	28.7	-2.3	59.8	-1.5
3	3-phenylpropanal → (2R)-2-hydroxy-4-phenylbutanenitrile	<i>DtHNL1</i>	50.9	45.8	84.8	48.3	94.5	48.1	93.2	48.0
		Control	18.0	0.4	36.3	0.2	53.0	0.0	80.2	0.2
4	3-phenylprop-2-enal → (2R)-2-hydroxy-4-phenylbut-3-enenitrile	<i>DtHNL1</i>	5.3	82.3	46.6	95.2	88.2	94.5	98.2	92.6
		Control	-17.0	-	-11.6	-	-12.6	-	-10.5	-
5	furan-2-carbaldehyde → (2S)-furan-2-yl-hydroxyacetone nitrile	<i>DtHNL1</i>	95.3	99.3	98.0	98.9	98.0	97.6	98.1	92.9
		Control	10.5	0.0	20.5	0.4	35.9	0.1	71.1	-0.1
6	1-phenylethanone → 2-hydroxy-2-phenylpropanenitrile	<i>DtHNL1</i>	9.8	n.d.	17.3	n.d.	23.5	n.d.	22.3	n.d.
		Control	6.6	-	6.6	-	6.9	-	7.0	-

Table 1. Synthesis of cyanohydrins with *DtHNL1*. The synthesis of cyanohydrins was performed with 3 mg of *DtHNL1*. 0.5 M aldehyde or 0.3 M of ketone were mixed with 2 M HCN in TBME (final volume 1 mL), 2% v/v triisopropylbenzene was added as internal standard. To monitor the non-enzymatic formation of cyanohydrins, independent control reactions were set up at the same conditions, but omitting the enzyme (Control). Reaction conditions: pH 4.0, 10 °C and 1,000 rpm. Samples were analyzed by GC after acetylation of the product. Conversion (conv) is based on the substrate consumption. Entry 5: change of product configuration as a consequence of the Cahn–Ingold–Prelog rule. Dashes indicate that no product peak was detected. n.d., not determined.

water phase ratio partly depend on the activity and stability of the biocatalyst. Here, we chose pH 4.0 and a 2:1 ratio v/v organic phase/aqueous phase. Water is necessary for the activity, as described above. *DtHNL1* was sufficiently stable at pH 4.0 for 24 hours (Fig. 2d) and showed ca. 80% of its maximum activity at this pH (Fig. 2a). We used 3 mg of *DtHNL1* (equal to 0.026 mol% for aldehyde substrates or 0.043 mol% for the ketone substrate) and monitored the cyanohydrin synthesis as summarized in Table 1.

DtHNL1 accepted different aldehyde types, and at least one ketone (1-phenylethanone). It showed high selectivity for the (*R*)-products and high reaction rates. The maximum conversion and $\geq 99\%$ enantiomeric excess were obtained in less than 2 hours for the synthesis of (*R*)-mandelonitrile and (*2S*)-furan-2-yl-hydroxyacetone nitrile (Table 1, entries 1 and 5). As expected for a natural non-engineered enzyme, benzaldehyde was the best substrate, and yield and enantiomeric purity of the product were not affected by decreasing the catalyst loading. We obtained 97.6% conversion and $\geq 99\%$ enantiomeric excess (*ee*) in 24 hours, using more than hundred times less catalyst (0.2×10^{-3} mol% of *DtHNL1*). Depending on the substrate structure, the enzyme competes more or less with the chemical condensation reaction for substrate availability, as indicated by the control values in Table 1. Low values for 1-phenylethanone and 3-phenylprop-2-enal reflect little or no background reaction (Table 1, entries 4 and 6).

Low enantiomeric excess was obtained for (*2R*)-2-hydroxy-4-phenylbutanenitrile (Table 1, entry 3) as a consequence of slow enzymatic transformation combined with fast unselective background. Consistent results were observed with less biocatalyst loading: 90.4% conversion, 43.5% *ee* and 97.0% conversion, 50.4% *ee* were retrieved with 1.7×10^{-3} mol% and 8.7×10^{-3} mol% of *DtHNL1* respectively, after 24 hours. These results indicate that *DtHNL1* wild-type displays intrinsically low stereoselectivity for 3-phenylpropanal cyanohydrin formation, due to flexibility of the propanal moiety in the active site and possibly two binding modes. The decrease of *ee* over time for (*R*)-mandelonitrile, (*2R*)-2-(2-chlorophenyl)-2-hydroxyacetone nitrile, (*2R*)-2-hydroxy-4-phenylbut-3-enenitrile and (*2S*)-furan-2-yl-hydroxyacetone nitrile (Table 1, entries 1,2,4,5) is attributed to the cyanohydrin racemization in buffer³⁸. Finally, *DtHNL1* can convert aromatic ketones such as 1-phenylethanone, however, in low analytical yield (<24%), which reveals ketones as more difficult substrates for this enzyme (Table 1, entry 6). Lower activities for ketones as substrates in comparison to aldehydes were also observed for α/β -hydrolase fold HNLs³⁹ and GMC oxidoreductase fold HNLs⁴⁰. The cupin fold HNL from *Granulicella tundricola*, e.g. does not catalyze the addition of cyanide to ketones¹¹.

