Elisa Lanfranchi Msc.

NEW HYDROXYNITRILE LYASE FROM WHITE RABBIT'S FOOT FERN: The route from discovery to perspectives for applications

DOCTORAL THESIS

to achieve the university degree of

Doktorin der Naturwissenschaften

submitted to

Graz University of Technology

Supervisor

Ao.Univ.-Prof. Mag.rer.nat. Dr.rer.nat. Anton Glieder

Institute of Molecular Biotechnology

Co-Supervisor: Dr. Margit Winkler

Graz, November 2015

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present doctoral thesis.

7.10.15
Date

Elisa Bonfanti
Signature

"El gran llibre, sempre obert
I que cal esforçar-se a llegir,
és el de la Naturalesa"

Antoni Gaudí

"The great book, always open and which
we should make an effort to read,
is that of Nature"

Antoni Gaudí

La Sagrada Familía, Barcelona

Acknowledgments

At the end of my studies I would like to thank all people that contributed either with dedicated participation and efforts as well as an incredible support along the (almost) four years.

Thanks to my parents for the support during my studies as well as my choice of moving to Graz.

Thanks to my supervisor Toni Glieder for giving me the opportunity to do my PhD in his group as well as working for such appealing and challenging project. Thanks for the freedom and the possibility to carry on my ideas and encourage to be independent.

A big thanks to Margit! Co-supervisor, ex-group leader, project manger, colleague…. (whatever she wants to be defined). Thanks for her everyday life support, for boosting me for the best work and the incredible dedication to improve my writing skills.

Thanks to my master student Eva. She gave a real practical support for this thesis. Thanks for her patience during my super stressed days, especially the last minute desperate whatsapp messages before report deadlines.

Thanks to the summer students Kasia and Claudia, who gave a contribution to complete parts of the experiments.

Thanks to the structural biology people! Karl, Tea and the master student Matthias. Their strong contribution brought an additional value to this work. Thanks for the close collaboration during writing and for setting up a great story together. Last but not least thank for answering all my random questions.

Thanks to all other people, which contributed to this work, by teaching, supervising, Chinese translations and technical support: Kerstin, Mandana Karin, Barbara, Ruth, Gerhard, Jürgen and Ivan.

Thanks to all the third floor group! The survivors and all the people which passed over the years! Margit, Thorsten, Katrin, Aleks, Kamila, Clautsch, Barbara, Ela, Viktorija, Michi, Holly, Gerli and all students especially Bernd, Kathi, Filippo and Jacek as well as the current students Carol, Eli, Manfred and Lukas.

A big thanks to the amazing Glieder group!

Thanks to the old crew, when I started at the second floor long time ago, which made me feeling welcome immediately! Martina, Laura, Luki, Andrea M., Patrick, Rama, Thomas, Flo, Aleks, Anna, Andreas and Julia.

Thanks also to the many people currently spread over the floors: Martina, Mali, Luki, Jasmine, Christian, Mohammed, Patri, Rafi, Flo, Anna and Astrid. + Rui and Fitri.

An extra thanks to Flo, Thomas, Anna and Andrea M. for the amazing time together chatting, drinking, dancing 90ies music etc…. Graz life won't be the same without you!

Thanks to all the people from ACIB/IMBT, a place where the working atmosphere is always nice and collaborative.

Thanks to Andrea and Gloria for their incredible help in the beginning.

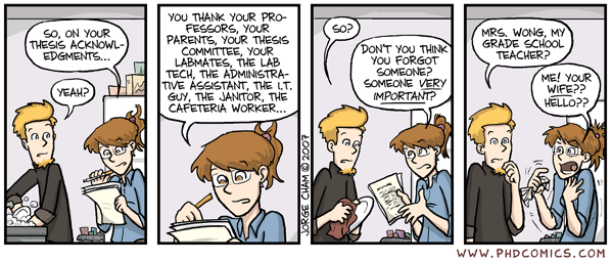
Since the private life also contributes for a successful work, I would like to thank all the friends from Graz! Particularly: Nina, Elvira, Giulio, Filippo, Federica, Francesco, Giacomo, Erika, Kamila, Robert, Rauhl, Beppe, Roberto, Marco and Gianluca.

A special thanks to Andrea and Anna for the awesome friendship, enjoy the time together, the nice discussions as well as their incredible support during my hyper-stressed days.

Thanks to my sister Arianna, my brother Stefano and all my Italian friends, who encouraged me from 750 km far. Especially: Benny, Fabio, Coppy, Ele, Cri, Marco, Eli and Manuel.

Thanks to Tribeka, the cafeteria where part of this thesis has been written.

Last but not least, thanks to PhDcomics, which make me smile for every shade of the PhD life…and… if you feel your name is missing, this comic strip is dedicated to you:



"Piled Higher and Deeper" by Jorge Cham www.phdcomics.com

Abstract

Enzyme discovery strongly contributes to the development of new successful and competitive biocatalytic processes. In this context, the investigation of new hydroxynitrile lyases (HNLs) for cyanohydrin synthesis is still a hot topic, due to the small number of so far discovered and fully characterized HNLs. Herein, the way from the discovery of a new HNL from white rabbit's foot fern (*Davallia tyermannii*) to perspectives for applications is described. A versatile assay for bioprospecting HNLs was developed from scratch and applied together with transcriptomic and proteomic data. The interconnection of the different approaches enabled the identification of a truly new enzyme with an unpredictable sequence and protein fold, which was not known so far for proteins with HNL activity. *Dt*HNL's tertiary structure and catalytic mechanism were elucidated by colleagues of the University of Graz and revealed a unique enzyme within the entire protein superfamily. Finally, the obtained high recombinant expression level and biocatalytic properties of *Dt*HNL open perspectives for a new generation of biocatalysts for stereoselective cyanohydrin synthesis.

Zusammenfassung

Die Entdeckung neuer Enzyme leistet einen wichtigen Beitrag zur Entwicklung neuer erfolgreicher und wettbewerbsfähiger Biokatalyseprozesse. Aufgrund der geringen Anzahl der bisher entdeckten und vollständig charakterisierten Oxynitrilasen (HNLs), ist die Untersuchung neuer HNLs immer noch ein aktuelles Thema. In dieser Dissertation, ist der Weg von der Entdeckung bis zur Anwendung einer neuen HNL aus dem Farn *Davallia tyermannii* beschrieben. Ein vielseitiger Assay zur Entdeckung von HNLs in der Natur wurde von Grund auf entworfen und zusammen mit Transkriptom und Proteom-Daten angewandt. Die Verknüpfung der verschiedenen Ansätze ermöglichte die Identifizierung eines neuen Enzyms mit einer Aminosäuresequenz und Proteinfaltung wie sie bisher für Proteine mit HNL-Aktivität noch nicht bekannt war. Die Aufklärung der Tertiärstruktur der neuen *Dt*HNL und ihres katalytischen Mechanismus durch Kolleginnen der Universität Graz offenbarte die Einzigartigkeit des Enzyms innerhalb der gesamten Protein-Superfamilie. Schlussendlich eröffnen die hohen rekombinanten Expressionslevels und die biokatalytischen Eigenschaften der *Dt*HNL Perspektiven für eine neue Generation von Biokatalysatoren zur stereoselektiven Cyanhydrinproduktion.

Keywords

Enzyme discovery, hydroxynitrile lyase, plant hydroxynitrile lyase, fern hydroxynitrile lyase, *Davallia teyermanii*, cyanohydrin, Bet v 1, biocatalysis, enzyme characterization, enzyme immobilization, protein structure, catalytic mechanism, *Pichia pastoris*, recombinant expression.

Table of content

Introduction and aims of this thesis

Hydroxynitrile lyases (HNLs) catalyze the enantioselective cleavage of α-cyanohydrins into the respective aldehyde or ketone and hydrocyanic acid. They are assigned to the EC numbers 4.1.2.10, 4.1.2.11, 4.1.2.46 and 4.1.2.47 and alternative names are: oxynitrilase, mandelonitrile lyase, hydroxynitrilase[1]. HNLs have been identified in different natural sources such as plants, bacteria[2,3] as well as animals[4]. In plants they play a role in the defense system upon herbivorous attacks. HNLs are very diverse enzymes, which belong to different protein superfamilies. They do not share any feature except for catalyzing the same chemical reaction, therefore they are defined as non homologous isofunctional enzymes and a result of convergent evolution. Nowadays we know five different classes of HNL, however the number will likely increase, due to those sequences and/or structures that have not been identified yet[1,5].

HNLs are an appealing tool for the synthesis of α-cyanohydrins, key building blocks for follow up chemical and chemo-enzymatic reactions. The condensation reaction has three important characteristics: a chiral center is formed, the carbon chain is prolonged by one carbon atom and a further versatile functional group is introduced (the nitrile). Therefore, cyanohydrins find applications in pharmaceutical, agrochemical and cosmetic industries[1]. Moreover, HNLs are an interesting target for the detoxification of cyanogenic food crops[6] and an attracting nature-inspired approach for wheat seed protection[7].

Thousands of plant species show cyanogenic activity and they potentially express a HNL, however, the number of isolated and characterized HNLs is limited to few tens, and even less are systematically employed for biocatalysis[5]. The reason concerns the difficulties to identify the sequence of truly new HNLs, due to their highly diverse nature. The majority has been isolated from their natural host by several steps of protein purification followed by further work for sequence determination, to end up in a laborious and time consuming process.

The major aims of this thesis were:

1. Discover the sequence of a novel HNL from fern
2. Acquire knowledge of the new enzyme by enzymological characterization, structure and reaction mechanism determination
3. Apply the new recombinant HNL to biocatalysis

Cyanogenesis is a widespread phenomenon in the plant kingdom and it is based on the fast release of poisoning HCN in response to herbivore attacks. Cyanogenic glycosides are synthetized and stored in the cell. After tissue disruption, a β-glycosidase cleaves the sugar from the aglycone moiety and provides the cyanohydrin substrate to the HNL. Finally, HCN is released[8]. The degradation of cyanohydrins is a simple reaction and it spontaneously occurs in an aqueous system at pH higher than 5. HNLs can catalyze the same reaction also at lower pH values, and more rapidly. Becker and Pfeil already proposed in 1965 the essential features for HNL activity: A general base is necessary to abstract a proton from the hydroxyl group of the cyanohydrin; a second important condition is the stabilization of the cyanide ion, therefore a positively charged residue is necessary[9]. HNLs derive from different ancestors and they are a product of convergent evolution, and they fulfill the above described requirements in different manners.

FAD- dependent HNLs belong to the **GMC oxidoreductase superfamily**. These are *N*-glycosylated proteins and (*R*)-selective. Members have been found in *Rosaceae* such as *Prunus amygdalus* (*Pa*HNL)[10], *Prunus serotina* (*Ps*HNL)[11], *Prunus mume* (*Pm*HNL)[12] and *Eriobotrya japonica* (*Ej*HNL)[13,14]. The structure of *Pa*HNL was solved and the reaction mechanism elucidated[15]: the reaction follows a general acid/base catalysis, where His497 acts as the general base, abstracting the proton from the hydroxyl group. CN^- is stabilized by an overall positive charge in the active site, as a result of several positively charged residues present in the active site. The Flavin cofactor does not seem involved in catalysis, however it is necessary for the protein stability[1].

(*S*)-Selective HNLs have been found in the *Euphorbiaceae* family, for instance in *Manihot esculenta* (*Me*HNL), *Hevea brasiliensis* (*Hb*HNL) and *Baliospermum montanum* (*Bm*HNL). They are more than 50% identical and belong to the **alpha/beta hydrolase superfamily**[16,17]. The cyanohydrin is coordinated by Ser80, Thr11 and Lys236 (*Hb*HNL numbering). An hydrogen bond network is established between Asp207, His235 and Ser80, which acts as general base and abstracts the proton from the hydroxyl group. The stabilization of CN^- is accomplished by Lys236, facilitating the C-C bond cleavage. In the last step, the cyanide molecule is protonated by His235 and HCN is released[18]. Interestingly, **esterases** and hydroxynitrile lyases share the same structure and catalytic

triad, however esterases lack the Lys and Thr in the active site. Padhi and coauthors demonstrated that two amino acid substitutions Gly12Thr and Met239Lys are sufficient for the conversion of the plant esterase SABP2 into a hydroxynitrile lyase[19]. The HNL from *Arabidopsis thaliana* (*At*HNL) is the only isolated (*R*)-selective HNL with an alpha/beta hydrolase fold. Interestingly, it is also the only identified protein with HNL activity in a non cyanogenic plant[20]. *At*HNL is 45% identical and 67% similar to *Hb*HNL: the triad Ser-His-Asp is conserved, however the Lys is replaced by a Met and other residues in the active site differ as well. The protein structure of *At*HNL was solved and the reaction mechanism elucidated[21]. His236 and Asn12 coordinate the hydroxyl group of the substrate, while the backbone amide group of Ala13 and Phe82 establish an hydrogen bond with the nitrile group. The triad Asp208-His236-Ser81 acts as the general base, but the proton is abstracted from the hydroxyl group by His236 and not from the Ser. The cyanide molecule is stabilized by the backbone amide group of Ala13 and Phe82 and it is finally protonated by the catalytic Ser81, which in turn is protonated by Hys236 (opposite of *Hb*HNL)[21].

HNL from *Sorghum bicolor* (*Sb*HNL) reveals an **alpha/beta hydrolase fold**, however its sequence is substantially different from the above described HNLs and more similar to **serine carboxypeptidases**[22]. *Sb*HNL is an *N*-glycosylated heterotetramer, the result of a post-translational cleavage of a large polypeptide. The architecture of the active site differs from the other (*S*)-HNLs and a different catalytic mechanism was proposed: Ser158 and the carboxylate group of Trp270 coordinate the hydroxyl group of the cyanohydrin. The latter was indicated as the general base in the first step of deprotonation. The cyanide anion is stabilized by a water molecule. Finally, the same water molecule transfers a proton from Trp270 to CN^- and hydrocyanic acid is formed[22].

The tertiary structure of the HNL from *Linum usitatissimum* (*Lu*HNL) has not been published yet, however its sequence shows similarity to homologous **Zn^{2+} containing alcohol dehydrogenases**[23]. *Lu*HNL is a homodimer, where each subunit contains two zinc ions. Trummler and coauthors claimed that zinc plays a significant role in protein stability, but it is not involved in catalysis. Moreover, substitution of conserved residues revealed *Lu*HNL and Zn^{2+} ADHs have same features for ion coordination and maintaining an active

structure[23]. Interestingly, *Lu*HNL has different stereoselectivity for cyanohydrin synthesis depending on the substrate[1].

One of the latest discovered HNLs classes belongs to the **cupin superfamily**. Cupin-like HNLs do not occur in plants but in bacteria such as *Granulicella tundricola* (*Gt*HNL) and *Acidobacterium capsulatum* (*Ac*HNL)[2]. The tertiary structure of *Gt*HNL was solved and an in-depth study concerning the role of the metal ion of these metalloproteins was done. *Gt*HNL activity is strictly dependent on the presence of manganese, furthermore two His residues were identified close to the metal ion and they have been proposed as the general base and positive charge for HNL activity[24].

The first arthropod HNL from the millipede *Chamberlinius hualienensis* (*Chua*HNL) was isolated and its sequence determined. *Chua*HNL is an 18 kDa glycosylated protein. Its tertiary structure has not been solved yet and the amino acid sequence is not similar to any other above described HNL, hence, a new HNL class is expected[4].

The last group is rather heterogeneous and it consists of different HNLs with unknown amino acid sequence and protein fold. HNLs from *Passiflora edulis* (*Pa*HNL)[25], *Ximenia americana* (*Xa*HNL)[26], *Prunus amygdalus turcomanica* (*Pat*HNL)[27] and *Phlebodium aureum* (*Pha*HNL)[28] are some examples. The proteins were isolated from their natural hosts by several purification steps. Subsequently, purified enzyme preparation was employed for biochemical characterizations and cyanohydrin synthesis. Based on the results, they have been described as new proteins, with distinct properties compared to those of known HNLs.

In addition to the fascinating diversity concerning fold and enzymology, HNLs are strongly demanded for sustainable and green production of enantiomerically pure cyanohydrins. **Chapter 1** is an overview about HNL applications in biocatalysis. The review strongly focuses on the recent patent literature and highlights processes, where cyanohydrins are either the final product or an intermediate for the synthesis of highly functionalized chiral compounds.

The search for novel compounds for new pharmaceuticals, or environmental friendly agrochemicals and more sustainable industrial processes is in continuous progress.

Therefore, the investigation of different HNLs can contribute to broaden the substrate range or to fulfill industrial need such as protein stability or biocatalyst availability. However, the discovery of new HNLs is strongly limited by their diversity, therefore by exploitation of common bioinformatic tools (e.g. blastp or CD search) it is not possible to identify new members of this class of enzymes. HNLs are mainly identified from scratch by a long and tedious process of protein purification and subsequent sequence identification. The latter is indeed necessary for the production of the recombinant biocatalyst as well as protein engineering.

Chapter 2 describes a new method for fast HNL identification, without the obligation of classical chromatography protein purification. The principle is based on the visualization of HNL activity after protein electrophoresis and it takes the advantages of Blue Native PAGE (BN PAGE)[29]. Proteins are separated according to their native molecular weight, while the enzymatic activity remains intact. This method is applicable to any kind of protein preparation from different sources, therefore suitable for bioprospecting HNLs. This strategy was validated by screening a variety of plants and tissues as well as different HNLs. Finally, the unknown HNL from *Prunus domestica* kernels was chosen as proof of concept for sequence determination.

Previously an HNL from the fern *Phlebodium aureum* was discovered by Wajant and coworkers in 1995[28]. The protein was purified from the leaves of the cyanogenic fern and a specific activity of 19,000 U/mg revealed *Pha*HNL to be one of the fastest enzymes for (*R*)-mandelonitrile degradation. *Pha*HNL is a homomultimer of 20 kDa subunits, it exists in three isoforms and as it did not resemble any characteristics of known HNLs, Wajant *et al.* expected it to belong to a new enzyme class[28]. To date, *Pha*HNL protein sequence and fold are still unsolved, making it an attractive target for enzyme discovery. However, cyanogenesis in *pteridophytes* is not restricted to *P. aureum*, and it is commonly found in a number of ferns as *Davalliaceae*[30] as well as bracken fern *Pteridium aquilinum*[31]. The white rabbit's foot fern *Davallia tyermannii* has been in the horticultural trade for long time and it is not subjected to any legal restrictions as the tropical fern *P. aureum*. With confirmed cyanogenic activity of *D. tyermannii* leaves and fiddleheads, the method developed in chapter 2 was applied for the identification of *Dt*HNL. In order to determine

the *Dt*HNL sequence, the established workflow required a combination of proteomics data and transcriptome or genome information. The transcriptome is actually preferred, since it consists of all the expressed genes, whereas non-translated regions are not considered. Because little information on fern genes was available in public databases, the *D. tyermannii* transcriptome was sequenced. The approach finally allowed to discover four different *Dt*HNL isoenzymes and determine their gene and amino acid sequence without any previous knowledge as discussed in detail in the first part of Chapter 3.

The second part of Chapter 3 consists of a comprehensive study of the novel *Dt*HNL isoenzymes with a major focus on enzymology. The four recombinant *Dt*HNL isoenzymes were purified by affinity chromatography and *Dt*HNL activity and stability were investigated at different pH values and temperatures. Furthermore, kinetic parameters in synthesis and cyanogenesis direction were determined. A high resolution x-ray structure of *Dt*HNL was solved. *Dt*HNL is a homodimer and belongs to the Bet v 1 protein superfamily[32], confirming indeed the discovery of a new HNL class. Furthermore, the crystals of *Dt*HNL together with substrate/product ligands were obtained. Residues interacting with the substrate were substituted by site-specific mutagenesis, and substrate binding architecture and catalytic mechanism were proposed.

The acquired knowledge of *Dt*HNL was applied for the investigation of homologous HNLs in the bracken *P. aquilinum* (Chapter 3) and within the Bet v 1 superfamily as well (Chapter 3, Supplementary Results). The two sections discuss the uniqueness of *Dt*HNL within the protein superfamily and the expectation of a new HNL type expressed by *P. aquilinum*. The scouting of the Bet v 1 superfamily was performed *in silico* by sequence, and structure alignment aided by the 3DM system.

The last part of Chapter 3 concerns applications of *Dt*HNL in biocatalysis. The purified enzyme was employed for the synthesis of different cyanohydrins in a biphasic reaction system. *Dt*HNL efficiently converted several aldehydes into the respective cyanohydrins in high enantio-purity in short time with the wild type enzyme. The reaction was less efficient when a ketone was used as the substrate. This study revealed *Dt*HNL as a highly promising biocatalyst for industrial cyanohydrin synthesis.

Biocatalyst availability is an important aspect to consider for an overall competitive process and it can be achieved with a well-established recombinant production system. To date, there are various choices of different host organisms both prokaryotic or eukaryotic systems which address different needs. The methylotrophic yeast *Pichia pastoris* (formally *Komagatella phaffii*)[33] is one of the major used eukaryotic host organisms for protein production due to several advantages. It can reach high cell densities during fermentation, it is easy to manipulate and able to make post-translational modifications e.g. protein glycosylation, disulfide bond formation and proteolytic processing[34]. Finally, it was declared as "Generally Recognized As Safe" (GRAS) by the Food and Drug Administration (FDA). The recombinant protein production of *Dt*HNL in *Pichia pastoris* is described in Chapter 4. High levels of the target protein were achieved under the control of the strong and tightly regulated AOX1 promoter[35]. Nevertheless, *Dt*HNL isoenzymes were also well expressed by *E. coli*. The same chapter addresses a second important aspect, the robustness of the biocatalyst. Specifically, optimal synthesis conditions to reach enantiomerically pure cyanohydrins require stable enzymes at pH lower than 4.5 in biphasic or microaqueous systems. Cross-linked enzyme aggregate (CLEA) of *Dt*HNL was developed from *P. pastoris* cell free lysate. CLEA-*Dt*HNL stability under acidic conditions and improvements for (*R*)-mandelonitrile and (*R*)-hydroxypivaldeyde synthesis are discussed. This study contributes to the development of a powerful and robust biocatalyst for cyanohydrin synthesis.

In Chapter 5 the procedure discussed in chapter 2 for HNL detection in protein mixture is described in a form of step by step protocol.

References

1. Winkler, M., Glieder, A. & Steiner, K. in *Comprehensive Chirality* **7**, 350–371 (2012).
2. 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).
3. Hussain, Z. et al. Characterization of two bacterial hydroxynitrile lyases with high similarity to cupin superfamily proteins. *Appl. Environ. Microbiol.* **78**, 2053–2055 (2012).
4. 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).
5. Dadashipour, M. & Asano, Y. Hydroxynitrile lyases: insights into biochemistry, discovery, and engineering. *ACS Catal*. **1**, 1121–1149 (2011).
6. Narayanan, N. N., Ihemere, U., Ellery, C. & Sayre, R. T. Overexpression of hydroxynitrile lyase in cassava roots elevates protein and free amino acids while reducing residual cyanogen levels. *PLoS One* **6**, e21996 (2011).
7. Halter, J. G. et al. Induced cyanogenesis from hydroxynitrile lyase and mandelonitrile on wheat with polylactic acid multilayer-coating produces self-defending seeds. *J. Mater. Chem. A* **2**, 853–858 (2014).
8. Poulton, J. E. Cyanogenesis in plants. *Plant Physiol*. **94**, 401–5 (1990).
9. Becker, W. & Pfeil, E. Uber das Flavinenzym *D*-Oxynitrilase. *Biochem. Z*. **346**, 301–321 (1966).
10. Dreveny, I., Gruber, K., Glieder, A., Thompson, A & Kratky, C. The hydroxynitrile lyase from almond: a lyase that looks like an oxidoreductase. *Structure* **9**, 803–15 (2001).
11. Cheng Ping, I. & Poulton, J. E. Cloning of cDNA of *Prunus serotina* (*R*)-(+)-mandelonitrile lyase and identification of a putative FAD-binding site. *Plant Cell Physiol*. **34**, 1139–1143 (1993).
12. Fukuta, Y. et al. Characterization of a New (*R*)-hydroxynitrile lyase from the japanese apricot *Prunus mume* and cDNA cloning and secretory expression of one of the isozymes in *Pichia pastoris*. *Biosci. Biotechnol. Biochem*. **75**, 214–220 (2011).

13. Zhao, G.-J., Yang, Z.-Q. & Guo, Y.-H. Cloning and expression of hydroxynitrile lyase gene from *Eriobotrya japonica* in *Pichia pastoris*. *J. Biosci. Bioeng*. **112**, 321–5 (2011).
14. Ueatrongchit, T., Kayo, A., Komeda, H., Asano, Y. & H-Kittikun, A. Purification and characterization of a novel (*R*)-hydroxynitrile lyase from *Eriobotrya japonica* (Loquat). *Biosci. Biotechnol. Biochem*. **72**, 1513–1522 (2008).
15. Dreveny, I., Andryushkova, A. S., Glieder, A., Gruber, K. & Kratky, C. Substrate binding in the FAD-dependent hydroxynitrile lyase from almond provides insight into the mechanism of cyanohydrin formation and explains the absence of dehydrogenation activity. *Biochemistry* **48**, 3370–7 (2009).
16. 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).
17. Guterl, J. K. et al. Uneven twins: Comparison of two enantiocomplementary hydroxynitrile lyases with α/β-hydrolase fold. *J. Biotechnol*. **141**, 166–173 (2009).
18. Gruber, K., Gartler, G., Krammer, B., Schwab, H. & Kratky, C. Reaction mechanism of hydroxynitrile lyases of the alpha/beta-hydrolase superfamily: the three-dimensional structure of the transient enzyme-substrate complex certifies the crucial role of LYS236. *J. Biol. Chem*. **279**, 20501–10 (2004).
19. Nedrud, D. M. et al. Uncovering divergent evolution of α/β-hydrolases: a surprising residue substitution needed to convert *Hevea brasiliensis* hydroxynitrile lyase into an esterase. *Chem. Sci*. **5**, 4265–4277 (2014).
20. Andexer, J. et al. An R-selective hydroxynitrile lyase from *Arabidopsis thaliana* with an α/β-hydrolase fold. *Angew. Chemie - Int. Ed*. **46**, 8679–8681 (2007).
21. Andexer, J. N. et al. Hydroxynitrile lyases with α/β-hydrolase fold: two enzymes with almost identical 3D structures but opposite enantioselectivities and different reaction mechanisms. *Chembiochem* **13**, 1932–9 (2012).
22. 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).

23. Trummler, K., Fo, S., Pfizenmaier, K. & Wajant, H. Expression of the Zn^{2+}-containing hydroxynitrile lyase from flax (*Linum usitatissimum*) in *Pichia pastoris*—utilization of the recombinant enzyme for enzymatic analysis and site-directed mutagenesis. *Plant Science* **139**, 19–27 (1998).
24. Hajnal, I. et al. Biochemical and structural characterization of a novel bacterial manganese-dependent hydroxynitrile lyase. *FEBS J*. **280**, 5815–5828 (2013).
25. 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).
26. Kuroki, G. W. & Conn, E. E. Mandelonitrile lyase from *Ximenia americana* L.: stereospecificity and lack of flavin prosthetic group. *Proc. Natl. Acad. Sci. U.S.A*. **86**, 6978–81 (1989).
27. 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).
28. Wajant, H., Forster, S., Selmar, D., Effenberger, F. & Pfizenmaier, K. Purification and characterization of a novel (*R*)-mandelonitrile lyase from the fern *phlebodium* a*ureum*. *Plant Physiol*. **109**, 1231–1238 (1995).
29. Schägger, H. & von Jagow, G. Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal. Biochem*. **199**, 223–231 (1991).
30. Kofod, H. & Fson, R. E. Y. J. Cyanogenesis in species of the fern genera C*ystopteris* and *Davallia*. *Phytochemistry* **8**, 1509–1511 (1969).
31. Hadfieldt, P. R. H. & Dyer, A. F. Cyanogenesis in gametophytes and young sporophytes of bracken. *Biochem. Sys. Ecol*. **16**, 9–13 (1988).
32. Radauer, C., Lackner, P. & Breiteneder, H. The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol. Biol*. **8**, 286 (2008).
33. Kurtzman, C. P. Biotechnological strains of *Komagataella* (*Pichia*) *pastoris* are *Komagataella phaffii* as determined from multigene sequence analysis. *J. Ind. Microbiol*. *Biotechnol.* **36**, 1435–1438 (2009).

34. Cereghino, G. P. L., Cereghino, J. L., Ilgen, C. & Cregg, J. M. Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. *Curr. Opin. Biotechnol*. **13**, 329–332 (2002).
35. Cregg, J. M., Madden, K. R., Barringer, K. J., Thill, G. P. & Stillman, C. A. Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. *Mol. Cell. Biol*. **9**, 1316–1323 (1989).

Chapter 1

Mini-Review: Recent Developments in Hydroxynitrile Lyases for Industrial Biotechnology

Elisa Lanfranchi[1], Kerstin Steiner[1], Anton Glieder[1], Ivan Hajnal[1], Roger A. Sheldon[2], Sander Van Pelt[2] and Margit Winkler[1]*

Published in: Recent Patents on Biotechnology. 7 (2013) 3, 197-206.

[1]Austrian Centre Industrial Biotechnology, Petersgasse 14 8010 Graz, Austria
[2]CLEA Technologies B.V. Delftechpark 34 2628XH Delft, The Netherlands

Corresponding Author:
Margit Winkler
Austrian Centre Industrial Biotechnology, Petersgasse 14/III, 8010 Graz, Austria
Email: margit.winkler@acib.at; Tel: +43-316-8739333; Fax +43-316-8739308

Send Orders for Reprints to reprints@benthamscience.net

Recent Patents on Biotechnology, 2013, *7*, 197-206

Mini-Review: Recent Developments in Hydroxynitrile Lyases for Industrial Biotechnology

Elisa Lanfranchi[a], Kerstin Steiner[a], Anton Glieder[a], Ivan Hajnal[a], Roger A. Sheldon[b], Sander van Pelt[b] and Margit Winkler[a,*]

[a]*ACIB GmbH (Austrian Centre of Industrial Biotechnology GmbH), c/o Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria;* [b]*CLEA Technologies B.V., Delftechpark 34, 2628XH Delft, The Netherlands*

Received: February 02, 2013 Revised: April 12, 2013 Accepted: June 21, 2013

Abstract: Hydroxynitrile lyases (HNLs) catalyze the cleavage as well as the formation of cyanohydrins. The latter reaction is valuable for the stereoselective C-C bond formation by condensation of HCN with carbonyl compounds. The resulting cyanohydrins serve as versatile building blocks for a broad range of chemical and enzymatic follow-up reactions. A significant number of (*R*)- and (*S*)-selective HNLs are known today and the number is still increasing. HNLs not only exhibit varying substrate scope but also differ in sequence and structure. Tailor-made enzymes for large-scale manufacturing of cyanohydrins with improved yield and enantiomeric excess are very interesting targets, which is reflected in a solid number of patents. This review will complement and extend our recent review with a strong focus on applications of HNLs for the synthesis of highly functionalized, chiral compounds with newest literature, recent and current patent literature.

Keywords: Asymmetric synthesis, biocatalysis, biotransformation, cascade reactions, C-C coupling, chiral building blocks, cyanogenesis, cyanohydrin, enantioselectivity, enzyme engineering, HNL, hydrocyanic acid, hydroxynitrile lyase, industrial process, mandelonitrile lyase, plant enzymes, oxynitrilase, stereoselective.

INTRODUCTION

Hydroxynitrile lyases (HNLs, EC 4.1.2.10, EC 4.1.2.11, EC 4.1.2.46 and EC 4.1.2.47; alternative names: oxynitrilase, hydroxynitrilase, mandelonitrile lyase, hydroxymandelonitrile lyase) catalyze the enantioselective cleavage and stereoselective synthesis of cyanohydrins. Cyanogenesis has been reported not only in many plant species like *Rosaceae, Linaceae, Euphorbiaceae, Clusiaceae, Olacaceae, Gramineae* and *Polypodiaceae* but also in other organisms like bacteria, fungi, lichen, arthropods or insects [1,2]. However, HNLs were also discovered in non-cyanogenic plants such as *Arabidopsis thaliana* [3].

In 1837, Wöhler and Liebig reported hydroxynitrile lyase activity for the first time as they investigated the formation of hydrocyanic acid (HCN, alternative name: prussic acid) – the cyanogenesis reaction [4]. The reverse reaction was first observed by Rosenthaler in 1908, who used emulsin – an emulsion derived from bitter almonds (*Prunus amygdalus*) – to prepare (*R*)-mandelonitrile. The first stereoselective enzyme catalyzed addition of HCN to benzaldehyde was thus discovered [5]. The enzyme responsible for this activity was an (*R*)-hydroxynitrile lyase. Since then, a number of other HNLs, both (*R*)- and (*S*)-selective were discovered. Interestingly, hydroxynitrile lyase activity is found from polypeptides with completely unrelated sequential and structural features and seems to be the result of convergent evolution [6]. HNLs from *Rosaceae* are FAD (flavin adenine dinucleotide) containing monomeric glycoproteins. Their natural substrate is mandelonitrile and this is a relatively conserved group of enzymes. In contrast, the group of FAD-independent HNLs is diverse and varies largely in physicochemical properties (primary sequence, molecular weight, oligomerization state, glycosylation, structure and natural substrate). HNLs differ in their stereospecificity and substrate range and have therefore found distinct applications for asymmetric syntheses of chiral building blocks not only in research laboratories but also at industrial level [7-9].

Chemical as well as enzyme catalyzed condensation of HCN to carbonyl compounds is reversible and the equilibrium is dependent on pH, solvent composition and the available concentrations of carbonyl compound, HCN and α-hydroxynitrile product (alternative name: cyanohydrin). The reverse reaction, the so called cyanogenesis of a hydroxynitrile, is mostly used for the discovery of new HNLs, whereas the C-C bond forming condensation is a synthetically very useful reaction: a new chiral center is formed, the carbon chain is prolonged by one carbon atom and an additional versatile functional group – the nitrile – is introduced. The cyanohydrins product serve as starting material for many enzymatic and chemical follow-up reactions [10]. For industrial applications it is often essential to immobilize the enzyme in order to optimize its performance and facilitate its recovery and reuse [11]. This can involve adsorption on solid supports such as silica or ion exchange resins or carrier-free

*Address correspondence to this author at the ACIB GmbH (Austrian Centre of Industrial Biotechnology GmbH), Petersgasse 14/III, A-8010 Graz, Austria; Tel: +43-316-8739333; Fax: +43-316-873-9308;
E-mail: margit.winkler@acib.at

2212-4012/13 $100.00+.00

immobilization by precipitation from aqueous buffer and subsequent cross-linking with a bifunctional reagent, such as glutaraldehyde, to afford so-called cross-linked enzyme aggregates (CLEAs). The latter have been widely applied in biocatalytic transformations [12,13], including enantioselective hydrocyanations with a variety of both (*R*) and (*S*)-selective HNLs [14]. Recently, several aspects of HNL research were reviewed [9,10,15-23]. In this article we are reviewing mainly the current patent literature and scientific literature from the last two years. For an overview, Table **1** summarizes HNL enzymes covered in current patents, including enzyme source, substrate range and selectivity.

R-SELECTIVE HYDROXYNITRILE LYASES

The group of (*R*)-selective HNLs is known for several decades and consists of the largest number of variants. Most of these variants were characterized as preparations of plant parts [25,35-37] or as purified enzymes. For a long time, defatted almond flour or the soluble protein fraction obtained from it was used as biocatalyst. Interestingly, another protein fraction - without HNL activity - derived from almond flour can be used as a stabilizer for this (*R*)-HNL [38].

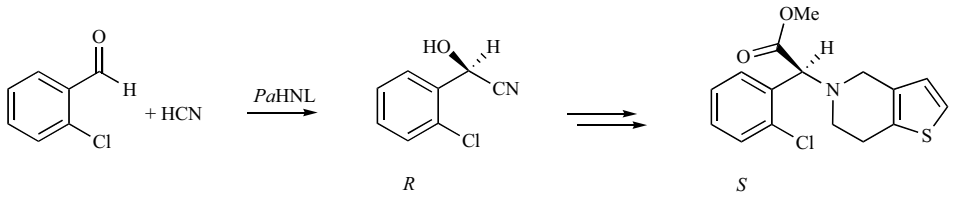
Table 1. Hydroxynitrile Lyase Sources Specifically Mentioned in Patent Literature.