Discussion

Several bioinformatic tools have been developed for enzyme discovery, especially for wide screening of metagenomic libraries and sequenced genomes of bacteria and fungi. These tools are mainly based on similarities between sequences or common features such as characteristic protein motifs. Therefore, only homologous genes can be identified, even if they are distantly related.

Species varieties evolved different solutions to address the same issue. For example, alkane hydroxylations are catalyzed by completely dissimilar enzymes such as methane monooxygenase⁴¹, cytochrome P450s (CYP153A6⁴² or CYP52 from *Candida*⁴³, alkB from *Pseudomonas*⁴⁴), or fungal peroxxygenases⁴⁵, indicating that convergent evolution is well represented in nature for important biochemical reactions. Plants adopt several defense tactics against herbivory, chemical or mechanical attack. This is also valid at the molecular level, which is exemplified by

hydroxynitrile lyases. Nature developed numerous strategies to catalyze the same chemical reaction – cyanogenesis – and the only common feature between the different classes is acid/base catalysis⁴⁶.

We discovered a protein with hydroxynitrile forming activity unique within the Bet v1 superfamily. The most similar characterized protein is the lachrymatory factor synthase from onion where the sequence identity is less than 25%. Other family members are polyketide cyclases² and abscisic acid receptor; most other sequences are annotated as unknown proteins. Additional examples of sequences with described function in the superfamily are norcoclaurine synthases⁴⁷ or proteins with *in vitro* RNAase activity³⁵. However, they are significantly distant from *DtHNL* and they were classified differently in Pfam (pf00407, *DtHNL* pf10604).

Based on structural analyses and mutation studies, we identified six residues responsible for the substrate binding and catalysis. Specifically, Arg69 and Tyr101 are directly involved in the catalysis together with a water molecule. Tyr117, Asp85, Ser87 and Tyr161 are also relevant for enzymatic activity. The catalytic residues, Tyr117 and Tyr161 are conserved in isotig02775, however, they are not sufficient for the activity. Aspartate in position 85 and serine in position 87 seem to be most critical (Isotig02775 numbering: 92 and 94) to confer HNL activity. These residues are occupied by alanine and glutamic acid in isotig02275 and the respective protein was inactive. Engineering the two critical residues (Ala92Ser and Ala92Asp-Glu94Ser) resulted in inactive mutants. Interestingly, *DtHNL* is the only sequence in the entire Bet v1 superfamily which exhibits the six residues simultaneously. A glutamic acid in position 87 (*DtHNL* numbering) is strongly conserved in a protein subset created with 3DM⁴⁸, where Arg69, Tyr101 and Tyr117 were fixed (Supplementary Fig. 10g), and this residue is likely not compatible with HNL activity in a Bet v1 fold protein. It appears to be unlikely to identify another protein in the Bet v1 protein superfamily with HNL activity, based on today's knowledge (Supplementary Results 10).

P. aquilinum expresses an HNL (*PtaHNL*) with similar protein size, however, this enzyme appears to belong to yet another protein family, as confirmed by the list of putative HNL sequences obtained by our approach (Supplementary Dataset 2).

From the biocatalytic point of view, we discovered promising enzymes for cyanohydrin synthesis. Recombinant *DtHNL* isoenzymes display excellent specific activities and are tolerant to low pH conditions, the optimal environment for cyanohydrins. *DtHNL1* efficiently converts different aldehydes into the respective cyanohydrins, allowing for short reaction times. Excellent *ee* can be achieved by keeping reaction times short and adjusting the biocatalyst amount. Enantiomerically pure products can be obtained with good yields already with the wild type enzyme for entries 1, 2, 4 and 5 (Table 1), whereas engineering of the protein would be necessary for improving the stereoselectivity of *DtHNL1* for (*R*)-3-phenylpropionaldehyde cyanohydrin production.

In summary, *DtHNL*'s novelty and properties open perspectives for the development of a new class of biocatalysts. With our work, we extend the knowledge about HCN release strategies in nature and associate a new function to the Bet v1 superfamily.

The discovery approach reported herein can speed up the process to identify additional enzymes with hydroxynitrile lyase function significantly. Tedious purification steps are avoided, and there is no need for degenerate primers for the amplification of the genes from gDNA or cDNA, which can be a problematic issue for organisms with big genomes. Finally, the concept can be extended to any enzyme of interest, as long as a suitable enzymatic assay is available. Thus, it gives the opportunity to identify novel sequences for a desired function in a reasonable time and the restriction to identify related proteins is no longer a limitation.

Methods

General. The fern *Davallia tyermannii* was purchased in a local shop. *Pteridium aquilinum* leaves and croziers were harvested from a local forest (Styria, Austria). Racemic mandelonitrile was purchased from abcr GmbH & Co. KG. (*R*)-Mandelonitrile was purchased from Sigma-Aldrich or kindly donated by DSM Fine Chemicals Austria. All other chemicals were purchased from Sigma-Aldrich or Carl Roth GmbH, if not stated otherwise. Material for molecular biology and protein analysis was obtained from Thermo Fisher Scientific or Promega, if not specifically mentioned. Gibson Assembly[®] enzymes were purchased from New England Biolabs and BioZym. ÄKTA purifier (GE Healthcare) was employed for protein purification. Protein purification columns were purchased from GE Healthcare. For protein electrophoresis, an XCell SureLock[®] Mini-Cell equipped with a PowerEase[®] 500 Programmable Power Supply (Thermo Fisher Scientific) was used. Spectrophotometric measurements were performed with a Synergy Mx plate reader (BioTek) or Cary Series Agilent Technologies spectrophotometer. The nucleotide sequence of proteins reported herein have been submitted to GenBank (Supplementary Table 12).

Transcriptome generation and sequencing. The total RNA was isolated from *Davallia tyermannii* and *Pteridium aquilinum* following the protocols provided by the Spectrum[™] Plant Total RNA Kit (Sigma Aldrich) and RNAqueous[®] Kit (Ambion[®], Thermo Fisher Scientific), respectively. Quality assessment to ensure RNA integrity was performed with an Agilent 2100 Bioanalyzer (Agilent Technologies) and agarose gel electrophoresis (1% agarose gel, running conditions: 80 V, 40 min). Normalized transcriptome sequencing was obtained by the commercial service from Microsynth AG as follows: library generation for the 454 FLX sequencing was carried out according to standard protocols (Roche/454 life sciences, Branford, CT 06405, USA). The concatenated inserts were sheared randomly by nebulization to fragments ranging in size from 400 bp to 900 bp. These fragments were end polished and the 454 A and B adaptors that are required for the emulsion PCR and sequencing were ligated to the ends of the fragments. The resulting fragment libraries were sequenced on both halves of a picotiterplate on the GS FLX using the Roche/454 Titanium chemistry. Sequence data can be accessed via the EMBL-EBI European Nucleotide Archive under the study accession number PRJEB10896 (*D. tyermannii*) and PRJEB10897 (*P. aquilinum*).

High-quality reads were selected using Newbler sequence filtering at default settings. The quality controlled reads were assembled into individual isotigs using the Roche/454 Newbler software (454 Life Sciences Corporation, version 2.6.0) with default settings (minimum read length 20, duplicate reads excluded, expected depth 0, seed step 12, seed length 16, seed count 1, minimum overlap length 40 bp, minimum overlap identity 90%, alignment identity score 2, alignment difference score -3).

Protein isolation from fern leaves and purification. Disruption of *D. tyermannii* leaves using the P-PER™ Plant Protein Extraction Reagent was chosen for routine protein isolation according to the manufacturer's protocol. PD-10 desalting columns (GE Healthcare) were used for buffer exchange (50 mM sodium phosphate buffer, pH 5.7). The partial purification of the HNLs was carried out by using anion exchange chromatography (HiTrap QFF 1 mL column, from HiTrap IEX Selection Kit). The column was previously equilibrated with 20 mM sodium phosphate buffer, pH 5.7. The elution was performed with the following parameters: gradient from 0 to 1 M NaCl in 20 column volumes, flow 1 mL/min and 1 mL elution fractions were collected. All purification fractions were tested for HNL activity using Feigl-Anger test paper³⁰ in a 384-well plate in 100 mM citrate buffer pH 4.0 and 3 mM racemic mandelonitrile as the substrate. The mixture was incubated for 20 min. Finally, pH of positive fractions was determined with a pH indicator. The fractions between 100 and 200 mM NaCl elution showed HNL activity. Each active fraction was concentrated 10 times through centrifugation via MCWO 10 Vivaspin 500 (Sartorius) and the buffer was exchanged to 25 mM potassium phosphate buffer, pH 6.0. Protein samples were stored at -20°C .

BN PAGE and in gel HNL activity detection. Specifically, 15 μL of each concentrated purification fraction or flow through or 2 μL of total protein extract were applied on a NativePAGE™ Novex® 4–16% Bis-Tris protein gel, and HNL activity assay was performed after the electrophoretic run. The procedure was performed similar to that described previously³³. Afterwards, the gel was stained by silver staining as described⁵⁰ with the following modifications: fix step over-night and 30 min incubation in water after the first ethanol wash step, in order to recover the original gel dimensions, because the gel reduced its size after incubation in ethanol. The gel was stored in 12% acetic acid. Bands of interest were excised and stored at -20°C in 10% ethanol.

Mass spectrometry and transcriptome data integration. Excised protein bands were tryptically digested, dissolved in 0.1% formic acid and separated by nano-RP-HPLC using a 70 min gradient. The samples were ionized in the nanospray source equipped with nanospray tips and analyzed in a Thermo LTQ-FT mass spectrometer operated in positive ion mode, applying alternating full scan MS (m/z 400 to 2,000, 50,000 resolution) in the ion cyclotron and MS/MS by collision induced dissociation of the five most intense peaks in the ion trap with dynamic exclusion enabled.

The LC-MS/MS data were analyzed by searching the translated *D. tyermannii* transcriptome and known contaminants with Proteome Discoverer 1.3 and Mascot 2.3 (1% FDR, min. two rank 1 peptides with min. Mascot ion score 20 and precursor mass error lower than 10 ppm required for protein identification). Hits were subjected to BLAST alignment against NCBI non-redundant public protein database.

Isoenzyme gene isolation. Genomic DNA from *D. tyermannii* was isolated with the PowerPlant® Pro DNA Isolation Kit (MO BIO Laboratories Inc.) according to the provided manual. Specific primers were designed for amplification of the isoenzymes (Supplementary Table 13) and genes were amplified by PCR. PCR products were isolated from a 1% agarose gel after electrophoresis and sequenced (LGC Genomics). Results obtained by Sanger and Roche/454 sequencing were compared and corrected in case of inconsistency.