Enzyme Source and Abbreviation	Spec.	Substrate range	Heterologous Expression System	Reference
Arabidopsis thaliana (cress) *At*HNL	R	Aliphatic and aromatic (substituted), heteroaromatic, α,β-unsaturated aldehydes, and methylketones	*E. coli*	[24]
Chaenomeles speciosa (flowering quince) *Cs*HNL	R	Aliphatic and aromatic ketones	--	[25]
Cyclonia oblonga (quince)	--	--	--	[26]
Eriobotrya japonica (loquat) *Ej*HNL	R	Aliphatic and aromatic heteroaromatic and aldehydes	*P. pastoris*	[25]
Hevea brasiliensis (rubber tree) *Hb*HNL	S	Aliphatic, aromatic, heteroaromatic, α,β-unsaturated aldehydes and methylketones	*E. coli, P. pastoris, S. cerevisiae*	[27-29]
Linum usitatissimum (flax) *Lu*HNL	R, S	R: aliphatic aldehydes and methylketones, S: aromatic (with at least on C separating) ketones	*E. coli, P. pastoris*	[30]
Malus communis (apple)	R	Only aromatic aldehydes tested	--	[26]
Malus pumila (paradise apple)	--	--	--	[26]
Manihot esculenta (cassava) *Me*HNL	S	Aliphatic, aromatic, heteroaromatic, α,β-unsaturated aldehydes, methylketones and aromatic ketones	*E. coli, P. pastoris, S. cerevisiae Leishmania tarentolae*	[29,31,32]
Passiflora edulis (passion fruit) *Pe*HNL	R	Aromatic aldehydes, aliphatic less	--	[25]
Prunus amygdalus (almond) *Pa*HNL	R, S	Aliphatic, aromatic, heteroaromatic, α,β-unsaturated aldehydes and methylketones	*P. pastoris*	[26,33]
Prunus armeniaca (apricot) *Par(s)*HNL	R	Sterically demanding aromatic aldehydes	--	[26]
Prunus avium (wild cherry)	R	Aliphatic and aromatic aldehydes	--	[26]
Prunus domestica (plum)	R	Aliphatic and aromatic aldehydes	--	[26]
Prunus laurocerasus (cherry Laurel)	--	--	--	[26]
Prunus lyonii (Catalina cherry)	R	Only mandelonitrile tested	--	[26]
Prunus mume (Japanese apricot) *Pm*HNL	R	Aliphatic, aromatic, heteroaromatic and polycyclic aldehydes, methyl and cyclic ketones	*P. pastoris*	[26,34]
Prunus persica (peach)	R	Aliphatic and aromatic aldehydes	--	[25,26]
Prunus serotina (black cherry) *Ps*HNL	R	Aliphatic and aromatic aldehydes	--	[26]
Prunus sp.	--	--	--	[30]
Sorbus aucuparia	R	--	--	[25]
Sorghum bicolor (millet) *Sb*HNL	S	Aromatic, heteroaromatic aldehydes and methylketones	--	[30]

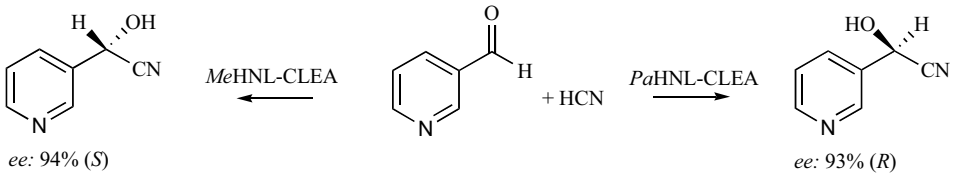
To allow for applications on a large scale, heterologous expression in microbial hosts is indispensable. However, because the *Prunus* HNL posed difficulties in functional recombinant expression in *E. coli*, the first recombinant (*R*)-HNL was *Lu*HNL from *Linum usitatissimum*, which has a different structure and substrate spectrum [39]. Seemingly, the expression of *Lu*HNL in *E. coli* was straight forward and was not patented; neither was the enzyme itself which was first purified in 1988 [40]. HNL isoenzymes 1 and 5 from *Prunus amygdalus* were the first (*R*)-HNLs from *Rosaceae*, that were functionally expressed in the eukaryotic expression system *Pichia pastoris* and patented [33]. Genes of other *Prunus* HNL isoenzymes have been known before [41] but initially successful recombinant expression of these glycosylated and disulfide bond containing enzymes had not been shown. In contrast to *Lu*HNL, HNLs from *Rosaceae* have a broad substrate spectrum accepting aromatic aldehydes as well as aliphatic aldehydes [42] and ketones. Today, an increasing number of (*R*)-HNL genes are known. Recently, Zhao *et al.* identified the coding sequence of the (*R*)-hydroxynitrile lyase from *Eriobotrya japonica* (loquat) and successfully expressed and secreted it from *Pichia pastoris* [43]. *Ej*HNL comprises 552 amino acids including a 25 amino acid signal peptide and shows 60-70% sequence identity to other HNLs from *Rosaceae* (i.e. *Prunus amygdalus*, *Prunus mume* and *Prunus serotina*). The recombinant protein showed similar catalytic properties as the natural *Ej*HNL reaching up to 98.6% *ee* in the biosynthesis of (*R*)-mandelonitrile from benzaldehyde [44].

(*R*)-2-Chloro-benzaldehyde cyanohydrin was reported by Breuer *et al.* to be a chiral building block for an oral antiplatelet agent [45]. This compound – and other substituted benzaldehyde cyanohydrin, derivatives were synthesized on preparative scale in good yield and enantiomeric excesses (98% and 83% for (*R*)-2-Cl-benzaldehyde cyanohydrin respectively) using defatted almond flour, the natural source of *Pa*HNL Scheme **1** [46]. Weis *et al.* used a *Pa*HNL isoenzyme 5 preparation, that was obtained as highly pure secreted enzyme by heterologous expression in *P. pastoris,* as a catalyst for (*R*)-2-Cl-benzaldehyde hydrocyanation and obtained 99.7% product yield with 90% *ee*, respectively [47]. For the same substrate, chimeric recombinant proteins based on type 1 HNL (*R*)-hydroxynitrile lyase and type 5 HNL (*R*)-hydroxynitrile lyase were patented in Japan. The combination of the high specific activity of HNL1 with the favorable pH and temperature stability of HNL5 yielded industrially versatile enzymes [48]. Additional improvement in activity and selectivity was achieved by protein engineering. The secreted *Pa*HNL5α_L1Q, A111G variant showed significantly higher specific activity for (*R*)-mandelonitrile formation compared to the secreted wild-type enzyme. Moreover, it exhibited satisfactory stability at pH 2.6 [49]. In a further example, mutein *Pa*HNL5α_N3I I108M A111G 4_E10 was cultivated on the 5 liter scale and the crude supernatant was concentrated. With the us of 16.7 mg of crude protein, 1 mol of 2-chloro-benzaldehyde was converted to 86.2% to the product in 98.5% *ee* within 4 h [26]. The above mentioned wealth of achieved improvements for the preparation of (*R*)-2-Cl-mandelonitrile was mainly related to the protein itself. However, also reaction engineering can have a valuable impact on the productivity of a process. When 2-Cl-benzaldehyde and HCN are added continuously to a stirred mixture of *Pa*HNL in methyl-*tert*-butylether (MTBE), 41.4% of the total weight of (*R*)-mandelonitrile can be obtained in 91.5% *ee* [50]. Furthermore, *Pa*HNL, e.g. can be immobilized on porous silica gel and then used in a homogenous mixture of aqueous buffer, MTBE and HCN. The total water content influences the reaction efficiency, whereby approximately 50% of water leads to quantitative conversion of 2-Cl-benzaldehyde [51].

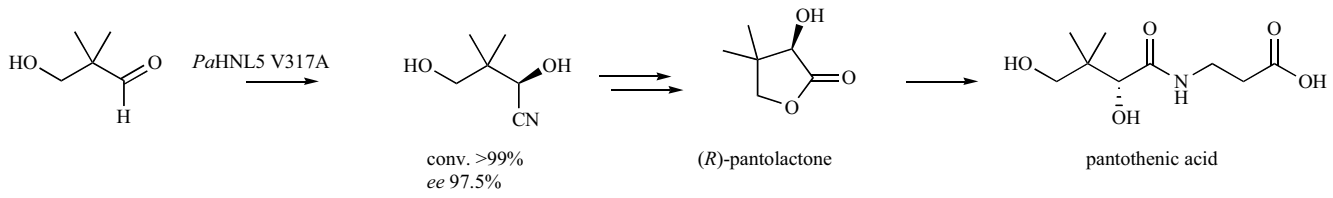
*Pa*HNL was also successfully immobilized as cross-linked enzyme aggregates by cross-linking with glutaraldehyde. The resulting *Pa*HNL CLEAs were highly effective catalysts for the enantioselective hydrocyanation of aldehydes under microaqueous conditions and could be recycled at least ten times without significant loss of activity [52]. Because such HNL-CLEAs perform exceptionally well in organic solvents, they can afford higher enantioselectivities than observed with the free enzymes owing to the essentially complete suppression of competing non-enzymatic hydrocyanation under these conditions. For example Roberge and coworkers obtained high enantioselectivities in the hydrocyanation of pyridine-3-carboxaldehyde Scheme **2** [53,54]. The latter is a difficult substrate for enantioselective hydrocyana-



Scheme (1). *Pa*HNL mediated synthesis of (*R*)-2-Cl-mandelonitrile.



Scheme (2). HNL-CLEAs as catalyst for the preparation of 3-pyridinecarbaldehyde cyanohydrins.



Scheme (3). Preparation of the chiral key-intermediate towards vitamin B5 synthesis.

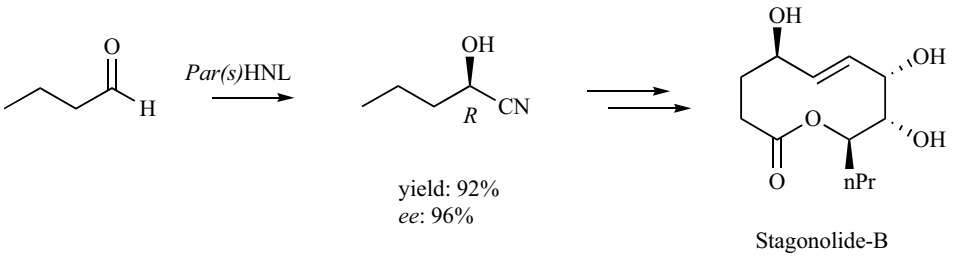
tion owing to the relatively facile non-enzymatic background reaction as a result of the electron-attracting properties of the pyridine ring [55].

Another target molecule that was triggered with HNL technology is (*R*)-pantothenic acid (vitamin B5). Its chiral key-intermediate – (*R*)-pantolactone – is accessible by hydrocyanation of hydroxypivalaldehyde (3-hydroxy-2,2-dimethylpropanal, see Scheme **3**. The two methyl groups adjacent to the aldehyde functionality imply a steric demand in the active site. Nevertheless, some HNLs formed the cyanohydrin in decent yield and *ee.* Recombinant *Lu*HNL (*Linum usitatissimum*) gave the desired product in 47% yield and 73% *ee* [56]. The special challenge with the substrate was to provide it to the enzyme in monomeric form: acidic conditions were beneficial and therefore only enzymes tolerant to low pH can be useful for this transformation on an industrial scale; native *Pa*HNL showed too low stability and afforded only 84% yield with 89% *ee* under thoroughly optimized conditions with large amounts of enzyme. Further improvement of yields and *ee*s could, however, only be achieved by enzyme engineering of the recombinant enzyme *Pa*HNL5 which is more stable at low pH than native *Pa*HNL. Following a semi-rational approach, a superior *Pa*HNL isoenzyme 5 mutant was discovered, in which Val317 was replaced by alanine. Surprisingly, the improvement was ascribed to a reduction of stabilizing hydrophobic interactions with the less favored (*S*)-enantiomer and not due to changes in steric hindrance. Summarizing, this *Pa*HNL5 V317A mutein afforded the cyanohydrin intermediate with >99% conversion and 97.5% *ee* in laboratory scale experiments [57].

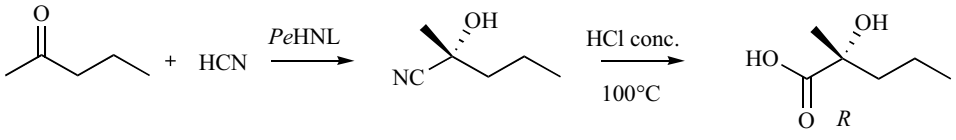
Stagonolides isolated from the fungal pathogen *Stagonospora cirsii* are nonenolides which show phytotoxic activities against some weeds and can be used as herbicides. The multi-step total synthesis of stagonolide B is started with the conversion of cheap commercially available *n*-butanal and HCN by the (*R*)-selective hydroxynitrile lyase from *Prunus armeniaca* (Par(s)HNL) see Scheme **4** [58].

Aldehydes are the classical substrates for HNL catalyzed hydrocyanation reactions; however, also ketones are transformed to their respective cyanohydrins. If the ketone substrate is not symmetric, the product represents a chiral quarternary carbon center, which is a valuable synthetic moiety that is difficult to obtain by classical chemical means. Up to now, only ketones with at least one small substituent (methyl) and cyclic ketones have been reported as HNL substrates. Notably, ketones are less reactive than aldehydes and the reaction equilibrium is often not in favor of the desired product; consequently, the reaction mixture may consist of high quantities of the substrates instead of the desired tertiary alcohol. The HNL of *Eriobotrya japonica* was recently used to convert 2-pentanone to its respective cyanohydrin which was then further hydrolyzed to give (*R*)-2-hydroxy-2-methylpentanoic acid in 32% overall yield and 24% *ee*. A similar yield was obtained with *Passiflora edulis* HNL, albeit the *ee* of 87% was significantly better Scheme **5** [25].

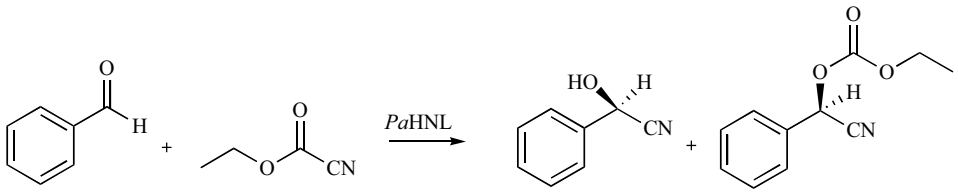
HCN as a reagent requires special care because it is volatile and highly toxic. A well ventilated fumehood, HCN detectors and special training for the operators are prerequisites for research in this field, especially when HCN is used in neat form. To overcome the risks of HCN handling, cyanides can also be provided as acetone cyanohydrin, trimethylsilyl



Scheme (4). Chemoenzymatic synthesis of Stagonolide-B from *n*-butanal.



Scheme (5). From methylketone to chiral quarternary carbon center.



Scheme (6). Chemo-enzymatic cascade with ethyl-cyanoformate as CN source and acylation agent.

cyanide and cyanoformate and the HNL catalyzed hydrocyanation reaction proceeds as transcyanation from these cyanide donors [59]. A beautiful transcyanation example is the use of ethyl cyanoformate as both an activated acyl donor as well as the source of cyanide for *Pa*HNL Scheme **6** [60]. Despite the sophisticated concept, mixtures of both the desired carbonate and mandelonitrile were obtained [59,60].

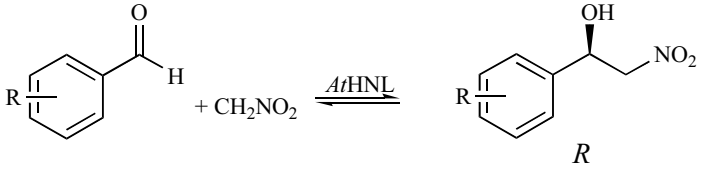
Until recently, the stereopreference of HNLs was ascribed to their three dimensional structure; it was believed that α,β-hydrolase fold enzymes exclusively display (*S*)-selectivity (*vide infra*). The discovery of an (*R*)-selective HNL with high sequence similarity to (*S*)-selective enzymes from *Hevea brasiliensis* and *Manihot esculenta* came as a surprise, especially because its origin was a non-cyanogenic plant: *Arabidopsis thaliana*. Substrate range and biophysical parameters of *At*HNL were determined [3,24,61] and the structure was solved (Pdb ID: 3DQZ) [62]. Although the enzyme is highly similar to the structures of (*S*)-selective HNLs (*Hb*HNL and *Me*HNL), including the catalytic Ser-His-Asp triad in the active site, several differences in the active site were identified, which are potentially for the diverging binding of the substrate and thus result in a different reaction mechanism. Based on docking studies and site-directed mutagenesis, a reaction mechanism for (*R*)-selective HNLs with α,β-hydrolase fold was proposed. In the first step, the cyanohydrin hydroxyl group is directly deprotonated by the His236 residue of the catalytic triad. The hydroxyl group also interacts with Asn12 to facilitate the deprotonation. Compared to *Hb*HNL, a Lys residue which stabilizes the negative charge of the released cyano group is missing. In contrast, the cyano group could interact with the backbone amide groups of Ala13 and Phe82 and thereby be stabilized by an oxyanion hole similar to those described in serine hydrolases. Finally, the cyanide is protonated indirectly by His236 with Ser81 as a mediator [62].

Interestingly, despite the similarity to *Hb*HNL and *Me*HNL, *At*HNL is significantly less stable at acidic pH (below pH 5.4) than the (*S*)-selective HNLs [61], which is rather deplorable because the non-enzymatic background reaction is more abundant at higher pH and thus reduces the *ee* of the product. This background reaction is also dependent on the water content of the reaction medium. Thus, a lot of effort was recently directed to stabilize *At*HNL at low pH or, alternatively, to reduce the water content in the reaction systems. The introduction of eleven surface mutations of *At*HNL and exchange of these amino acids for the corresponding amino acids from the more stable *Me*HNL resulted in a variant, which is active at pH 4.5 (23.6 U/mg initial rate activity in the cleavage of mandelonitrile, a half-life of about 13 h at 20°C, and 70% conversion with 99.3% *ee* at pH 4.5 and 0°C in a citrate-phosphate/MTBE two-phase system) [63]. In another approach, buffer-saturated (pH 6.5) MTBE was used in the reaction system either with precipitated protein or celite-immobilized or solgel-entrapped *At*HNL [64]. While fresh solgel-entrapped *At*HNL showed the most prominent improvement in activity and *ee* in the first round of reaction, celite-immobilized *At*HNL was much more stable during recycling (in an organic solvent-resistant, fine-woven nylon mesh 'teabag') and storage. It was shown that the water-content was crucial for optimal activity. In a subsequent paper excellent enantioselectivity was achieved using *E. coli* whole cells expressing *At*HNL in a mono-phasic micro-aqueous reaction system with MTBE. Importantly, the cells were recyclable using the 'tea-bag' approach without loss of enantioselectivity, however, decreasing conversion rates [65].

Recently, Asano and his group showed that *At*HNL also catalyzes the first (*R*)-selective HNL-catalyzed Henry reaction of aromatic aldehydes Scheme **7**. They achieved the highest enantioselectivity for benzaldehyde and $MeNO_2$ in a biphasic system at pH 7 with 50% *n*-butyl acetate (20% yield and 90% *ee*). It was shown that substituents on the aromatic ring did not influence the reactivity substantially. The best enantioselectivity was obtained with 2-naphthaldehyde, however, with low yield (7% yield, >99.9% *ee*) [66]. Aliphatic aldehydes were converted sluggishly, which was not surprising, as *At*HNL is also less active towards aliphatic aldehydes in standard hydrocyanation reactions [3].

S-SELECTIVE HYDROXYNITRILE LYASES

In 1989, Conn *et al.* discovered a hydroxynitrile lyase from the rubber tree [67] that turned out to be the first HNL exhibiting (*S*)-selectivity [68]. This HNL from *Hevea brasiliensis* (*Hb*HNL) and a similar enzyme from *Manihot esculenta* (*Me*HNL) belong to the α/β-hydrolase fold-type en-



Scheme (7). *R*-selective *At*HNL catalyzed asymmetric nitroaldol condensations.

zyme class with the typical catalytic triad aspartic acid, serine and histidine in the active site. These two enzymes exhibit high sequence identity and show a relatively broad substrate tolerance towards aliphatic, alicyclic, unsaturated, aromatic, and heteroaromatic aldehydes and ketones. The cDNA sequence coding for the enzyme *Me*HNL was first published in 1994 [69]. The primary sequence of *Hb*HNL as well as homologous enzymes with at least 80% protein sequence identity were protected by a patent application [27], followed by the enantioselective process towards the preparation of (*S*)-cyanohydrins [70]. From the reaction engineering viewpoint, the use of immobilized enzymes has advantages over a reaction in homogenous system: immobilized enzyme is easily removed from the reaction mixture, can be reused and facilitate product purification. *Me*HNL was, for example immobilized on nitrocellulose and used for the preparation of an array of (*S*)-cyanohydrins derived from both aldehydes and ketones [71]. Sintered clay, silica, alumina and silica-alumina carriers with pore sizes of 10-80 nm and a surface area of more than 20 m^2g^{-1} can also be used as carriers for *Me*HNL [72]. Porous silica gel, used in a homogenous reaction system consisting of MTBE and aqueous buffer with a water content of 50%, can be re-used up to 16 times for the preparation of (*S*)-mandelonitrile without loss of activity and selectivity [51]. Notably, there is no general HNL immobilization technique, because each enzyme works optimal on different carriers and reaction conditions and each setup needs separate optimization [73].

For industrial application, it is more convenient to overexpress the desired enzymes in heterologous systems than to use isolated enzyme from plants. Semba *et al.* claimed the expression of the *Me*HNL coding gene derived from cassava in yeasts. In particular, *Me*HNL was intracellularly expressed in *Saccharomyces cerevisiae* and *Pichia*, using vectors with the GAP or the AOX1 promoter, respectively [32]. In order to increase heterologous expression levels also in *E. coli* and thereby facilitating mutation studies, codon optimization proved useful. The codon use in plants differs from that in the prokaryotic host *E. coli* and therefore, the gene sequence of *Me*HNL was optimized: all codons which are used by *E. coli* with frequencies lower than 5% were replaced with more frequently used codons. Additional improvement of functional expression by correct folding was achieved by low temperature induction (17°C) [31,74].

In case of substrates with bulky substituents, conversions or enantiomeric excesses (*ees*) of wild-type (*S*)-HNLs were not satisfying. One reason appeared to be the rather narrow hydrophobic substrate entrance channel to the active site. This problem was tackled by the exchange of bulky amino acid residues of *Hb*HNL and *Me*HNL with less space demanding residues such as the tryptophan in position 128 for alanine or leucine [75-77]. In addition to this rational approach, random mutagenesis revealed a range of positions which yielded improved enzyme activity for the cleavage of the standard test substrate *rac*-mandelonitrile. Also site-saturation mutagenesis at selected positions and even the insertion of an amino acid between position 128 and 129 in the *Me*HNL sequence resulted in improved variants [29]. Enzyme engineering was also performed to improve *Me*HNL's heat tolerance. The interaction of helix D3' with helix A and β-sheet 2 was increased by the introduction of specific mutations. Substitution of glycine 165 by acidic amino acids such as aspartate or glutamate in helix D3' increased hydrogen bonding with basic residues. Alternatively, hydrophobic interaction was increased by substitution of valine 173 by leucine. These variants as well as other mutants and multiple mutants in the above mentioned regions showed increased heat stability and also stability against organic solvents [28].

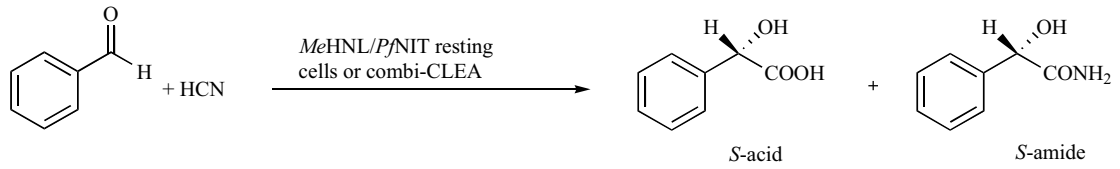
Histidine 103 was identified as a hotspot by random mutagenesis of wild-type *Me*HNL [78]. Specifically, substitutions to hydrophobic amino acids with aliphatic side chains (Leu, Val, Ile, Met and Cys) showed enhanced solubility during expression in *E. coli* (or a prokaryotic cell free expression system) and consequential higher activity compared to WT *Me*HNL, which is expressed mainly in inclusion bodies in *E. coli* [77,79]. However, no improvement of expression level and activity was observed when the same mutants were expressed in *P. pastoris* or *Leishmania tarentolae* or a eukaryotic cell free expression system, in which WT *Me*HNL already expressed as soluble protein [79]. Expression level of *Me*HNL in *E. coli* was about 7-fold improved by changing the valine in position 2 for a lysine [78]. Also substitutions for Ile, Asn, Arg and Gln showed a beneficial effect on the expression level which directly translated into increased activity of cell free extracts. A combination of V2I and H103L resulted in a 10-fold improved strain as compared to the wild-type (WT). Moreover, substitution of other lysine residues at the surface of the protein by prolines increased the expression level further [78]. In the same patent application mutants of enzymes with more than 90% sequence identity to *Hb*HNL as well as to *Me*HNL are claimed [78]. Analysis of several solvent-exposed Lys residues in *Me*HNL by site-saturation mutagenesis revealed that Pro-substitution at position 176, 199 and 224 resulted in higher solubility of the mutants. Combination of all three to a triple mutant added up to the highest activity due to an increase in solubility [80].

Recently, the first HNL, which was engineered from an esterase, was published [81]. Esterases and some HNLs belong to the α/β-hydrolase superfamily and they even share the typical conserved Asp, His and Ser catalytic triad in their catalytic site. However, the reaction mechanisms are different. Recently, two groups introduced HNL activity into esterases. Schwab *et al.* converted the bacterial esterase EstC from *Burkholderia gladioli* into an HNL by introduction of only one single mutation (S276K), which was sufficient to generate HNL activity and to abolish esterase activity [82]. Kazlauskas *et al.* switched the plant esterase SABP2 to an HNL with two point mutations (G12T and M239K). The enzyme showed strongly reduced esterase activity, but clearly detectable HNL activity with racemic mandelonitrile (20 mU/mg, k_{cat}/K_M = 72 $min^{-1}M^{-1}$). Enantioselectivity is expected to be low in cyanogenesis direction because the enzyme showed comparable activity for (*R*)- and (*S*)-mandelonitrile cleavage with 12.6 and 15.5 mU/mg, respectively. This result was confirmed in the synthesis direction, where the (*S*)-product showed 20% *ee* [81]. The sequence of the plant gene SABP2 is claimed in several patent applications not related to HNL chemistry but increased resistance of plants overexpressing such genes.

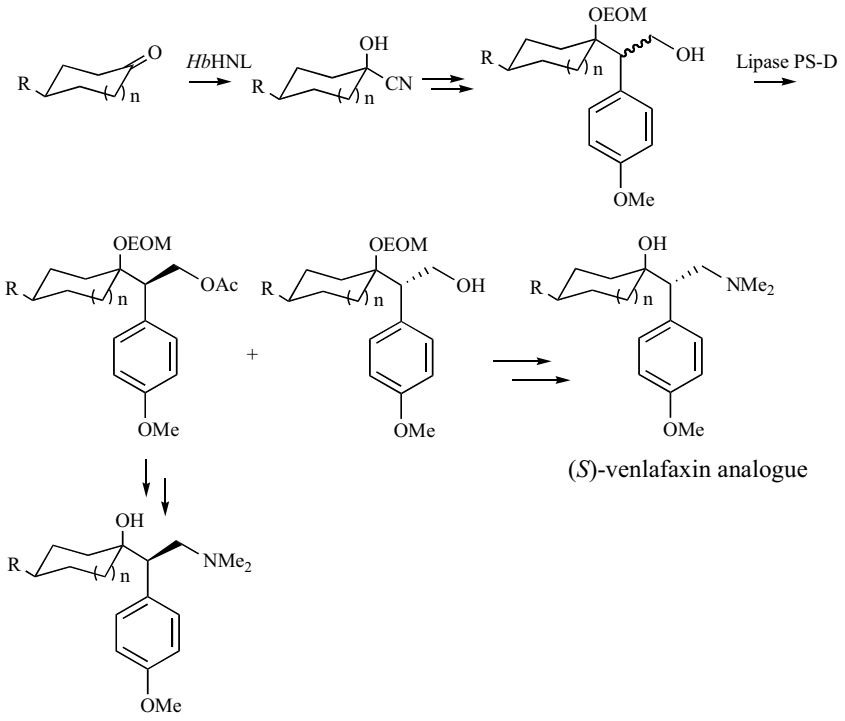
CASCADE REACTIONS

The reaction products of hydroxynitrile lyases, cyanohydrins, can be hydrolyzed by nitrilases to α-hydroxycarboxylic acids and, occasionally, to amide as side product. Thus, the two enzymes can be coupled to a cascade reaction Scheme **8**. Several approaches including engineering of the nitrilase, application of whole cells co-expressing both enzymes and the use of combi-CLEAs in combination with optimization of the reaction conditions to increase the yield of mandelic acid from benzaldehyde have been described [10]. Similarly, a combi-CLEA consisting of a HNL and a nitrile hydratase was used for the one pot cascade synthesis of enantiopure aliphatic α-hydroxycarboxylic amides from the corresponding aldehydes [83]. *E. coli* resting cells co-expressing *Me*HNL and an arylacetonitrilase from *Pseudomonas fluorescens* EBC191 were also used to synthesize (*S*)-atrolactic acid with 92% *ee* and atrolactamide as side product from acetophenone. Interestingly, with this coupled system a much higher conversion of acetophenone was achieved than with *Me*HNL alone [76,84], presumably because the equilibrium was shifted. The formation of amide was reduced to 4% by applying the Tyr54Val nitrilase variant [85]. In the most recent approach, whole cells of the same expression system were subsequently used in a two-phase system with water-immiscible ionic liquids [1-butyl-1-methylpyrrolidinium-bis(trifluoromethanesulfonyl)imide (BMpl NTf2), 1-butyl-3-methylimidazoliumbis(trifluoromethanesulfonyl)imide (BMim NTf2) or 1-butyl-3-methylimidazolium hexafluorophosphate (BMim PF6)] [86]. Depending on the reaction conditions 80-100% of conversion was reached with up to 95% *ee* for (*S*)-mandelic acid and up to 99% *ee* for the (*S*)-mandeloamide. Here, amide formation dominated mandelic acid formation.

A chemoenzymatic approach was recently applied for the preparation of the enantiomers of Venlafaxine hydrochloride (brand name EffexorXR® marketed by Pfizer) and analogs thereof. Both enantiomers are antidepressant agents targeting different receptors. The asymmetric synthesis started with the conversion of cyclohexanone derivatives to their corresponding cyanohydrin by the (*S*)-selective *Hb*HNL followed by several synthetic steps and finally a lipase catalyzed kinetic resolution [87] Scheme **9**.



Scheme (8). (*S*)-selective hydrocyanation in cascade with nitrile hydrolysis using unselective nitrilase.



Scheme (9). Chemoenzymatic synthesis of both enantiomers of Venlafaxine hydrochloride.

REACTION ENGINEERING

A drawback of HNL catalyzed C-C bond formation is the non-enzymatic background reaction, which can significantly compromise the enantiomeric excess (ee) of the reaction. This non-enzymatic reaction needs to be eliminated but at the same time enzyme activity must be conserved.

One way of suppression is to perform the synthesis of cyanohydrins in an acidic environment, in particular pH≤4.5, at temperatures such that the competing chemical reaction and racemization are negligible compared with the enzymatic synthesis. To overcome losses of biocatalyst activity under these conditions the reaction can be performed at low temperature in a range from 5° to 8°C [88]. The currently most widely applied method for biotransformations with HNLs is the use of biphasic systems [73]. Organic substrates are usually soluble in the organic layer; the substrate concentration can be high without inhibiting the enzyme or decreasing its stability, due to the separation from the enzyme in the aqueous layer. Chemical side-reactions are suppressed in the near water-free organic layer and, moreover, the organic solvent can be easily removed and the enzyme in the aqueous phase can be reused. Different aldehydes and ketones can be converted to the respective (*S*)-cyanohydrins in biphasic system consisting of aqueous solution of (*S*)-HNL and a water immiscible or slightly miscible organic phase [89]. A very similar procedure was patented for aromatic (*R*)-benzaldehyde cyanohydrins [46]. Alternatively, the production of the 2-chloro-benzaldehyde and *n*-butyraldehyde cyanohydrins was very efficient in an emulsion system. After 3 hours of reaction time in continuous emulsion 100% of conversion was reached with 89.1% *ee*, compared to the traditional two-phase system which gave 98.7% conversion and 86% *ee* after 4.5 hours [90].

CURRENT AND FUTURE DEVELOPMENTS

First genes for (*S*)- and (*R*)-HNLs were identified and expressed about two decades ago and the discovery of new enzymes is still ongoing. Many international patent applications still cover the production, improved variants, new enzymes from different plant origin and applications of both (*S*)- and (*R*)-HNLs for chemical synthesis. However, besides patent protected major enzyme sources several enzymes such as *Lu*HNL, engineered esterases with HNL activity and new bacterial HNLs are targets of basic research. HNL catalysis offers exciting new opportunities for broad applications in industrial production.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

This work has been supported by the FP7 project KYROBIO, the Austrian BMWFJ, BMVIT, SFG, Standortagentur Tirol and ZIT through the Austrian FFG-COMET-Funding Program.

ABBREVIATIONS

CIP	=	Cahn-Ingold-Prelog
CLEAs	=	Cross-linked enzyme aggregates
ee	=	Enantiomeric excess
FAD	=	Flavin adenine dinucleotide
HCN	=	Hydrocyanic acid
HNL	=	Hydroxynitrile lyase
MTBE	=	Methyl-*tert*-butylether
WT	=	Wild type

REFERENCES

[1] Poulton JE. Cyanogenesis in plants. Plant Physiol 1990; 94: 401-5.
[2] Zagrobelny M, Bak S, Moeller B. Cyanogenesis in plants and arthropods. Phytochemistry 2008; 69: 1457-68.
[3] Andexer J, von Langermann J, Mell A *et al.* An R-selective hydroxynitrile lyase from *Arabidopsis thaliana* with an alpha/beta-hydrolase fold. Angew Chem Int Ed 2007; 46: 8679-81.
[4] Wöhler F, Liebig J. Ueber die Bildung des Bittermandelöls. Ann Pharm 1837; 22: 1-24.
[5] Rosenthaler L. Enzyme effected asymmetrical synthesis. Biochem Zeitschrift 1908; 14: 238-53.
[6] 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 1995; 109: 1231-8.
[7] Gruber-Khadjawi M, Fechter MH, Griengl H. Cleavage and formation of cyanohydrins. In: Drauz K, Gröger H, May O, Eds Enzyme Catalysis in Organic Synthesis. Weinheim: Wiley-VCH 2012: pp. 947-990.
[8] Hilterhaus L, Liese A. Industrial application and processes using carbon–carbon lyases. In: Drauz K, Gröger H, May O, Eds Enzyme Catalysis in Organic Synthesis. Weinheim: Wiley-VCH 2012: pp. 991-1000.
[9] Purkarthofer T, Skranc W, Schuster C, Griengl H. Potential and capabilities of hydroxynitrile lyases as biocatalysts in the chemical industry. Appl Microbiol Biotechnol 2007; 76: 309-20.
[10] Winkler M, Glieder A, Steiner K. Hydroxynitrile lyases: from nature to application. In: Carreira EM, Yamamoto H, Eds. Comprehensive Chirality, Synthetic Methods VI - Enzymatic and Semi-Enzymatic. Amsterdam: Elsevier 2012: Vol. 7, pp. 350-71.
[11] Sheldon RA. Enzyme immobilization: the quest for optimum performance. Adv Synth Catal 2007; 349: 1289-307.
[12] Sheldon RA. Cross-linked enzyme aggregates as industrial biocatalysts. Org Proc Res Devel 2011; 15: 213-23.
[13] Sheldon RA. Characteristic features and biotechnological applications of cross-linked enzyme aggregates (CLEAs). Appl Microbiol Biotechnol 2011; 92: 467-77.
[14] Sheldon RA. Industrial Applications of Asymmetric Synthesis using Cross-Linked Enzyme Aggregates. In: Carreira EM, Yamamoto H, Eds. Comprehensive Chirality, Synthetic Methods VI - Enzymatic and Semi-Enzymatic. Amsterdam: Elsevier 2012: Vol. 9, pp. 353-66.
[15] Effenberger F, Förster S, Wajant H. Hydroxynitrile lyases in stereoselective catalysis. Curr Opin Biotechnol 2000; 11: 532-9.
[16] Sharma M, Sharma NN, Bhalla TC. Hydroxynitrile lyases: at the interface of biology and chemistry. Enz Microbial Technol 2005; 37: 279-94.
[17] Lu W-Y, Lin G-Q. Chiral synthesis of pharmaceutical intermediates using oxynitrilases. In: Tao J, Lin G-Q, Liese A, Eds. Biocatalysis for the Pharmaceutical Industry, Wiley VCH 2009: pp. 89-109.
[18] Andexer JN, Langermann J, Kragl U, Pohl M. How to overcome limitations in biotechnological processes - examples from hydroxynitrile lyase applications. Trends Biotechnol 2009; 27: 599-607.
[19] Holt J, Hanefeld U. Enantioselective enzyme-catalysed synthesis of cyanohydrins. Curr Org Synth 2009; 6: 15-37.
[20] Bornscheuer UT, Kazlauskas RJ. Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways. Angew Chem Int Ed 2004; 43: 6032-40.