Cloning. The list of primers, gBlocks® Gene Fragments and synthetic genes is reported in Supplementary Tables 13 and 14. Electrocompetent *E. coli* strains as described below were transformed by electroporation and positive clones were selected on LB agar plates with the appropriate antibiotics. *E. coli* TOP 10F' strain was used for vector amplification and protein expression after transformation of pMS470 vector. *E. coli* BL21 Star (DE3) strain was employed for expression of *DtHNL1-4*.

Isotigs screening. Synthetic genes were ordered after codon optimization for expression in *E. coli* (GeneArt® Gene synthesis, Thermo Fisher Scientific). The genes were cloned into the pMS470 vector (*NdeI/HindIII*). Clones were selected on ampicillin (100 mg/L).

***DtHNL1* expression.** A synthetic gene of isotig02643 was cloned into the pEHISTEV⁵¹ vector (*NcoI/HindIII*), in order to add the His-TEV tag at the *N*-terminus of the protein. Clones were selected on kanamycin (50 mg/L).

***DtHNL2, 3 and 4* expression.** Genes amplified from gDNA were cloned into the pJET1.2 vector (CloneJET PCR Cloning Kit, Thermo Fisher Scientific) and clones were selected on ampicillin (100 mg/L). Plasmids were isolated and used as a template for a second PCR with primers designed for cloning into the pEHISTEV vector (*NcoI/HindIII*), in order to add the His-TEV tag at the *N*-terminus of the protein. Clones were selected on kanamycin (50 mg/L). Sequences were confirmed by Sanger sequencing (LGC Genomics).

***DtHNL1* and *PtaIso02775* mutants.** Optimized nucleotide sequences coding for parts of the protein and containing the desired mutation(s) were purchased as gBlock® Gene Fragments (Integrated DNA Technologies). pEHISTEV containing *DtHNL1* or *PtaIso02775* was amplified with appropriate primers and the two fragments (Vector and the specific gBlock®) were assembled by the Gibson Assembly® Method. Clones were selected on kanamycin (50 mg/L). The sequences were confirmed by Sanger sequencing (Microsynth AG).

Protein Expression. Protein expression was performed in shake flasks in LB medium supplemented with the specific antibiotic at 37 °C. Induction was performed by addition of 0.5 mM IPTG at OD₆₀₀ 0.7, followed by incubation at 25 °C for 20 hours. After the cultivation, cell pellets were suspended in the appropriate buffer (*Isotig screening*: 50 mM potassium phosphate buffer, pH 6.0. *Protein purification*: 20 mM sodium phosphate, 0.5 M NaCl, 10 mM imidazole, pH 7.4).

Protein purification. Cells were disrupted by sonication (80% duty cycle, 7 output, 6 min). Affinity chromatography was performed with a HisTrap FF 5 mL column with standard protocol at 4 °C. Start buffer: 20 mM sodium phosphate, 0.5 M NaCl, 10 mM imidazole at pH 7.4. Elution buffer: 20 mM sodium phosphate, 0.5 M NaCl, 500 mM imidazole, pH 7.4. Elution method: gradient from 0 to 100% elution buffer, 20 column volumes. Fractions containing *DtHNL* were combined and desalted (HiPrep 26/10 desalting column). Purified protein fractions were stored in 50 mM sodium phosphate pH 6.5, at –80 °C.

Size-exclusion chromatography was performed with ÄKTA Avant 25 (GE Healthcare) equipped with a Superdex 200 10/300 GL column (GE Healthcare) at 4 °C. The column was pre-equilibrated with 150 mM NaCl, 10 mM Tris-HCl pH 8. The protein (0.5 ml; 1 mg/ml in 10 mM Tris-HCl pH 8) was loaded onto a column with a flow rate of 0.1 mL/min. The absorbance of the eluent was monitored at 280 and 254 nm. A Gel-Filtration-Standard (BioRad) was diluted 10x and separated under the same conditions.

Determination of *DtHNL* activity. *Standard assay.* The enzymatic activity was quantified as described⁵². The standard reaction was carried out in 96-well plates in 50 mM sodium citrate-phosphate buffer pH 5.0 and 15 mM racemic or (*R*)-mandelonitrile as the substrate, previously dissolved in 3 mM sodium citrate-phosphate buffer pH 3.5. Control reactions contained storage buffer instead of protein. Benzaldehyde formation was detected at 280 nm for 10 min. One unit is defined as the amount of enzyme that catalyzes the formation of 1 μmol of benzaldehyde in 1 min.

Activity at different pH values was performed using the following 50 mM buffers: HCl-potassium chloride pH 2.0 and 2.5; sodium citrate-phosphate pH 2.5–6.5.

Michaelis-Menten curves were determined at standard conditions using concentrations of (*R*)-mandelonitrile from 0.009 mM to 18 mM and 0.01 μg of *DtHNL*. K_m and v_{max} were calculated on basis of non-linear regression using Sigma Plot™ 11.0. Obtained values are based on the average of three independent experiments.

Activity at different temperatures. Determination of the optimal reaction temperature was performed in cuvettes. A final volume of 1 mL contained 100 μL of purified protein and 700 μL of pre-warmed 50 mM citrate phosphate buffer, pH 5.0. The reaction was initiated by the addition of 200 μL of substrate solution (60 mM racemic mandelonitrile in 3 mM citrate phosphate buffer, pH 3.5). Enzymatic activity was measured from 10 to 50 °C at 280 nm for 10 min.

Enzyme stability. *pH stability.* 1 mg/mL of purified protein was incubated in 50 mM sodium citrate phosphate buffer pH 2.5, 4.0 at 5–8 °C. At certain time-points, an aliquot of enzyme was diluted to 0.01 mg/mL. *HNL* activity was measured under standard conditions.

Rapid qualitative hydroxynitrile lyase assay. The assay detects the cyanogenesis reaction and is based on detection of released HCN *via* a Feigl–Anger test paper³⁰. Generally, the reaction was performed in 100 mM sodium citrate buffer at pH 4.5 with 13 mM mandelonitrile. The reaction was carried out at room temperature until blue spots were detected.

SeMet-*DtHNL1* expression and purification. *E. coli* BL21 (DE3) Star expressing tagged *DtHNL1* were cultivated in a shake flask in minimal medium (M9 salts 5X, 2% (w/v) glucose, 2 mM MgSO₄, 0.01 mg/mL thiamine, 0.01 mg/mL FeCl₃) supplemented with 50 mg/L kanamycin and 50 mg/L selenomethionine. Induction was performed at OD₆₀₀ 0.5 by addition of 0.5 mM IPTG, and the culture was incubated at 25 °C for 38 hours. After harvesting and cell disruption, SeMet-*DtHNL1* was purified by affinity chromatography (NiSepharose 6 Fast Flow resin, GE Healthcare). Elution was performed with 20 mM sodium phosphate, 0.5 M NaCl, 300 mM imidazole, pH 7.4. Fractions containing SeMet-*DtHNL1* were combined and desalted (PD10 Desalting columns, GE Healthcare). Protein was stored at –20 °C in 50 mM potassium phosphate buffer pH 6.0. For crystallization, the buffer was exchanged to 10 mM Tris-HCl pH 8 by diluting and re-concentrating the enzyme in 2 mL Ultra Centrifugal Filters (Amicon). The final protein concentration of native *DtHNL1* was 4 mg ml^{–1} and of SeMet *DtHNL1* was 3 mg ml^{–1}.

Crystallization and structure determination. Crystallization experiments were performed with an ORYX 8 robot (Douglas Instruments) using the sitting drop vapor-diffusion method in 96-well plates at 16 °C. Screening was performed using commercial screens Morpheus Screen MD 1–46, JCSG + MD1–37 (Molecular Dimensions) and Index HT HR2–144 (Hampton Research). Optimization of crystallization conditions was performed manually by the sitting drop vapor-diffusion method in Crystal Clear Duo crystallization frames at 16 °C.

Native crystals of *DtHNL1* were obtained by mixing 0.5 μL 4 mg/mL protein sample (in 10 mM Tris-HCl pH 8.0) with 1 μL reservoir solution (0.9 M NaNO₃; Na₂HPO₄; (NH₄)₂SO₄ mix, 0.1 M Tris-bicine buffer pH 8.5 and 30% (w/v) polyethylene glycol monomethyl ether 550 & polyethylene glycol 20k; Morpheus condition C9). Additionally, native crystals were also grown by mixing 1 μL 4 mg/mL protein sample (in 10 mM Tris-HCl pH 8.0) with 0.5 μL reservoir solution (0.1 M 2-(4-(2-hydroxyethyl)-1-piperazinyl) ethanesulfonic acid pH 7.5 and 10% (w/v) polyethylene glycol; JSCG condition B4). Crystals appeared after 2–3 days. SeMet-*DtHNL1* crystals were obtained in 0.2 M sodium thiocyanate, 20% (w/v) polyethylene glycol 3350. A 1:1 ratio of protein and screening

solutions was used, using protein concentration of 3 mg/mL (in 10 mM Tris-HCl pH 8.0). Crystals appeared within 2–3 days. After supplementation of 30% glycerol, the crystals were flash-cooled in liquid-nitrogen.

Soaking experiments were performed with the native *DtHNL1* crystals (grown as described above). Crystalline 4-hydroxybenzaldehyde - HBA, (*R*)-mandelonitrile - MXN or benzoic acid - BEZ were added to a crystallization drop with a small CryoLoop. After an incubation period of 30 s, 1 min, 5 min and 15 min, crystals were harvested, flash-cooled in liquid nitrogen and used for data collection.

All datasets were collected at 100 K at beamlines ID29 and BM14 at the ESRF (Grenoble, France) and at beamline XRD1 at Elettra (Trieste, Italy). Data were processed using the XDS program package⁵³ or iMos-flm⁵⁴/SCALA⁵⁵. The AutoSol Program^{56,57} and the AutoBuild Program⁵⁸ from the PHENIX software suite⁵⁹ were used to define the selenium heavy metal-atom sites using a SeMet-*DtHNL1* SAD data set, as well as to build an initial model. The resulting model was completed manually in Coot⁶⁰ and refined with PHENIX. Difference electron density too large for a water molecule was observed in the putative active site of the enzyme (Supplementary Fig. 7a), which could not be fitted using known buffer components or compounds of the crystallization conditions. Therefore, we did not interpret this portion of the electron density.