[21] North M. Synthesis and applications of non-racemic cyanohydrins. Tetrahedron-Asymm 2003; 14: 147-76.
[22] Sukumaran J, Hanefeld U. Enantioselective C-C bond synthesis catalysed by enzymes. Chem Soc Rev 2005; 34: 530-42.
[23] Dadashipour M, Asano Y. Hydroxynitrile lyases: insights into biochemistry, discovery, and engineering. ACS Catalysis 2011; 1: 1121-49.
[24] Eggert T, Andexer JN. (*R*)-hydroxynitrile lyase from Brassicaceae. US20100112648, 2010.
[25] Asano Y. (*R*)-hydroxynitrile lyase and method for using the same. JP4746084, 2009.
[26] Gaisberger R, Glieder A, Liu ZB, Pscheidt B. R-HNL random variants and their use for preparing optically pure, sterically hindered cyanohydrins. US 20100143986, 2010.
[27] Hasslacher M, Schall M, Schwab H, Hayn E, Kohlwein S, Griengl H. (*S*)-hydroxynitrilelyase from *Hevea brasiliensis*. WO9703214, 1997.
[28] Ichige E, Semba H, Shijuku T, Harayama S. Novel modified S-hydroxynitrile lyase. US20080124784, 2008.
[29] Senba T, Ichige E. New highly-active modified-type S-hydroxynitrile lyase. JP4494286, 2006.
[30] Effenberger F, Förster S, Wajant H. Method for the production of hydroxynitrile lyases. EP1246909, 2006.
[31] Semba H, Ichige E, Mukouyama M. Method for production of S-hydroxynitrile lyase by use of *Escherichia coli*. US7214523, 2007.
[32] Semba J, Tsuchiura-shi I. Process for producing S-hydroxynitrile lyase. EP1016712, 2005.
[33] Schwab H, Glieder A, Kratky C and *et al*. Genes containing a DNA sequence coding for hydroxynitrile lyase, recombinant proteins derived therefrom and having hydroxynitrile lyase activity, and use thereof. US6861243, 2005.
[34] Asano Y. Synthetase for optically active cyanohydrin derived from *Prunus mume* fruit and method for production of optically active substance using the same. JP2006238763, 2006.
[35] Asano Y, Tamura K, Doi N and *et al.* Screening for new hydroxynitrilases from plants. Biosci Biotechnol Biochem 2005; 69: 2349-57.
[36] Hernandez L, Luna H, Ruiz-Teran F, Vazquez A. Screening for hydroxynitrile lyase activity in crude preparations of some edible plants. J Mol Catal B-Enzym 2004; 30: 105-8.
[37] Hickel A, Heinrich G, Schwab H, Griengl H. Screening for hydroxynitrile lyases in plants. Biotechnol Techniques 1997; 11: 55-8.
[38] Senba T, Dobashi Y. (*R*)-hydroxynitrile lyase composition and stabilization method. JP2004113053, 2004.
[39] Trummler K, Wajant H. Molecular cloning of acetone cyanohydrin lyase from flax (*Linum usitatissimum*) - definition of a novel class of hydroxynitrile lyases. J Biol Chem 1997; 272: 4770-4.
[40] Xu LL, Singh BK, Conn EE. Purification and characterization of acetone cyanohydrin lyase from *Linum usitatissimum*. Arch Biochem Biophys 1988; 263: 256-63.
[41] Cheng IP, Poulton JE. Cloning of cDNA of *Prunus-serotina* (*R*)-(+)-mandelonitrile lyase and identification of a putative FAD-binding site. Plant Cell Physiol 1993; 34: 1139-43.
[42] Effenberger F, Ziegler T, Förster S. Verfahren zur Herstellung von optisch aktiven Cyanhydrinen. EP0276375, 1992.
[43] Zhao GJ, Yang ZQ, Guo YH. Cloning and expression of hydroxynitrile lyase gene from *Eriobotrya japonica* in *Pichia pastoris*. J Biosci Bioeng 2011; 112: 321-5.
[44] Ueatrongchit T, Komeda H, Asano Y. Parameters influencing asymmetric synthesis of (*R*)-mandelonitrile by a novel (*R*)-hydroxynitrile lyase from *Eriobotrya japonica*. J Mol Catal B-Enzym 2009; 56: 208-14.
[45] Breuer M, Ditrich K, Habicher T and *et al.* Industrial methods for the production of optically active intermediates. Angew Chem Int Ed 2004; 43: 788-824.
[46] Kirschbaum B, Wilbert G, Effenberger F. Process for preparing optically active cyanohydrins and secondary products. US6653498, 2003.
[47] Weis R, Poechlauer P, Bona R and *et al.* Biocatalytic conversion of unnatural substrates by recombinant almond R-HNL isoenzyme 5. J Mol Catal B-Enzym 2004; 29: 211-8.
[48] Dobashi Y, Senba T. Chimeric recombinant (*R*)-hydroxynitrile lyase. JP2009106167, 2009.
[49] Skranc W, Glieder A, Gruber K, Weis R, Luiten R. R-hydroxynitrile lyases having improved substrate tolerance and the use thereof. US7572608, 2009.
[50] Ogura K, Sato E, Mori H, Nogami H, Kato O, Saito T, Kokubu N, Ninomiya Y. Method for producing optically active cyanohydrin and method for producing optically active alpha-hydroxy carboxylic acid. JP2012105671, 2012.
[51] Semba H, Dobashi Y. An enzyme reaction method and a method for enzymatically producing an optically active cyanohydrin. EP1160329, 2011.
[52] van Langen LM, Selassa RP, van Rantwijk F, Sheldon RA. Cross-linked aggregates of (*R*)-oxynitrilase: a stable, recyclable biocatalyst for enantioselective hydrocyanation. Org Lett 2005; 7: 327-9.
[53] Roberge C, Fleitz F, Pollard D, Devine P. Synthesis of optically active cyanohydrins from aromatic ketones: evidence of an increased substrate range and inverted stereoselectivity for the hydroxynitrile lyase from *Linum usitatissimum*. Tetrahedron Asymm 2007; 18: 208-14.
[54] Roberge C, Fleitz F, Devine P. (*R*)- and (*S*)-cyanohydrin formation from pyridine-3-carboxaldehyde using CLEA-immobilized hydroxynitrile lyases. In: Whittall J, Sutton PJ, Eds. Practical Methods for Biocatalysis and Biotransformations. Wiley: 2010; pp. 114-8.
[55] Saito T, Kubo E, Mori H. Preparation of optically-active amino alcohols and their hydrochlorides by reduction of pyridinecarbaldehyde cyanohydrins. JP2012217375, 2012.
[56] Trummler K, Roos J, Schwaneberg U and *et al.* Expression of the Zn^{2+}-containing hydroxynitrile lyase from flax (*Linum usitatissimum*) in *Pichia pastoris*-utilization of the recombinant enzyme for enzymatic analysis and site-directed mutagenesis. Plant Sci 1998; 139: 19-27.
[57] Pscheidt B, Liu ZB, Gaisberger R and *et al.* Efficient biocatalytic synthesis of (*R*)-pantolactone. Adv Synth Catal 2008; 350: 1943-8.
[58] Das T, Mahapatra T, Nanda S. Total synthesis of stagonolide B. Tetrahedron Lett 2012; 53: 1186-1189.
[59] Purkarthofer T, Skranc W, Weber H and *et al.* One-pot chemoenzymatic synthesis of protected cyanohydrins. Tetrahedron 2004; 60: 735-9.
[60] Skranc W, Poechlauer P, Wirth I, Neuhofer R, Mayrhofer H. Process for preparing protected enantiomer-enriched cyanohydrins by in-situ derivatization. US6909011, 2005.
[61] Guterl JK, Andexer JN, Sehl T andfk, *et al.* Uneven twins: comparison of two enantiocomplementary hydroxynitrile lyases with alpha/beta-hydrolase fold. J Biotechnol 2009; 141: 166-73.
[62] Andexer JN, Staunig N, Eggert T and *et al.* Hydroxynitrile lyases with alpha/beta-hydrolase fold: two enzymes with almost identical 3D structures but opposite enantioselectivities and different reaction mechanisms. Chem Bio Chem 2012; 13: 1932-9.
[63] Okrob D, Metzner J, Wiechert W, Gruber K, Pohl M. Tailoring a stabilized variant of hydroxynitrile lyase from *Arabidopsis thaliana*. Chem Bio Chem 2012; 13: 797-802.
[64] Okrob E, Paravidino M, Orru RVA and *et al.* Hydroxynitrile lyase from *Arabidopsis thaliana*: identification of reaction parameters for enantiopure cyanohydrin synthesis by pure and immobilized catalyst. Adv Synth Catal 2011; 353: 2399-408.
[65] Scholz KE, Okrob D, Kopka B and *et al.* Synthesis of chiral cyanohydrins by recombinant *Escherichia coli* cells in a micro-aqueous reaction system. Appl Environ Microbiol 2012; 78: 5025-7.
[66] Fuhshuku K, Asano Y. Synthesis of (*R*)-beta-nitro alcohols catalyzed by R-selective hydroxynitrile lyase from *Arabidopsis thaliana* in the aqueous-organic biphasic system. J Biotechnol 2011; 153: 153-9.
[67] Selmar D, Lieberei R, Biehl B, Conn EE. Alpha-hydroxynitrile lyase in *Hevea brasiliensis* and its significance for rapid cyanogenesis. Physiol Plantarum 1989; 75: 97-101.
[68] Klempier N, Griengl H, Hayn M. Aliphatic (*S*)-cyanohydrins by enzyme-catalyzed synthesis. Tetrahedron Lett 1993; 34: 4769-72.
[69] Hughes J, Decarvalho JPC, Hughes MA. Purification, characterization, and cloning of alpha-hydroxynitrile lyase from cassava (*Manihot esculenta* Crantz). Arch Biochem Biophys 1994; 311: 496-502.
[70] Kirchner G, Wirth I, Werenka C, Griengl H, Schmidt M. Enzymatic processes for preparing (*S*)-cyanohydrins. US6337196, 2002.
[71] Effenberger F, Wajant H, Förster S, Roos J. Process for the preparation of (*S*)-cyanhydrines. EP0799894, 2004.
[72] Semba H, Dobashi Y. Immobilised enzyme and method for synthesizing an optical active cyanohydrin. EP1116789, 2004.
[73] Avi M, Griengl H. Biocatalysis in biphasic systems: Oxynitrilases. In: Carrea G, Riva S, Eds. Organic Synthesis with Enzymes in Non-Aqueous Media. Weinheim: Wiley-VCH 2008: pp. 211-26.

[74] Semba H, Ichige E, Imanaka T, Atomi H, Aoyagi H. Efficient production of active form of recombinant cassava hydroxynitrile lyase using *Escherichia coli* in low-temperature culture. Methods Mol Biol 2010; 643: 133-44.

[75] Avi M, Wiedner R, Griengl H, Schwab H. Improvement of a stereoselective biocatalytic synthesis by substrate and enzyme engineering: 2-hydroxy-(4 '-oxocyclohexyl)acetonitrile as the model. Chem Eur J 2008; 14: 11415-22.

[76] Bühler H, Effenberger F, Förster S, Roos J, Wajant H. Substrate specificity of mutants of the hydroxynitrile lyase from *Manihot esculenta*. ChemBioChem 2003; 4: 211-6.

[77] Lauble H, Miehlich B, Förster S and *et al.* Structure determinants of substrate specificity of hydroxynitrile lyase from *Manihot esculenta*. Protein Sci 2002; 11: 65-71.

[78] Asano Y, Akiyama T, Yu F, Sato E. Substitutional variants of hydroxynitrile lyase with high specific activity and methods of use. US8030053, 2011.

[79] Dadashipour M, Fukuta Y, Asano Y. Comparative expression of wild-type and highly soluble mutant His103Leu of hydroxynitrile lyase from *Manihot esculenta* in prokaryotic and eukaryotic expression systems. Protein Expr Purif 2011; 77: 92-7.

[80] Asano Y, Dadashipour M, Yamazaki M, Doi N, Komeda H. Functional expression of a plant hydroxynitrile lyase in *Escherichia coli* by directed evolution: creation and characterization of highly *in vivo* soluble mutants. Prot Eng Des Sel 2011; 24: 607-16.

[81] Padhi SK, Fujii R, Legatt G and *et al.* Switching from an esterase to a hydroxynitrile lyase mechanism requires only two amino acid substitutions. Chem Biol 2010; 17: 863-71.

[82] Feichtenhofer S, Höffken WW, Schwab H. Conversion of esterase EstC from *Burkholderia gladioli* into a hydroxynitrile lyase. 2009; Poster at Enzyme Engineering XX, Groningen, The Netherlands.

[83] van Pelt S, van Rantwijk F, Sheldon RA. Synthesis of aliphatic (*S*)-alpha-hydroxycarboxylic amides using a one-pot bienzymatic cascade of immobilised oxynitrilase and nitrile hydratase. Adv Synth Catal 2009; 351: 397-404.

[84] von Langermann J, Mell A, Paetzold E, Daußmann T, Kragl U. A hydroxynitrile lyase in organic solvent-free systems to overcome thermodynamic limitations. Adv Synth Catal 2007; 349: 1418-24.

[85] Baum S, Williamson DS, Sewell T, Stolz A. Conversion of sterically demanding alpha,alpha-disubstituted phenylacetonitriles by the arylacetonitrilase from *Pseudomonas fluorescens* EBC191. Appl Environ Microbiol 2012; 78: 48-57.

[86] Baum S, van Rantwijk F, Stolz A. Application of a recombinant *Escherichia coli* whole-cell catalyst expressing hydroxynitrile lyase and nitrilase activities in ionic liquids for the production of (*S*)-mandelic acid and (*S*)-mandeloamide. Adv Synth Catal 2012; 354: 113-22.

[87] Bhuniya R, Nanda S. Asymmetric synthesis of both the enantiomers of antidepressant venlafaxine and its analogues.Tetrahedron Lett 2012; 53: 1990-2.

[88] Niedermayer U, Kragl U, Kula M-R, Wandrey C, Makryaleas K, Drauz K. Enzymatisches Verfahren zur Herstellung von optisch aktiven Cyanhydrinen. EP0326063, 1994.

[89] Poechlauer P, Schmidt M, Neuhofer R and *et al*. Enzymatic process for the preparation of (*S*)-cyanohydrins. US6225095, 2001.

[90] Pöchlauer P, Wirth I, Mayrhofer H, Neuhofer R. Method for the production of optically active cyanohydrins using R-oxynitilase. US6717006, 2004

Chapter 2

Bioprospecting for Hydroxynitrile Lyases by Blue Native PAGE and HCN Detection

Elisa Lanfranchi[1], Eva-Maria Köhler[1], Barbara Darnhofer[1,2,3], Kerstin Steiner[1], Ruth Birner-Gruenbergher[1,2,3], Anton Glieder[1,4], and Margit Winkler[1]*

Published in: Current Biotechnology, 4 (2015) 2, 111-117

[1]Austrian Centre Industrial Biotechnology, Petersgasse 14 8010 Graz, Austria
[2]Medical University of Graz, Institute of Pathology, Stiftingtalstrasse 24, 8010 Graz, Austria
[3]Omics Center Graz, BioTechMed Graz, Stiftingtalstraße 24, 8010 Graz, Austria

Corresponding Author:
Margit Winkler
Austrian Centre Industrial Biotechnology, Petersgasse 14/III, 8010 Graz, Austria
Email: margit.winkler@acib.at; Tel: +43-316-8739333; Fax +43-316-87393

Bioprospecting for Hydroxynitrile Lyases by Blue Native PAGE Coupled HCN Detection

Elisa Lanfranchi[1], Eva-Maria Köhler[1], Barbara Darnhofer[1,2,3], Kerstin Steiner[1], Ruth Birner-Gruenberger[1,2,3], Anton Glieder[1,4] and Margit Winkler[*,1]

[1]*ACIB GmbH, Graz, Austria;* [2]*Institute for Pathology, Medical University of Graz, Graz, Austria;* [3]*Omics Center Graz, BioTechMed, Graz, Austria;* [4]*Institute of Molecular Biotechnology, Graz University of Technology, NAWI Graz, Graz, Austria*

Elisa Lanfranchi

Margit Winkler

Abstract: Hydroxynitrile lyase enzymes (HNLs) catalyze the stereoselective addition of HCN to carbonyl compounds to give valuable chiral hydroxynitriles. The discovery of new sources of HNL activity has been reported several times as the result of extensive screening of diverse plants for cyanogenic activity. Herein we report a two step-method that allows estimation of not only the native size of the active HNL enzyme but also its substrate specificity. Specifically, crude protein extracts from plant tissue are first subjected to blue native-PAGE. The resulting gel is then directly used for an activity assay in which the formation of hydrocyanic acid (HCN) is detected upon the cyanogenesis reaction of any cyanohydrin catalyzed by the enzyme of interest. The same gel may be used with different substrates, thus exploring the enzyme's substrate scope already on the screening level. In combination with mass spectrometry, sequence information can be retrieved, which is demonstrated with the example of the so far unknown sequence of *Prunus domestica* HNL.

Keywords: Biocatalysis, blue native polyacrylamide gel electrophoresis, cyanogenesis, cyanohydrin, enzyme, hydroxynitrile lyase, oxynitrilase, screening.

INTRODUCTION

Several plants are known to utilize enzyme mediated cyanogenesis as a defense mechanism against herbivores and the proteins catalyzing this reaction are hydroxynitrile lyases (HNLs; alternative names: oxynitrilase, hydroxynitrilase, mandelonitrile lyase, hydroxymandelonitrile lyase) [1]. For industrial application, the reverse reaction is a highly interesting transformation that can be used for the stereoselective synthesis of cyanohydrins [2-5]. To date, HNL activity has been reported in bacteria [6] and many plant species [4,7,8], but only a relatively small number of HNLs are currently used on an industrial scale [2,5]. More often than not, the gene and protein sequences are yet unknown [9-14]. Without this information, heterologous expression - in order to access sufficient enzyme amounts - is not possible, and also adaptation of the enzymes to process requirements by protein engineering is not an option. Although a number of assays to identify new sources of HNL activity [15-19] or improved HNL variants [20,21] have been reported, there was no method that simultaneously shows HNL activity, the native size of the enzyme, its substrate specificity and moreover allows one to determine its primary sequence.

*Address correspondence to this author at the Austrian Centre of Industrial Biotechnology, Petersgasse 14, Graz, Austria; Tel: ++43 316 873 9333; Fax: ++43 316 873 9308; E-mail: margit.winkler@acib.at

In order to retain enzymatic activity for an in-gel assay, the separation of catalytically active enzyme from other proteins in complex samples by polyacrylamide gel electrophoresis (PAGE) must be carried out under non-denaturing conditions. Standard native PAGE is performed in the absence of SDS and a suitable buffer pH is necessary in order to keep the protein of interest negatively charged. Therefore, the knowledge of the isoelectric point of the protein is crucial [22]. However, this information is often not available in the course of enzyme discovery, since the protein sequences are unknown. Moreover, a crude protein extract is composed of different proteins, each with their own isoelectric point. Consequently, establishing an appropriate protocol can be time consuming and is distinct for each sample. Blue Native PAGE (BN-PAGE) allows the separation of the proteins based on their molecular weight, maintaining the proteins' native condition and oligomerization state. All proteins acquire negative charge due to the Coomassie G-250 reagent [23]. Therefore, the method can be applied to any unknown enzyme, and it is not only suitable for fairly pure enzyme preparations but also for the separation of crude protein extracts.

Herein we report a two step-method in which crude protein extracts e.g. from plant tissue is first subjected to BN-PAGE. The resulting gel is then directly used for an activity assay. Specifically, the formation of hydrocyanic acid (HCN) is detected upon the cyanogenesis reaction of any cyanohydrin catalyzed by the enzyme of interest. By combination of this method with mass spectrometry we

2211-551 /15 $58.00+.00

retrieved partial sequence information of the *Prunus domestica* hydroxynitrile lyase enzyme (*Pd*HNL).

MATERIALS AND METHODS

General

Different tissues of *Prunus domestica* and *Prunus laurus* were harvested from private gardens (Styria, Austria). *Malus domestica* seeds were obtained from apple fruits bought in a local shop. *Passiflora capparidifolia* leaves were kindly donated by the botanic garden of Graz. *Arabidopsis thaliana* HNL (*At*HNL) (lyophilized powder) was a kind gift from EVOCATAL GmbH. *Prunus amygdalus* HNL (*Pa*HNL) and acetone cyanohydrin were purchased from Sigma-Aldrich. Racemic mandelonitrile was purchased from abcr GmbH & Co. KG. *Hevea brasisliensis* HNL (*Hb*HNL) was obtained as lyophilized powder from DSM Fine Chemicals Linz. For gel electrophoresis, an XCell SureLock® Mini-Cell equipped with a PowerEase® 500 Programmable Power Supply (Life Technologies) was used.

Sample Preparation from Plant Tissues

Leaves, branches, sprouts and flowers were shock-frozen in liquid nitrogen and stored at -20°C. Seeds were stored at room temperature. Protein extracts were prepared from 80-100 mg of the different tissues with the P-PER® Plant Protein extraction kit (Thermo scientific). PD10 desalting columns (GE Healthcare) were used to exchange the buffer for sodium phosphate (25 mM, pH 6.6). The samples were concentrated by ultrafiltration (Vivaspin 20, 10.000 MWCO, Sartorius) and stored at -20°C.

Cloning of *Linum usitatissimum* HNL (*Lu*HNL) and Expression in *Pichia pastoris*

The *Lu*HNL sequence (gene accession number: AF024588) was codon optimized in order to increase its translation efficiency in *Pichia pastoris* (Graphical codon usage analyser http://gcua.schoedl.de), moreover, to remove an XhoI restriction site within the sequence. The synthetic gene (*Lu*HNL_opt), carrying EcoRI-NotI sites, was purchased in the pUC57 vector (GenScript). The *Lu*HNL_opt sequence was amplified by PCR using primers LuHNL-fw 5'-aat gcg aat tcg cca cca tgg ctt ctc ttc ctg tta gct ttg-3' and LuHNL_rev 5'-aat gcg cgg ccg ctc aat aat cgt tca act tga tc-3', and then it was cloned into the vector pPpB1. The correctness of the construct was confirmed by sequencing (LGC genomics). The plasmid was linearized with BglII, and 6 µg of DNA were transformed by electroporation (2.0 kV) into *Pichia pastoris* CBS 7435 MutS competent cells [24]. Clones were selected on YPD-medium agar plates (1% w/v yeast extract, 2% w/v peptone and 2% w/v glucose, 300 µg/ml zeocin) by incubation at 28°C for 3 days. Activity screening was performed in high through-put format as described previously [25]. The cell pellets were disrupted with the Y-PER™ Plus Dialyzable Yeast Protein Extraction Reagent yeast protein reaction reagent (Pierce). The protocol described by Krammer [21] was modified for *Lu*HNL activity detection. The hydroxynitrile lyase activity was assayed using 8 µL of protein extract in 72 µL of citrate buffer (300 mM, pH 4.0) and 20 µL of acetone cyanohydrin (1% v/v). The most active clone and the parental strain *Pichia pastoris* CBS 7435 MutS were cultivated in shake flasks as described previously [25]. The cells were harvested by centrifugation, suspended in potassium phosphate buffer pH 6.0 (50 mM) and disrupted using a French press. Cell free extracts were obtained by centrifugation at 20,000 rpm and 4°C for 40 min and stored at -20°C.