For all datasets from soaked crystals, molecular replacement was performed with Phaser-MR⁵⁶. The previously obtained SeMet-*DtHNL* structure was used as a search template. The resulting model was completed manually in Coot and refined with PHENIX. The occupancies of the ligands refined to values of 70–80%. Final structures were validated using Molprobity⁶¹. Detailed data processing and structure refinement statistics are summarized in Supplementary Table 6. The atomic coordinates and structure factors have been deposited in the Protein Data Bank under the accession codes 5E46 (*DtHNL1* SeMet), 5E4B (*DtHNL1*-MXN), 5E4D (*DtHNL1*-BEZ), 5E4M (*DtHNL1*-HBA).

Cyanohydrin Synthesis. Synthesis of cyanohydrins was carried out in a biphasic system as described in detail by Wiedner *et al.*⁵², except for 1-phenylethanone which was used in 0.3 M concentration. The aqueous phase contained 3 mg of purified *DtHNL1*. After acetylation, samples were analyzed by isothermal GC at 110 °C for 20 min. Retention times: internal standard triisopropylbenzene (IS) 1.6 min; 1-phenylethanone 4.1 min; 2-hydroxy-2-phenylpropanenitrile acetate 9.1 min. A negative control reaction (non-enzymatic background) was set up in the same conditions, with buffer instead of the enzyme solution. Apparent kinetic parameters were determined with 0.05 mg of *DtHNL1* (final concentration 0.1 mg/mL). Different concentrations of benzaldehyde were used (500–10 mM), while the amount of HCN was constant (2 M). The specific activity was obtained by determination of (*R*)-mandelonitrile formation during the first 30 min. In a biphasic system the benzaldehyde is predominantly in the organic phase and the *de facto* substrate concentration in the buffer phase was calculated by determination of the partition coefficient MTBE-H₂O of benzaldehyde. Each reaction was performed in duplicate as two different independent experiments. K_m and v_{max} were calculated on basis of non-linear regression by using Sigma Plot™ 11.0.

References

- Kassim, M. A. & Rumbold, K. HCN production and hydroxynitrile lyase: a natural activity in plants and a renewed biotechnological interest. *Biotechnol. Lett.* **36**, 223–8 (2014).
- Dadashipour, M. & Asano, Y. Hydroxynitrile lyases: insights into biochemistry, discovery, and engineering. *ACS Catal.* **1**, 1121–1149 (2011).
- Sharma, M., Sharma, N. N. & Bhalla, T. C. Hydroxynitrile lyases: At the interface of biology and chemistry. *Enzyme Microb. Technol.* **37**, 279–294 (2005).
- Hussain, Z. *et al.* Characterization of two bacterial hydroxynitrile lyases with high similarity to cupin superfamily proteins. *Appl. Environ. Microbiol.* **78**, 2053–5 (2012).
- Wiedner, R., Gruber-Khadjawi, M., Schwab, H. & Steiner, K. Discovery of a novel (*R*)-selective bacterial hydroxynitrile lyase from *Acidobacterium capsulatum*. *Comput. Struct. Biotechnol. J.* **10**, 58–62 (2014).
- Dadashipour, M., Ishida, Y., Yamamoto, K. & Asano, Y. Discovery and molecular and biocatalytic properties of hydroxynitrile lyase from an invasive millipede, *Chamberlinius hualienensis*. *Proc. Natl. Acad. Sci.* **112**, 10605–10610 (2015).
- Gruber-Khadjawi, M., Fechter, M. H. & Griengl, H. In *Enzyme Catalysis in Organic Synthesis* (eds Drauz, K., Gröger, H. & Griengl, H.) 947–990 (Wiley-VCH Verlag GmbH & Co. KGaA, 2012).
- Lanfranchi, E. *et al.* Mini-review: recent developments in hydroxynitrile lyases for industrial biotechnology. *Recent Pat. Biotechnol.* **7**, 197–206 (2013).
- Purkharthofer, T., Skranc, W., Schuster, C. & Griengl, H. Potential and capabilities of hydroxynitrile lyases as biocatalysts in the chemical industry. *Appl. Microbiol. Biotechnol.* **76**, 309–20 (2007).
- Winkler, M., Glieder, A. & Steiner, K. In *Comprehensive Chirality* (eds Carreira, E. M. & Yamamoto, H.) **7**, 350–371 (2012).
- Hajnal, I. *et al.* Biochemical and structural characterization of a novel bacterial manganese-dependent hydroxynitrile lyase. *FEBS J.* **280**, 5815–5828 (2013).
- Okrob, D., Metzner, J., Wiechert, W., Gruber, K. & Pohl, M. Tailoring a stabilized variant of hydroxynitrile lyase from *Arabidopsis thaliana*. *ChemBioChem* **13**, 797–802 (2012).
- Weis, R. *et al.* Biocatalytic conversion of unnatural substrates by recombinant almond *R*-HNL isoenzyme 5. *J. Mol. Catal. B Enzym.* **29**, 211–218 (2004).
- Hickel, A., Heinrich, G., Schwab, H. & Griengl, H. Screening for hydroxynitrile lyases in plants. *Biotechnol. Tech.* **11**, 55–58 (1997).
- Hernández, L., Luna, H., Ruíz-Terán, F. & Vázquez, A. Screening for hydroxynitrile lyase activity in crude preparations of some edible plants. *J. Mol. Catal. B Enzym.* **30**, 105–108 (2004).
- Asano, Y. *et al.* Screening for New hydroxynitrilases from plants. *Biosci. Biotechnol. Biochem.* **69**, 2349–2357 (2005).
- Kassim, M. A., Sooklal, S. A., Archer, R. & Rumbold, K. Screening for hydroxynitrile lyase activity in non-commercialised plants. *South African J. Bot.* **93**, 9–13 (2014).
- Omelchenko, M. V., Galperin, M. Y., Wolf, Y. I. & Koonin, E. V. Non-homologous isofunctional enzymes: a systematic analysis of alternative solutions in enzyme evolution. *Biol. Direct* **5**, 31 (2010).
- Dreveny, I., Gruber, K., Glieder, A., Thompson, A. & Kratky, C. The hydroxynitrile lyase from almond: A lyase that looks like an oxidoreductase. *Structure* **9**, 803–815 (2001).
- Guterl, J. K. *et al.* Uneven twins: Comparison of two enantiocomplementary hydroxynitrile lyases with α/β -hydrolase fold. *J. Biotechnol.* **141**, 166–173 (2009).

21. Nakano, S., Dadashipour, M. & Asano, Y. Structural and functional analysis of hydroxynitrile lyase from *Baliospermum montanum* with crystal structure, molecular dynamics and enzyme kinetics. *Biochim. Biophys. Acta - Proteins Proteomics* **1844**, 2059–2067 (2014).
22. Asano, Y. & Kawahara, N. A New S-Hydroxynitrile Lyase from *Baliospermum montanum*—Its structure, molecular dynamics simulation, and improvement by protein engineering. *Ind. Biotechnol.* **12**, 91–97 (2016).
23. Lauble, H., Miehlich, B., Förster, S., Wajant, H. & Effenberger, F. Crystal structure of hydroxynitrile lyase from *Sorghum bicolor* in complex with the inhibitor benzoic acid: A novel cyanogenic enzyme. *Biochemistry* **41**, 12043–12050 (2002).
24. Trummler, K., Fo, S., Pfizenmaier, K. & Wajant, H. Expression of the Zn²⁺-containing hydroxynitrile lyase from flax (*Linum usitatissimum*) in *Pichia pastoris*—utilization of the recombinant enzyme for enzymatic analysis and site-directed mutagenesis. **139**, 19–27 (1998).
25. Alagöz, D., Tükel, S. S. & Yildirim, D. Purification, immobilization and characterization of (R)-hydroxynitrile lyase from *Prunus amygdalus turcomanica* seeds and their applicability for synthesis of enantiopure cyanohydrins. *J. Mol. Catal. B Enzym.* **101**, 40–46 (2014).
26. Asif, M. & Bhalla, T. C. Hydroxynitrile lyase of wild apricot (*Prunus armeniaca* L.): Purification, characterization and application in synthesis of enantiopure mandelonitrile. *Catal. Letters* **146**, 1118–1127 (2016).
27. Ueatrongchit, T., Tamura, K., Ohmiya, T., H-Kittikun, A. & Asano, Y. Hydroxynitrile lyase from *Passiflora edulis*: purification, characteristics and application in asymmetric synthesis of (R)-mandelonitrile. *Enzyme Microb. Technol.* **46**, 456–465 (2010).
28. Nuylert, A., Ishida, I. & Asano, Y. Effect of glycosylation on the biocatalytic properties of hydroxynitrile lyase from the passion fruit *Passiflora edulis* - A Comparison of Natural and Recombinant Enzymes. *ChemBioChem.* **18**, 257–265 (2016).
29. Wajant, H., Forster, S., Selmar, D., Effenberger, F. & Pfizenmaier, K. Purification and characterization of a novel (R)-mandelonitrile lyase from the fern *Phlebodium aureum*. *Plant Physiol.* **109**, 1231–1238 (1995).
30. Krammer, B., Rumbold, K., Tschemmerneegg, M., Pöchlauer, P. & Schwab, H. A novel screening assay for hydroxynitrile lyases suitable for high-throughput screening. *J. Biotechnol.* **129**, 151–61 (2007).
31. Kofod, H. & Fson, R. E. Y. J. Cyanogenesis in species of the fern genera *Cystopteris* and *Davallia*. **8**, 1509–1511 (1969).
32. Lizotte, P. a. & Poulton, J. E. Catabolism of cyanogenic glycosides by purified vicianin hydrolase from squirrel's foot fern (*Davallia trichomanoides* Blume). *Plant Physiol.* **86**, 322–324 (1988).
33. Lanfranchi, E. *et al.* Bioprospecting for hydroxynitrile lyases by Blue Native PAGE coupled HCN detection. *Curr. Biotechnol.* **4**, 111–117 (2015).
34. Altschul, S. Basic Local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
35. Radauer, C., Lackner, P. & Breiteneder, H. The Bet v1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol. Biol.* **8**, 286 (2008).
36. Der, J. P., Barker, M. S., Wickett, N. J., DePamphilis, C. W. & Wolf, P. G. De novo characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. *BMC Genomics* **12**, 99 (2011).
37. Bracco, P., Busch, H., von Langermann, J. & Hanefeld, U. Enantioselective synthesis of cyanohydrins catalysed by hydroxynitrile lyases – a review. *Org. Biomol. Chem.* **14**, 6375–6389 (2016).
38. Ognyanov, V. I., Datcheva, V. K. & Kyler, K. S. Preparation of chiral cyanohydrins by an oxynitrilase-mediated transcyanation. *J. Am. Chem. Soc.* **113**, 6992–6996 (1991).
39. Andexer, J. *et al.* An R-selective hydroxynitrile lyase from *Arabidopsis thaliana* with an alpha/beta-hydrolase fold. *Angew. Chem. Int. Ed. Engl.* **46**, 8679–81 (2007).
40. Alagöz, D., Tükel, S. S. & Yildirim, D. Enantioselective synthesis of various cyanohydrins using covalently immobilized preparations of hydroxynitrile lyase from *Prunus dulcis*. *Appl. Biochem. Biotechnol.* **177**, 1348–1363 (2015).
41. Sigdel, S., Hui, G., Smith, T. J., Murrell, J. C. & Lee, J.-K. Molecular dynamics simulation to rationalize regioselective hydroxylation of aromatic substrates by soluble methane monooxygenase. *Bioorg. Med. Chem. Lett.* **25**, 1611–5 (2015).
42. Funhoff, E. G., Bauer, U., García-Rubio, I., Witholt, B. & van Beilen, J. B. CYP153A6, a soluble P450 oxygenase catalyzing terminal-alkane hydroxylation. *J. Bacteriol.* **188**, 5220–7 (2006).
43. Zimmer, T., Ohkuma, M., Ohta, A., Takagi, M. & Schunck, W. H. The CYP52 multigene family of *Candida maltosa* encodes functionally diverse n-alkane-inducible cytochromes P450. *Biochem. Biophys. Res. Commun.* **224**, 784–9 (1996).
44. Alonso, H. *et al.* Structural and mechanistic insight into alkane hydroxylation by *Pseudomonas putida* AlkB. *Biochem. J.* **460**, 283–93 (2014).
45. Hofrichter, M., Kellner, H., Pecyna, M. J. & Ullrich, R. Fungal unspecific peroxygenases: heme-thiolate proteins that combine peroxidase and cytochrome P450 properties. *Adv. Exp. Med. Biol.* **851**, 341–68 (2015).
46. Gruber, K. & Kratky, C. Biopolymers for biocatalysis: Structure and catalytic mechanism of hydroxynitrile lyases. *J. Polym. Sci. Part A Polym. Chem.* **42**, 479–486 (2004).
47. Lee, E.-J. & Facchini, P. Norcochlorine synthase is a member of the pathogenesis-related 10/Bet v1 protein family. *Plant Cell* **22**, 3489–3503 (2010).
48. Kuipers, R. K. *et al.* 3DM: systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins* **78**, 2101–13 (2010).
49. Damborský, J., Petrek, M., Banás, P. & Otyepka, M. Identification of tunnels in proteins, nucleic acids, inorganic materials and molecular ensembles. *Biotechnol. J.* **2**, 62–7 (2007).
50. Blum, H. *et al.* Improved silver staining of plant-proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* **8**, 93–99 (1987).
51. Liu, H. & Naismith, J. H. A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expr. Purif.* **63**, 102–111 (2009).
52. Wiedner, R. *et al.* Improving the properties of bacterial R-selective hydroxynitrile lyases for industrial applications. *ChemCatChem* **7**, 325–332 (2015).
53. Kabsch, W. *XDS Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
54. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 271–281 (2011).
55. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 72–82 (2006).
56. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
57. Terwilliger, T. C. *et al.* Decision-making in structure solution using Bayesian estimates of map quality: The PHENIX AutoSol wizard. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **65**, 582–601 (2009).
58. Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64**, 61–69 (2007).
59. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).
60. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
61. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010).