Protein Quantification

Protein concentrations were quantified using the BCA protein assay kit (Thermo scientific).

Blue Native PAGE

NativePAGE™ Novex®4-16% Bis-Tris Gels (Life Technologies) were used for blue native PAGE. The preparation of the samples, anode and cathode buffer was carried out as described in the user manual. NativeMark™ Unstained Protein Standard (Life Technologies) was used as protein ladder. The electrophoresis was performed at 4°C and 150 V for 60 min and then 250 V for 90 min. Subsequently, the gel was washed twice with deionized water and used for the in-gel activity assay.

HNL Activity Determination

Detection paper was prepared according to the Krammer protocol [21]: 1% (w/v) solution of 4,4'-methylenebis (*N*,*N*-dimethyl-aniline) in chloroform was added slowly under stirring to a 1% (w/v) solution of copper (II) ethylacetoacetate in chloroform. Whatman® filter paper No1 (GE Healthcare) was soaked with the resulting dark green, clear solution and dried in a fume hood. The greenish detection paper may be stored in a cool, dry and dark place. The native gel was incubated in citrate buffer (100 mM, pH 4.5) at 4°C for 30 min. The HNL activity assay must be performed in a well ventilated hood. Whatman® filter paper No1 (GE Healthcare) was soaked in citrate buffer (moist but not wet) and placed in a square polypropylene vessel and covered with the equilibrated gel. Another Whatman® filter paper moist with cyanohydrin substrate solution (e.g. 30 mM cyanohydrin in 100 mM citrate buffer pH 4.5) was placed on top of the gel, followed by a mosquito net and detection paper. The assembly was fixed with a transparent layer on the top (e.g. a petri dish) and a transparent weight (e.g. a small glass bottle) in order to follow the development of blue spots visually (Fig. **1**). The entire procedure was performed at room temperature. Typically, the exposure time is within a few minutes, depending on the amount of applied HNL and its activity towards cleavage of the substrate. The development of the spots is dynamic, starting with a distinct small spot at the center of activity which can extend to big blurred spots upon prolonged exposure. After this treatment, the same gel can be thoroughly washed, and assembled again with a new paper soaked in another substrate and a fresh HCN detection paper.

Staining

Subsequently to the visualization of the HNL activity, the gel may be stained with Coomassie Blue according to the Invitrogen User Manual or other methods such as silver staining [26]. Note that SimplyBlue™ SafeStain (Life

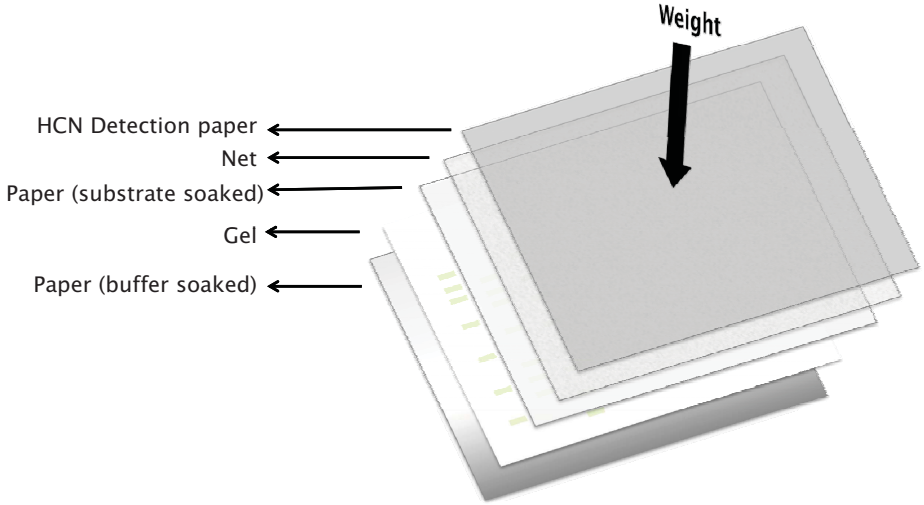


Fig. (1). Schematic representation of the assembly for the HNL activity coupled blue native gel assay.

technologies) is not compatible with NativePAGE™ Novex®4-16% Bis-Tris Gels.

Mass Spectrometry

LC-MS/MS was used for protein identification. Excised protein bands were digested with modified trypsin (Promega) [27]. Peptide extracts were dissolved in 0.1% formic acid, 5% acetonitrile and separated by nano-HPLC (Dionex Ultimate 3000) equipped with a µ-precolumn (C18, 5 µm, 100 Å, 5 x 0.3 mm) and an Acclaim PepMap RSLC nanocolumn (C18, 2 µm, 100 Å, 500 x 0.075 mm) (Thermo Fisher Scientific). Samples were concentrated on the enrichment column for 2 min at a flow rate of 20 µL min^{-1} with 0.5% trifluoroacetic acid isocratically. The separation was carried out on the nanocolumn at a flow rate of 200 nL min^{-1} using the following gradient, where solvent A is 0.05% trifluoroacetic acid in water and solvent B is a mixture of 80% acetonitrile in water containing 0.05 % trifluoroacetic acid: 0-2 min 4% B, 2-35 min 4-28% B, 35-47 min 28-50% B, 47-48 min 50-95% B, 48-58 min 95% B, 58-58.1 min 95-4% B, and 58.1-70 min 4% B. The sample was ionized in the nanospray source equipped with stainless steel emitters (ES528, Thermo Fisher Scientific) and analyzed in a Thermo LTQ-FT mass spectrometer in positive ion mode by alternating full scan MS (m/z 200 to 2000) in the ICR cell and MS/MS by CID of the 5 most intense peaks in the ion trap. Proteome Discoverer 1.4 (Thermo Electron) and Mascot 2.4 (Matrix Science) were used for MS/MS data analysis by searching the NCBInr database (downloaded on 02.02.2013, 7847231743 residues, 22826945 sequences) and the transcriptome of *Prunus armeniaca* (NCBI Bioproject PRJNA237575, downloaded on 06.05.2014, 78537253 residues, 98283 sequences). Detailed search criteria were: enzyme: trypsin; maximum missed cleavage sites: 2; *N*-terminus: hydrogen; *C*-terminus: free acid; fixed Cys modification: carbamidomethylation; search mode: homology search; possible multiple oxidized methionine, maximum precursor charge: 3; precursor mass tolerance: +/- 0.05 Da; product mass tolerance: +/- 0.7 Da; and 5% false discovery rate. We filtered them according to stringent peptide acceptance criteria, including mass deviations of ±10 ppm, Mascot Ion Score of at least 17 and a position rank 1 in Mascot search. Nucleotide sequences were searched for homologues with BlastX (National Center for Biotechnology Information, USA).

RESULTS & DISCUSSION

Assay Establishment

Hydroxynitrile lyase activity has been observed in several members of the *Rosaceae* family and we chose *P. domestica* [28] kernels for assay establishment. Protein extracts of the kernels were prepared as described in Materials and Methods and applied to the BN-PAGE in four different dilutions. The commercial preparation of *Prunus amygdalus* HNL was used as a positive control, flanked by empty lanes. Bovine serum albumin served as a negative control for the HNL detection. Subsequently to the electrophoresis, the in-gel assay was assembled as shown in Fig. (**1**). The use of the mosquito net or any other gas permeable inert spacer is necessary to separate the detection paper from the wet gel, as wetting of the detection paper will result in blurred spots. It is recommended to use a transparent lid and weight in order to follow the development of the spots visually. The time of incubation in this assembly depends on the level of HNL activity: the higher the activity, the lower the incubation time. Fig. (**2**) shows the BN-PAGE gel that was subsequently stained with Coomassie Blue (Fig. **2**, **A1**) and the corresponding detection filter on which blue spots are visible at the position of functional HNL enzymes (Fig. **2**, **A2**). An overlay of the respective zones (Fig. **2**, **A3**) shows the activity of the positive control on lane 2 and also on lanes 3-5 in decreasing intensity, which reflects the dilution of the protein. About hundred µg of *P. domestica* kernel protein extract (lane 5) are a sufficient amount for enzyme activity detection, although no distinct protein band is visible at the area of activity even at higher concentrations. This reflects the high sensitivity of the HCN detection.

A repetition of the activity assay confirmed HNL activity after overnight storage of the gel in citrate buffer (100 mM pH 4.5) at 4°C (data not shown).

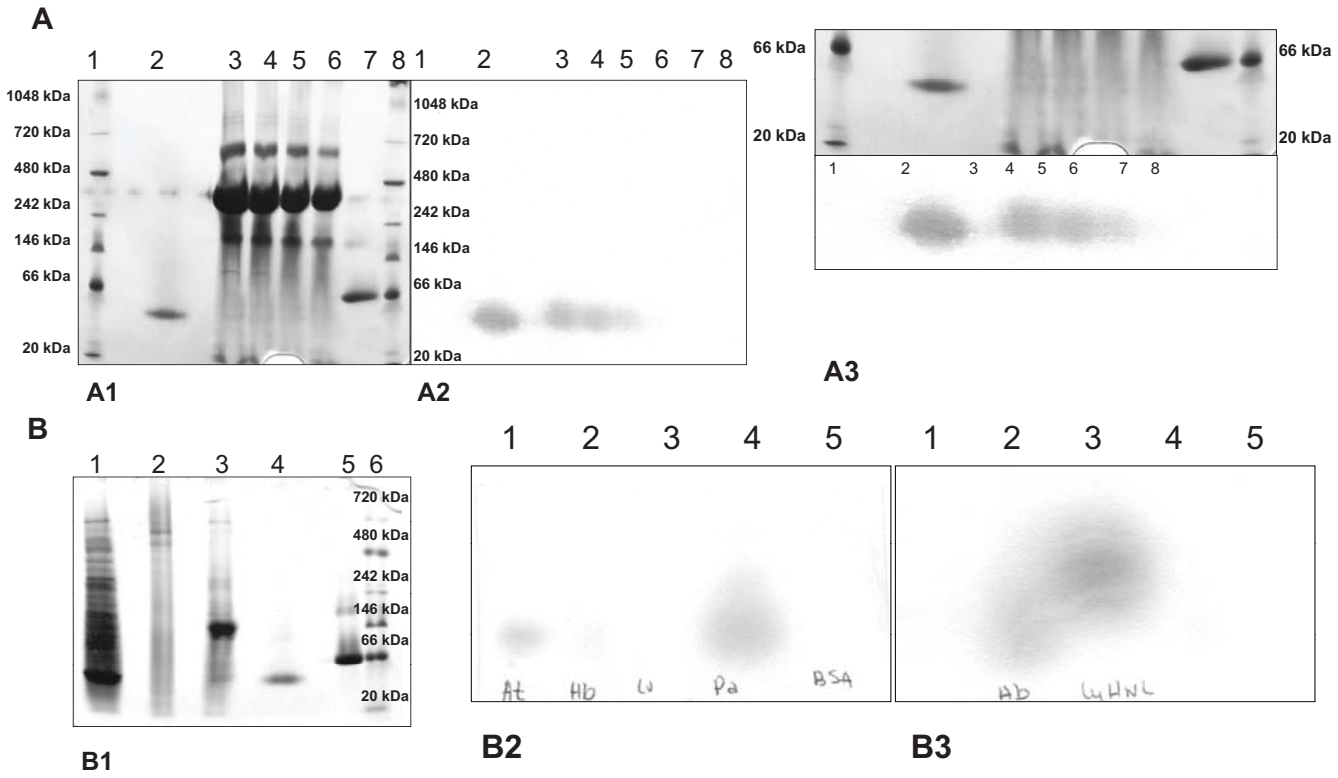


Fig. (2). (A) Blue Native PAGE and HNL activity visualization on *P. domestica* kernels extract: Blue Native PAGE (A1): NativeMark™ Unstained Protein Standard (1;8), positive control *Pa*HNL 7 µg (2), raw protein extract 7.4 mg/mL from *P. domestica* kernels 30 µL, 20 µL, 15 µL and 10 µL (3-6), negative control BSA 5 µg (7). Activity assay (**A2**): rac. mandelonitrile (30 mM in 100 mM citrate phosphate buffer, pH 4.5). Exposure time: 4 min. (**A3**): magnification of BN-PAGE and overlay with HNL activity detection. (**B**) Blue Native PAGE and HNL activity visualization on different known HNLs with two different substrates: Blue Native PAGE (**B1**): *At*HNL 14 µg (1), *Hb*HNL 7 µg (2), *Lu*HNL cell free extract 20 µg (3), *Pa*HNL 7 µg (4), negative control BSA 8 µg (5), NativeMark™ Unstained Protein Standard (6). Activity assay: rac. mandelonitrile (**B2**) and acetone cyanohydrin (**B3**) (30 mM in 100 mM citrate phosphate buffer, pH 4.5). Exposure time 8 min.

Method Versatility

The assay described herein can be applied on protein extracts obtained from different plant species and tissues (Fig. **3**). Needless to say, the assay can also be applied to less complex protein mixtures such as commercial enzyme preparations Fig. **2** (section **B**), heterologously overexpressed hydroxynitrile lyases Fig. **2** (section **B**) partially purified and pure protein samples from different sources. Furthermore, it is suitable for any kind of hydroxynitrile lyase and different hydroxynitrile substrates, which can be applied one after the other on the same gel (Fig. **2**, section **B**).

Specific activity of *At*HNL is low below pH 5 and it has a short half-life [29]. Therefore, the gel was incubated and kept in 25 mM sodium phosphate buffer, pH 6.6.

Different HNLs and Substrates

As an example, we applied the well studied HNL enzymes *Pa*HNL, *Hb*HNL, *At*HNL and *Lu*HNL to this assay. They belong to several protein families and show different substrate specificities. *Pa*HNL, *Hb*HNL and *At*HNL accept a broad range of substrates, either aromatic or aliphatic. However, *Pa*HNL and *At*HNL are poorly active with acetone cyanohydrin [30], which is the natural substrate of *Hb*HNL and *Lu*HNL [31,32]. Activity of *Lu*HNL for the cleavage of mandelonitrile has not been reported. The blue native PA gel (Fig. **2**, **B1**) was consecutively used with mandelonitrile (Fig. **2**, **B2**) and acetone cyanohydrin (Fig. **2**, **B3**) soaked paper. Fig. **2** (section **B**) confirms the substrate preference of these well known HNL enzymes.

Different Tissues and Plants

The assay described herein was applied on different tissues of the same plant such as young and ripe leaves, kernels, flowers, branches and sprouts and several plant species were tested (Figs. **2**, **3**). Protein extract samples from plants were screened for HNL activity with a rapid filter paper assay (data not shown) [21] and subsequently, they were applied to the BN-PAGE. Fig. **3** (Section **A**) reports the results of different *P. domestica* tissues. Leaves, flowers and sprouts showed HNL activity, whereas proteins isolated from branches did not display hydroxynitrile lyase activity (Fig. **3**, **A2**). This result is consistent with the HNL assay performed on the raw protein extract.

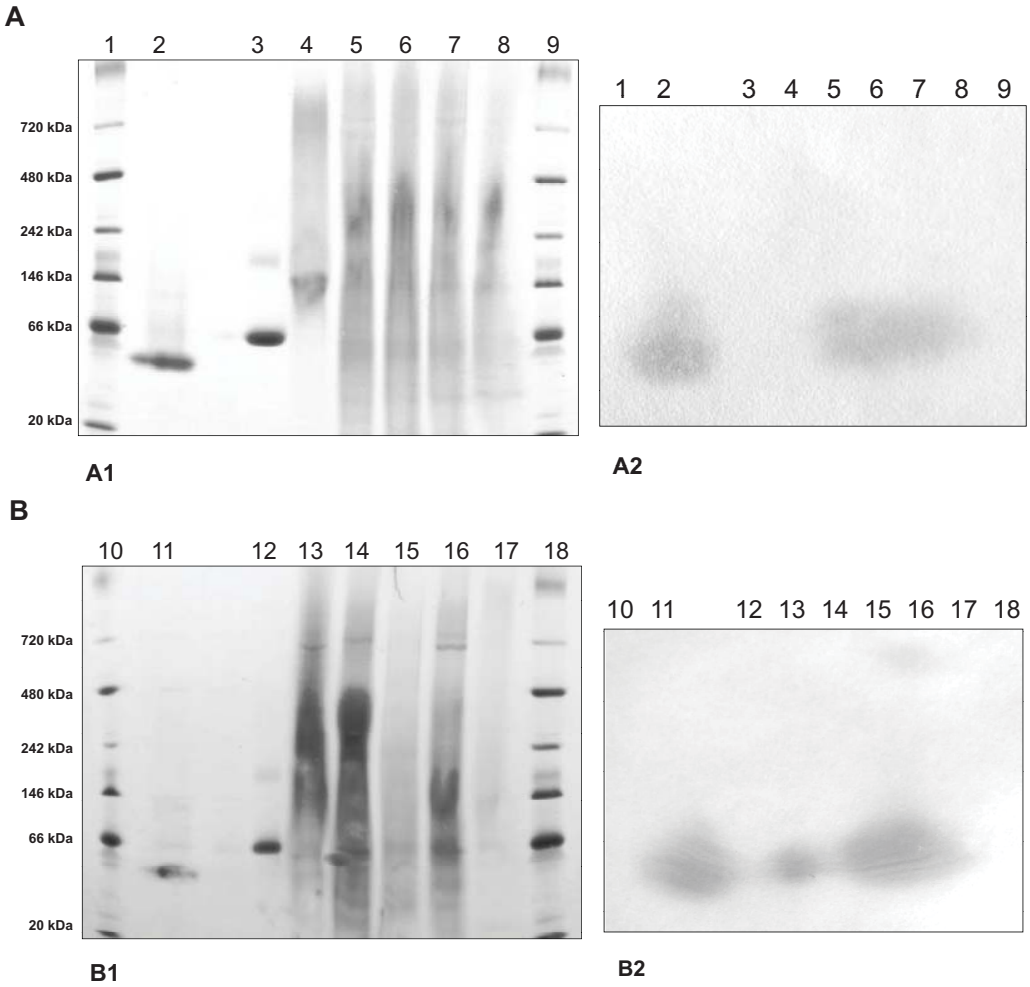


Fig. (3). BN-PAGE and HNL activity visualization on different plant protein extracts. (**A1**; **B1**) Blue Native PAGE: NativeMark™ Unstained Protein Standard (1;9;10;18), positive control *Pa*HNL (2;11); negative control BSA (3;12). *P. domestica* branches (4), flowers (5), sprouts (6), leaves (7), young leaves (8); *Malus domestica* (apple) kernels (13), *Passiflora capparidifolia* leaves (14), *Prunus laurus* ripe leaves (15), young leaves (16), green branches (17). (**A2**; **B2**) Activity assay: rac. mandelonitrile (30 mM in 100 mM citrate phosphate buffer, pH 4.5). Exposure time: six min (**A2**) and sixteen min (**B2**).

P. capparidifolia leaves did not show HNL activity after Blue Native PAGE (Fig. **3**, **B2**), although the protein extract prior to polyacrylamide gel electrophoresis displayed cyanogenesis activity (data not shown).

M. domestica kernels, *P. laurus* leaves and branches showed HNL activity as well, and the apparent molecular weight of the active species appears to be similar to that of the positive control (*Pa*HNL). Interestingly, longer exposure times lead to a weak band at high molecular weight (Fig. **3**, **B2**), which indicates the presence of a second HNL enzyme species, an HNL oligomer or an adduct of HNL with background protein(s).

Sequence Determination of P. domestica HNL

Although HNL activity in *Prunus domestica* has been reported [16], a primary protein sequence of a *P. domestica* HNL has not yet been elucidated. We excised a band that correlated to the activity signal on the HCN detection paper at approximately 60 kDa from a Coomassie stained BN-PAGE (Fig. **2**, **A3**) and subjected it to tryptic digestion. The LC-MS/MS data were analyzed by searching the NCBInr protein database. The sequence with highest similarity was *Prunus serotina* (black cherry) (*R*)-mandelonitrile lyase 1 (Swiss-Prot accession #P52706.1, 41 % sequence coverage, 13 identified peptides, 3 unique peptides). Since a *Prunus domestica* transcriptome is not available, data were matched against transcriptome data from *Prunus armeniaca* (apricot). The best hit (TSA: *Prunus armeniaca* Apricot (Seed) CL158. Contig5_LT-C transcribed RNA sequence, gi 590172944) showed 35% sequence coverage, 18 identified peptides and 19 unique peptides. BlastX was used to determine possible homologies in order to clarify a supposed function of the identified polypeptides. The resultant top rated hit, gi 590172944, is clearly an (*R*)-hydroxynitrile lyase, (blast score 1066, E-value 0.0) presenting a significant homology with other hydroxynitrile lyases described in public databases and a combination of the two homologous sequences from *P. serotina* and *P. armeniaca* allowed the unambiguous identification of large parts of the *Prunus domestica* HNL as shown in Fig. (**4**).

Ps	MEKSTMSAIL	LVLHLFVLLL	QYSEVHSLAT	TSNHDFSYLR	FAYDATDLEL
Par	MEKSTMSTIL	LVLHLFVLHL	QYSEVHSLAT	TSDHDFSYLS	FAYDATDLEL
Pd	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
Ps	EGSYDYVIV**G**	**GGTSGCPLAA**	**TLSEK**YK**VLV**	**LERGSLPTAY**	**PNVLTADGFV**
Par	EGSYDYVIVG	GGTSGCPLAA	TLSEKYK**VLV**	**LERGSLPTAY**	**PNVLTADGFV**
Pd	XXXXXXXX***XG***	***GGTSGCPLAA***	***TLSEK***XX***VLV***	***LERGSLPTAY***	***PNVLTADGFV***
Ps	**YNLQQEDDGK**	**TPVERFVSED**	**GIDNVR**GR**VL**	**GGTSMINAGV**	**YAR**ANTSIYS
Par	**YNLQQEDDGK**	**TPVERFVSED**	**GIDNVR**GR**VL**	**GGTSMINAGV**	**YAR**ANTSIYS
Pd	***YNLQQEDDGK***	***TPVERFVSED***	***GIDNVR***XX***VL***	***GGTSMINAGV***	***YAR***XXXXXXX
Ps	ASGVDWDMDL	VNKTYEWVED	TIVFKPNYQP	WQSVT**g**TAFL	EAGV**d**PNHGF
Par	SSGVDWDMDL	VNQTYEWVED	TIVYKPNSQS	WQSVTK**TAFL**	**EAGVHPNHGF**
Pd	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXX***TAFL***	***EAGVHPNHGF***
Ps	SLDHEAGTR**I**	**TGSTFDNK**GT	R**HAADELLNK**	GNSNNLR**VGV**	**HASVEKIIFS**
Par	**SLDHEAGTRI**	**TGSTFDNK**GT	R**HAADELLNK**	GNSNNLR**VGV**	**HASVEK**IIFS
Pd	***SLDHEAGTRI***	***TGSTFDNK***XX	***XHAADELLNK***	XXXXXXX***VGV***	***HASVEKIIFS***
Ps	**NAPGLTATGV**	**IYR**DSNGTPH	**r**AFVRSKGEV	IVSAGTIGTP	QLLLLSGVGP
Par	NA**d**GLTATGV	IYR**DSNGTPH**	**QAFVR**SKGEV	IVSAGTIGTP	QLLLLSGVGP
Pd	***NAPGLTATGV***	***IYRDSNGTPH***	***QAFVR***XXXXX	XXXXXXXXXX	XXXXXXXXXX
Ps	ESYLSSLNIP	VVL**SHPYVGQ**	**FLHDNPR**NFI	NILPPNPIEP	TIVTVLGISN
Par	ESYLSSLKIP	VVL**SHPYVGQ**	**FLHDNPR**NFI	NILPPNPIEP	TIVTVLGISN
Pd	XXXXXXXXXX	XXX***SHPYVGQ***	***FLHDNPR***XXX	XXXXXXXXXX	XXXXXXXXXX
Ps	DFYQCSFSSL	PFTTPPFSFF	PSTSYPLPNS	TFAHFASK**VA**	**GPLSYGSLTL**
Par	DFYQCSFSSL	PFTTPPFGFF	PSTSYPLPNS	TFAHFANKVA	GPLSYGSLTL
Pd	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXX***VA***	***GPLSYGSLTL***
Ps	**K**SSSNVRVSP	NVK**FNYYSNP**	**TDLSHCVSGM**	**KKIGELLSTD**	**ALKPYKVEDL**
Par	**t**SS-NVRVSP	NVK**FNYYSNP**	**TDLSHCVSGM**	**KKIGELLSTD**	**ALKPYKVEDL**
Pd	***K***XXXXXXXXX	XXX***FNYYSNP***	***TDLSHCVSGM***	***KKIGELLSTD***	***ALKPYKVEDL***
Ps	**PGIEGFNILG**	**IPLPKDQTDD**	**AAFETFCRES**	**VASYWHYHGG**	**CLVGKVLDGD**
Par	**PGIEGFNILG**	**IPLPKDQTDD**	**AAFETFCRES**	**VASYWHYHGG**	**CLVGKVLDGD**
Pd	***PGIEGFNILG***	***IPLPKDQTDD***	***AAFETFCRES***	***VASYWHYHGG***	***CLVGKVLDGD***
Ps	**FR**VTGI**d**ALR	**VVDGSTFPYT**	**PASHPQGFYL**	**MLGR**YVGIKI	LQERSASDLK
Par	**FRVTGINALR**	**VVDGSTFPYT**	**PASHPQGFYL**	**MLGR**YVGIKI	LQERSASDLK
Pd	***FRVTGINALR***	***VVDGSTFPYT***	***PASHPQGFYL***	***MLGR***XXXXXX	XXXXXXXXXX
Pd	ILDSLKSAAS	LVL			
Par	ILDSLKSAAS	FVF			
Pd	XXXXXXXXXX	XXX			

Fig. (4). Alignment of *P. serotina* (R)-mandelonitrile lyase 1 (Ps) and *P. armeniaca* (Par) sequences and the deduced *P. domestica* HNL (Pd) sequence. **bold:** Peptides retrieved from mass spectrometry; **bold and italics:** partial sequence deduced from matching peptides; **bold lower case letters**: mismatching amino acids.

CONCLUSION

Blue native PAGE in combination with HCN detection is a versatile method to visualize hydroxynitrile lyase activity after a non-denaturing protein gel electrophoresis in order to distinguish native HNL protein from inactive background proteins. The method reveals the molecular weight of the active enzyme. Mass spectrometric protein analysis of the active band can be done subsequently, in order to determine a partial protein sequence to pave the way for future applications of the respective enzyme. The assay is applicable to different protein preparations and more than one hydroxynitrile substrate can be used consecutively on the same native gel to give an indication about substrate preference of the investigated hydroxynitrile lyases.

LIST OF ABBREVIATIONS

HNL = Hydroxynitrile lyase

BN-PAGE = Blue native polyacrylamide gel electrophoresis

HCN = Hydrocyanic acid

MS = Mass spectrometry

LC = Liquid chromatography

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

We are grateful to Romana Wiedner and Thomas Stanzer for technical support. Barbara Petschacher and Christian Berg are kindly acknowledged for providing plant material and EVOCATAL for providing *At*HNL. 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) and the Federal Ministry of Science, Research and Economy (BMWFW), 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.

REFERENCES

[1] Poulton JE. Cyanogenesis in plants. Plant Physiol 1990;94(2):401-5.

[2] Purkarthofer T, Skranc W, Schuster C, Griengl H. Potential and capabilities of hydroxynitrile lyases as biocatalysts in the chemical industry. Appl Microbiol Biotechnol 2007;76(2):309-20.

[3] Gruber-Khadjawi M, Fechter MH, Griengl H. Cleavage and Formation of Cyanohydrins; In: Enzyme Catalysis in Organic Synthesis. Drauz K, Gröger H, May O, Eds. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2012 : pp. 947-90.

[4] Winkler M, Glieder A, Steiner K. Hydroxynitrile Lyases: From Nature to Application; In: Comprehensive Chirality C-X Bond Formation; Carreira EM, Yamamoto H, Eds. 2012; pp. 350-71.

[5] Lanfranchi E, Steiner K, Glieder A, *et al*. Recent Developments in Hydroxynitrile Lyases for Industrial Biotechnology. Recent Pat Biotechnol 2013;7(3):197-206.

[6] Hussain Z, Wiedner R, Steiner K, *et al*. Characterization of two bacterial hydroxynitrile lyases with high similarity to cupin superfamily proteins. Appl Environ Microbiol 2012;78(6):2053-5.

[7] Dadashipour M, Asano Y. Hydroxynitrile Lyases: Insights into Biochemistry, Discovery, and Engineering. ACS Catal 2011;1(9):1121-49.

[8] Kassim MA, Rumbold K. HCN production and hydroxynitrile lyase: a natural activity in plants and a renewed biotechnological interest. Biotechnol Lett 2014;36(2):223-8.

[9] Xu L-L, Singh BK, Conn EE. Purification and characterization of mandelonitrile lyase from *Prunus lyonii*. Arch Biochem Biophys 1986;250(2):322-8.

[10] Yemm RS, Poulton JE. Isolation and characterization of multiple forms of mandelonitrile lyase from mature black cherry (*Prunus serotina Ehrh.*) seeds. Arch Biochem Biophys 1986;247(2):440-5.

[11] Kuroki GW, Conn EE. Mandelonitrile lyase from *Ximenia americana L.*: stereospecificity and lack of flavin prosthetic group. Proc Natl Acad Sci 1989;86(18):6978-81.

[12] Bhunya R, Mahapatra T, Nanda S. *Prunus armeniaca* hydroxynitrile lyase (*Pars*HNL)-catalyzed asymmetric synthesis of cyanohydrins from sterically demanding aromatic aldehydes. Tetrahedron: Asymmetry 2009;20(13):1526-30.

[13] Tükel SS, Yildirim D, Alagöz D, Alptekin Ö, Yücebilgiç G, Bilgin R. Partial purification and immobilization of a new (R)-hydroxynitrile lyase from seeds of *Prunus pseudoarmeniaca*. J Mol Catal B Enzym 2010;66(1-2):161-5.

[14] 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 2010;46(6):456-65.

[15] Hickel A, Heinrich G, Schwab H, Griengl H. Screening for hydroxynitrile lyases in plants. Biotechnol Tech 1997;11(1):55-8.

[16] 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 2004;30(3-4):105-8.

[17] Asano Y, Tamura K, Doi N, Ueatrongchit T, H-Kittikun A, Ohmiya T. Screening for New Hydroxynitrilases from Plants. Biosci Biotechnol Biochem 2005;69(12):2349-57.

[18] Drochioiu G, Arsene C, Murariu M, Oniscu C. Analysis of cyanogens with resorcinol and picrate. Food Chem Toxicol 2008;46(11):3540-5.

[19] Kassim MA, Sooklal SA, Archer R, Rumbold K. Screening for hydroxynitrile lyase activity in non-commercialised plants. South African J Bot 2014;93:9-13.

[20] Andexer J, Guterl J-K, Pohl M, Eggert T. A high-throughput screening assay for hydroxynitrile lyase activity. Chem Commun 2006;9(40):4201-3.

[21] Krammer B, Rumbold K, Tschemmernegg M, Pöchlauer P, Schwab H. A novel screening assay for hydroxynitrile lyases suitable for high-throughput screening. J Biotechnol 2007;129(1):151-61.

[22] Arndt C, Koristka S, Bartsch H, Bachmann M. Native polyacrylamide gels. Methods Mol Biol 2012;869:49-53.

[23] Schägger H, von Jagow G. Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. Anal Biochem 1991;199(2):223-31.

[24] Lin-Cereghino J, Wong WW, Xiong S, Giang W, Luong LT, Vu J, *et al*. Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast *Pichia pastoris*. Biotechniques 2005;38(1):48-52.

[25] Weis R, Luiten R, Skranc W, Schwab H, Wubbolts M, Glieder A. Reliable high-throughput screening with *Pichia pastoris* by limiting yeast cell death phenomena. FEMS Yeast Res 2004;5(2):179-89.

[26] Heukeshoven J, Dernick R. Improved silver staining procedure for fast staining in PhastSystem Development Unit. I. Staining of sodium dodecyl sulfate gels. Electrophoresis 1988;9(1):28-32.

[27] Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem 1996;68(5):850-8.

[28] Solís A, Solís-Oba M, Pérez HI, Manjarrez N, Cassani J. Hydroxynitrile Lyases from *Prunus* Seeds in the Preparation of Cyanohydrins. Biosci Biotechnol Biochem 2011;75(5):985-6.

[29] Okrob D, Metzner J, Wiechert W, Gruber K, Pohl M. Tailoring a stabilized variant of hydroxynitrile lyase from *Arabidopsis thaliana*. Chembiochem 2012;13(6):797-802.

[30] Andexer J, von Langermann J, Mell A, *et al*. An R-selective hydroxynitrile lyase from *Arabidopsis thaliana* with an alpha/beta-hydrolase fold. Angew Chem Int Ed Engl 2007;46(45):8679-81.

[31] Hasslacher M, Schall M, Hayn M, Griengl H, Kohlwein SD, Schwab H. Molecular Cloning of the Full-length cDNA of (S) - Hydroxynitrile Lyase from *Hevea brasiliensis*: Functional Expression in *Escherichia coli* and *Saccharomyces cerevisiae* and Identification of an active site residue. J Biol Chem 1996; 271: 5884-91.

[32] Breithaupt H, Pohl M, Bönigk W, Heim P, Schimz K-L, Kula M-R. Cloning and expression of (R)-hydroxynitrile lyase from *Linum usitatissimum* (flax). J Mol Catal B Enzym 1999; 6(3): 315-32.

Received: December 1, 2014 Revised: January 26, 2015 Accepted: May 6, 2015

Chapter 3
Unique and Novel Enzymes Catalyze Cyanide Release from Fern

Elisa Lanfranchi[1], Tea Pavkov-Keller[1,2], Margit Winkler[1,3], Eva-Maria Koehler[1], Matthias Diepold[2], Kerstin Steiner[1], Barbara Darnhofer[1,4,5], Juergen Hartler[5,6], Tom Van Den Bergh[7], Henk-Jan Joosten[7], Mandana Gruber-Khadjawi[1], Gerhard G. Thallinger[1,5,6], Ruth Birner-Gruenberger[1,4,5], Karl Gruber[1,2], Anton Glieder[1,2]*

[1] Austrian Centre of Industrial Biotechnology, Petersgasse 14, 8010 Graz, Austria
[2] Graz University of Technology, Institute of Molecular Biotechnology, Nawi Graz, Petersgasse 14, 8010 Graz, Austria
[3] University of Graz, Institute of Molecular Biosciences, Nawi Graz, Humboldtstrasse 50, 8010 Graz, Austria
[4] Medical University of Graz, Institute of Pathology, Stiftingtalstrasse 24, 8010 Graz, Austria
[5] Omics Center Graz, BioTechMed Graz, Stiftingtalstraße 24, 8010 Graz, Austria
[6] Graz University of Technology, Institute of Knowledge Discovery, Bioinformatics Group, Petersgasse 14, 8010 Graz, Austria
[7] Bio-prodict BV, Nieuwe Marktstraat 54E 6511 AA Nijmengen, The Netherlands

Corresponding Author:
Anton Glieder
Graz University of Technology, Institute of Molecular Biotechnology, Nawi Graz, Petersgasse 14/V, 8010 Graz Austria
Email: a.glieder@tugraz.at Tel: +43 316 873 4077 Fax: +43 316 873 9302

Abstract

Hydroxynitrile lyases (HNLs) are highly diverse enzymes that catalyze the release of cyanide as a defense system of plants and animals. Moreover, HNLs play an important role in chemical manufacturing, due to their ability to catalyze the stereoselective synthesis of enantiomerically pure cyanohydrins. In spite of their biological and industrial importance, only a small number of HNL sequences have been published so far. The discovery of new representatives is challenging and due to their diversity classical homology approaches fail to discover sequences of truly novel enzymes. In this study we identified novel HNLs from the fern *Davallia tyermannii* by combining activity screening, and MS-MS data analysis with transcriptome sequencing. The four isoenzymes are the first described HNLs with a Bet v 1 fold. Moreover, a new catalytic mechanism for fast cyanide release was identified. *Dt*HNL's biochemical properties open perspectives for the development of a new competitive class of biocatalysts and show that nature in spite of the vast knowledge about sequenced genomes is still a valuable source of variability and novelty for enzyme discovery.

Introduction

Hydroxynitrile lyases(HNLs; EC 4.1.2.10, EC 4.1.2.11, EC 4.1.2.46 and EC 4.1.2.47; alternative names: oxynitrilase, hydroxynitrilase, mandelonitrile lyase, hydroxymandelonitrile lyase) catalyze the degradation of cyanohydrins into the corresponding aldehyde or ketone and hydrogen cyanide[1–3]. They have been mostly found in plants, but few very recent examples exist from bacteria[4,5] and animals[6] as well. Specifically, HNLs are involved in the cyanogenic pathway, a widespread defense mechanism against herbivores and pathogens. Hydrogen cyanide (HCN) is a poisonous and volatile chemical compound which rapidly diffuses into the environment. Plants produce cyanogenic glycosides *e.g.* prunasin or amygdalin as secondary metabolites and store them in different tissues such as kernels or leaves. Upon tissue disruption, the compound is processed in a two-step reaction. First, a β-glucosidase cleaves the sugar from the aglycone moiety, releasing the cyanohydrin, and subsequently HCN is formed in the HNL catalyzed reaction. Examples of cyanogenic plants expressing HNLs are known from *Rosaceae*, *Linaceae*, *Euphorbiaceae*, *Clusiaceae*, *Olaceae*, *Graminaceae* and *Polypodiaceae*. In addition, a protein with HNL activity has been found and characterized also in the non-cyanogenic plant *A. thaliana*[7].