Acknowledgements

This research has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 289646 (KYROBIO), from the OMICS Center Graz grant of the Austrian Ministry for Science, Research and Economy (BMWF) and the Federal Ministry of Science, Research and Economy (BMWF), the Federal Ministry of Traffic, Innovation and Technology (bmvit), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol, the Government of Lower Austria and ZIT - Technology Agency of the City of Vienna through the COMET-Funding Program managed by the Austrian Research Promotion Agency FFG. We also acknowledge the support of the beamline staff at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) and Elettra (Trieste, Italy) during data collection.

Author Contributions

E.L., M.W. and A.G. designed the study, E.L. and K.S. designed the discovery study, E.L. performed discovery experiments; E.L. and E.M.K. performed enzymatic characterization and mutations experiments; E.L., M.W., K.S. and M.G.K. designed the synthesis study; E.L. and M.G.K. performed synthesis experiments; R.B.G. designed the proteomics experiments. B.D. performed the proteomic experiments; E.L. B.D. and R.B.G. analyzed proteomics data, G.G.T. designed transcriptome assembly, J.H. performed transcriptome assembly, H.J.J. and T.V.D.B. generated a 3DM database, E.L. collected and analyzed bioinformatic data; K.G. and T.P.K. designed structural and mechanism studies, T.P.K. and M.D. performed structural experiments; T.P.K. and K.G. collected and analyzed structural data; E.L., T.P.K., K.G. and M.W. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing Interests: The authors declare no competing financial interests.

How to cite this article: Lanfranchi, E. *et al.* Enzyme discovery beyond homology: a unique hydroxynitrile lyase in the Bet v1 superfamily. *Sci. Rep.* 7, 46738; doi: 10.1038/srep46738 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017