HNLs are well-established tools for several biocatalytic industrial processes. They can perform the reverse condensation reaction, and synthesize α-cyanohydrins in a stereospecific manner. Cyanohydrins are suitable building blocks for a number of follow-up reactions, and the products find application in agrochemical, pharmaceutical and cosmetic industry[8–11]. Both (*R*)- and (*S*)-selective HNLs are employed as biocatalysts, and they often have been engineered in order to maximize conversion and enantiomeric purity of specific target products[12–14]. Although HNL activity has been shown in a wide variety of species[15–18], only a small number of HNLs was isolated and characterized, and hardly ever their sequences were identified. However, knowledge of an enzyme's amino acid sequence is essential for its recombinant production and the improvement of its biocatalytic function. Therefore, only few are employed for industrial processes and further enzyme sequence discovery is necessary.

HNLs are non-homologous isofunctional enzymes (NISE), a group of unrelated proteins that catalyze the same chemical reaction as a result of convergent evolution[19]. Nature evolved several different solutions to address the same widespread chemical reaction within the plant kingdom, and based on today's knowledge, we can expect more HNL families yet to be discovered. Classical homology approaches, however, fail in the discovery of completely new enzymes. This represents a significant limitation for the discovery of new HNLs with unprecedented structure and biocatalytic features, making HNLs an appealing target for enzyme discovery.

The original classification divided HNLs in two groups; FAD-dependent HNLs, typically (*R*)-selective from *Rosaceae*, and FAD-independent HNLs, more diverse in sequence and structure[11]. Nowadays, the scheme is more complex as new HNL types have been found: five different protein folds have been discovered so far and, they do not share any conserved motif. Crystal structures of HNLs belonging to cupin[20], GMC oxidoreductase[21], α/β-hydrolase[22,23] and peptidase S10[24] families have been determined. Moreover, by sequence homology, an HNL has been described as member of the zinc-binding dehydrogenase family[25]. Recently, a new HNL sequence from the invasive millipede *Chamberlinius hualiensis* has been released, but there is no indication about its tertiary structure[6].

Finally there is a number of characterized HNLs with yet unpublished amino acid sequences and protein folds, for example, *Pat*HNL (*Prunus amygdalus turcomanica*)[26], *Pe*HNL (*Passiflora edulis*)[27], and the fern HNL from *Phlebodium aureum* (*Pha*HNL)[28]. A biochemical characterization and experiments towards cyanohydrin synthesis were performed with purified proteins from the natural hosts.

We focused on HNLs from fern plants, due to the highly promising enzymatic properties reported by Wajant[28]. *Phlebodium aureum* is not the only cyanogenic fern, and it has been shown that several ferns contain cyanogenic glycosides.

Here, we describe the discovery of a new and even unique HNL from the white rabbit's foot fern *Davallia tyermannii* (alternative names: *Humata tyermannii*; *Humata tyermannii* T. Moore; *Davallia teyermanii*; *Davallia tyermannii* (T.Moore) Backer). A classical HNL identification workflow comprises several steps of protein purification

from the natural source, followed by degenerate primers designed based on result of mass spectrometry and subsequent sequence identification by PCR, to end up in a long, laborious and time consuming process. To our knowledge, it is the first time where several *omics* techniques were 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 *Dt*HNL and elucidated its catalytic mechanism. Finally, we investigated the presence of HNLs in different fern families.

Results

From the enzymatic activity to the sequence

The direct relation between cyanogenesis in plants and the presence of a hydroxynitrile lyase (HNL) has not been proven yet. Therefore, we confirmed HNL activity in the cyanogenic fern *Davallia tyermannii* by the addition of (*RS*)-2-hydroxy-2-phenylacetonitrile (racemic mandelonitrile) to a protein preparation and subsequent detection of the release of hydrogen cyanide (**Supplementary Fig. 1**). We chose mandelonitrile as it is the natural cyanohydrin identified in the genus *Davallia* and different other fern genera[28,29,30].

The screening workflow combined several techniques. We first collected the information about all expressed genes, which is well described by a normalized sequenced transcriptome. For this purpose, high quality mRNA from *D. tyermannii* leaves and croziers were isolated and the normalized cDNA library was subjected to 454 FLX sequencing (the quality of the transcriptome and assembling results are reported in **Supplementary Results 2**). Known HNL protein sequences were subjected to a blast[31] search. Low identity and sequence coverage in the alignments confirm that *Dt*HNL is distinct from known HNL classes (**Supplementary Results 3**). The sequenced transcriptome is only partially useful for enzyme discovery, when no information about the primary sequence, conserved motif or protein family is available. On the other hand, it provides the perfect database for analysis of active protein fractions. Thus, proteins from active tissues were subjected to anion exchange chromatography and subsequently to a BN PAGE coupled HCN detection assay[32]. We identified a protein indicative for HNL activity which correlated with a 20 kDa band (**Fig. 1a**). Although one purification step cannot yield pure protein, several elution fractions showed an enrichment of *Dt*HNL (**Fig. 1a**).

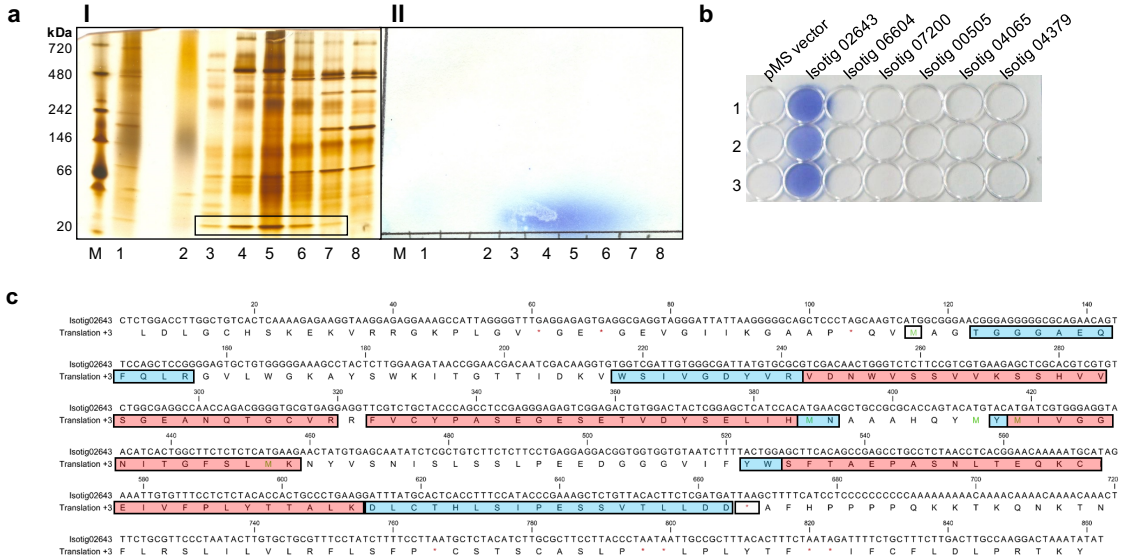


Figure 1. Identification of the *Dt*HNL sequence. A: BLUE NATIVE PAGE followed by HNL activity assay. aI: Concentrated fractions of anion exchange purification were applied separately on BN PAGE. NativeMark™ Unstained Protein Standard (Thermo Fisher Scientific) (M); total protein extract from *D. tyermannii* leaves (1); flow through (2); elution fractions (3-8); active bands at 20 kDa are highlighted in the box. **aII:** 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 10 F' strains expressing six putative HNL proteins. Each sample was tested in triplicate: 30 µL (1), 20 µL (2), 10 µL (3) of cell free lysate, respectively. *E. coli* TOP 10 F' 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 respective isotig or contig number found in the transcriptome. **c: nucleotide and amino acid sequence of isotig 02643 (*Dt*HNL1).** Parts detected by mass spectrometry are labeled. Highlighted 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.

We excised the bands that correlated to the activity signal at 20 kDa from a silver stained BN PAGE (**Fig. 1a**) and subjected them to in gel tryptic digestion and LC-MS/MS. 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 (**Appendix, Supplementary Dataset 1**). Finally, six candidates were recombinantly expressed in *E. coli* TOP 10F´ and tested for HNL activity (**Fig. 1b; Supplementary Table 3**).

The protein coded 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 peptides retrieved from mass spectrometry cover 72% of the translated ORF (Fig 1c). The transcriptome data revealed three highly similar sequences to the confirmed HNL: translated ORFs of isotig02641, isotig07602 and contig00751 show 93.5% identity with isotig02643 (Supplementary Fig. 5.1). 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* BL21 (DE3) Star, the HNL activity assay confirmed cyanogenic activity of all three isoenzymes (Supplementary Fig. 5.2). The four proteins coded by the ORF of isotig02643, 02641, 07602 and contig00751 were named *Dt*HNL1, *Dt*HNL2, *Dt*HNL3 and *Dt*HNL4, respectively.

Characterization of *Dt*HNL isoenzymes

Biochemical features of all four isoenzymes were determined after heterologous expression and purification of the His-tagged proteins by nickel affinity chromatography. The influence of pH and temperature on the enzymatic activity of *Dt*HNLs was determined using racemic mandelonitrile as substrate (Fig. 2a and 2b).

Maximum activity was recorded at pH 5.0. All isoenzymes were active at pH 2.5 but inactive at pH 2.0 (Fig. 2a). Enzymatic activity at low pH is a considerable advantage for HNLs in terms of their application, since cyanohydrins easily degrade at pH ≥ 5.0 as indicated by the background reaction line (Fig. 2a). Therefore, standard enzymatic syntheses of cyanohydrins are preferably performed at pH 4.0 or lower and stable HNLs in acidic environment are desired. We investigated *Dt*HNL stability at pH 2.5 and 4.0 by incubation at 8°C for 72 hours. Although the activity constantly decreased over time, residual activity was more than 50% after 24 hours (Fig. 2c). The enzymes were more stable at pH 4.0 except for *Dt*HNL1, which displayed only 38% of the initial activity after 48 h (Fig. 2d). An exhaustive table, comprehensive of further pH conditions, can be found at Supplementary Tables 5.1-5.4.

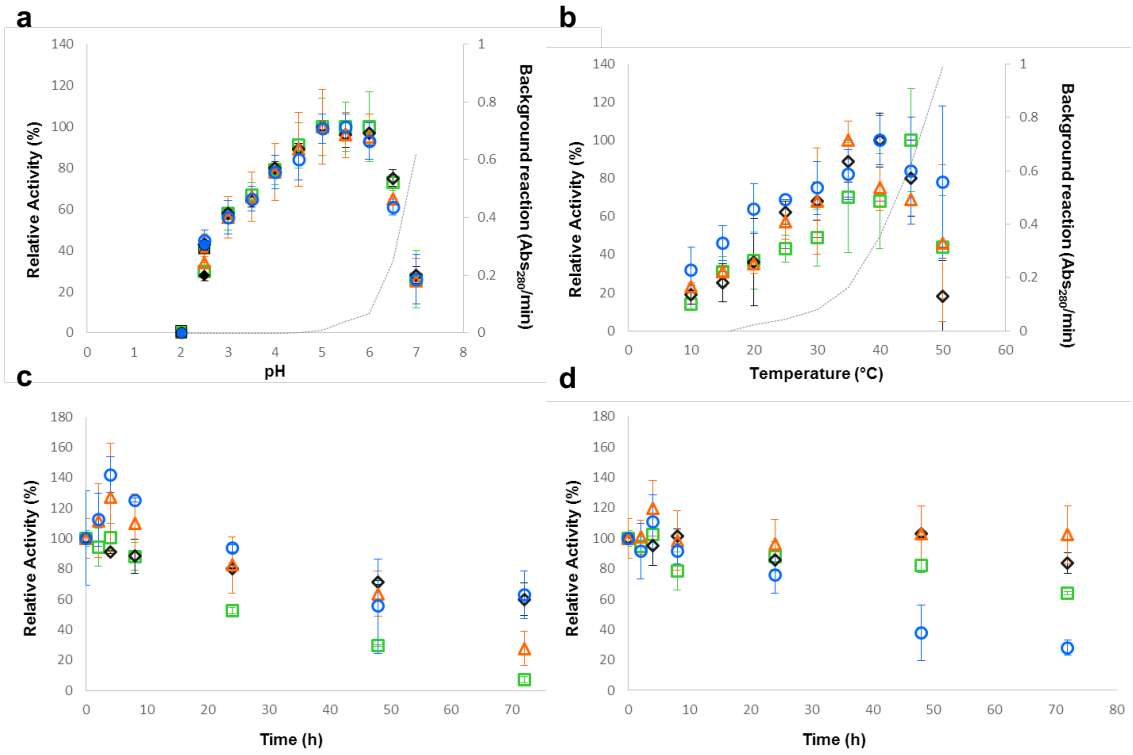


Figure 2. Biochemical properties of *Dt*HNL isoenzymes. *Dt*HNL1 (○); *Dt*HNL2 (△); *Dt*HNL3 (◊); *Dt*HNL4 (□); grey dashed line indicates the spontaneous degradation of racemic mandelonitrile (background reaction). Error bars indicate the standard deviation obtained from three independent experiments (exceptions are reported below). **a: activity at different pH values.** Initial rates recorded in HCl-potassium chloride buffer (pH 2.0 and 2.5) is reported with filled symbols, while empty symbols indicate initial rates in sodium citrate-phosphate buffer (pH 2.5 – pH 7.0). **b: activity at different temperatures. c: enzyme stability at pH 2.5.** Error bars indicate the standard deviation calculated from two independent experiments. **d: enzyme stability at pH 4.0.** Error bars indicate the standard deviation calculated from two independent experiments.

Figure 2b shows *Dt*HNL activity at different temperatures. The optimum was observed at 35°C for *Dt*HNL2 and 3, 40°C for *Dt*HNL1 and 45°C for *Dt*HNL4, respectively. However, there is a strong variability between the four isoenzymes and also within different replicates of the same isoform, indicated by the high values of standard deviation. One of the reasons is the high degradation rate of mandelonitrile, indicated by the grey dashed line (**Fig. 2b**). As a consequence, the substrate is only partially available for the enzymatic reaction.

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 *Dt*HNLs have

high affinity for (*R*)-mandelonitrile with K_m values of 0.297 ± 0.029 mM *Dt*HNL1, 0.451 ± 0.047 mM *Dt*HNL2, 0.639 ± 0.079 mM *Dt*HNL3 and 0.591 ± 0.089 mM *Dt*HNL4, respectively. These results are consistent with K_m values of reported (*R*)-HNL enzymes[2]. The turnover numbers were 141.9 s^{-1} *Dt*HNL1, 173.2 s^{-1} *Dt*HNL2, 319.8 s^{-1} *Dt*HNL3 and 231.9 s^{-1} *Dt*HNL4, respectively. Interestingly, the catalytic efficiency, defined as k_{cat}/K_m, is similar for the four proteins and the average is 438.5 ± 58.8 s^{-1} mM^{-1} **(Supplementary Fig. 6.4)**.

Enzyme kinetics for the synthesis of (*R*)-mandelonitrile was determined as well. The activity for cyanohydrin formation was measured at different concentrations of benzaldehyde and a constant excess of cyanide in a biphasic system. *Dt*HNL1 v_{max} is 376.9 ± 88.9 µmol min^{-1} mg^{-1}, similar to the maximal velocity of cyanogenesis from mandelonitrile 373 ± 7 µmol min^{-1} mg^{-1}. *Dt*HNL1 affinity for benzaldehyde was lower than the one of mandelonitrile (K_m 13 mM). Finally, k_{cat} was 71.6 s^{-1}, and the enzymatic efficiency k_{cat}/K_m was 5.3 s^{-1} mM^{-1}. Note that the values obtained for the synthesis of (*R*)-mandelonitrile are apparent constants, since the reaction was carried out in a biphasic system with strong agitation and emulsion formation, which can influence the enzyme kinetic.

*Dt*HNL structure and reaction mechanism

We determined the crystal structure of *Dt*HNL using SeMet-SAD to a resolution of 1.85 Å. The enzyme is a dimer and exhibits a Bet v 1-like fold. The Bet v I superfamily is composed of sequences related to the major Birch (*Betula verrucose*) pollen allergen Bet v 1. 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 v 1 superfamily this cavity plays important roles in the binding and metabolism of large, hydrophobic compounds such as lipids, hormones, and antibiotics[33]. We also determined the structures of complexes of *Dt*HNL with hydroxy-benzaldehyde, benzoic acid and (*R*)-mandelonitrile/benzaldehyde using soaking techniques (**Supplementary Table 7**). Clear electron density was observed for those ligands bound in the canonical binding cleft of the Bet v 1 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 (**Fig. 3b**). 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 Figure 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 *Dt*HNL.

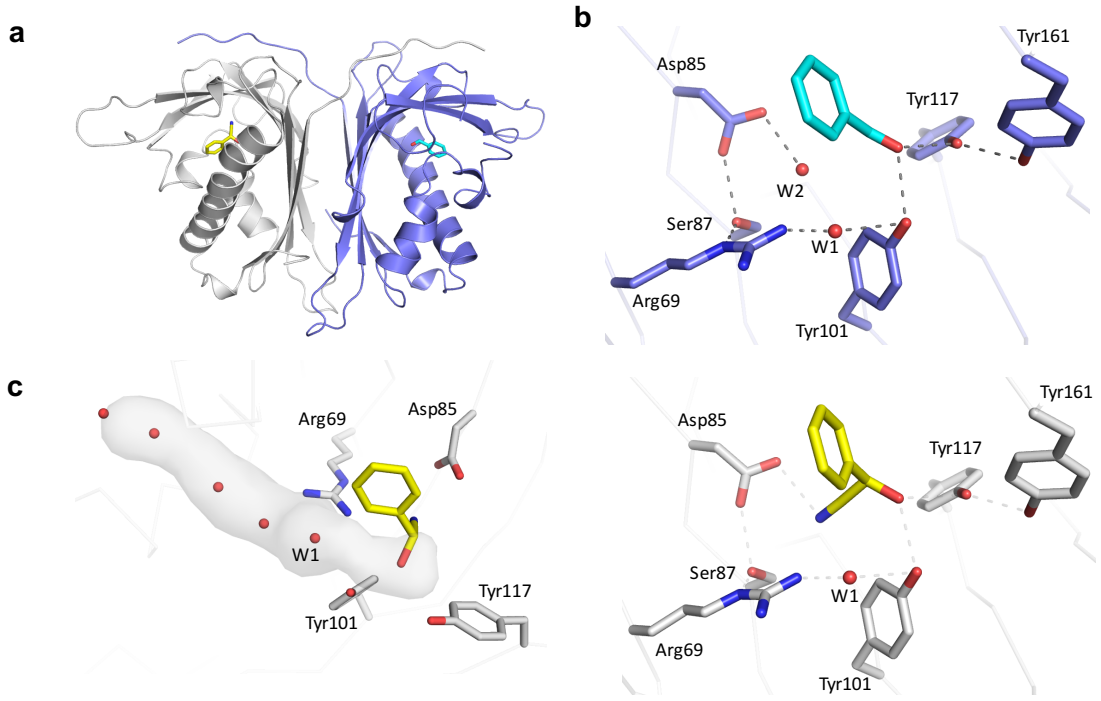


Figure 3. Structure and active site analysis. a: Overall structure of *Dt*HNL1 dimer. Individual monomers are shown in ribbon (grey, blue) and bound (*R*)-mandelonitrile (yellow) and benzaldehyde (cyan) are shown in stick representation. **b: The active site.** Plausible hydrogen-bonding network between the bound (*R*)-mandelonitrile and benzaldehyde, as present in the complex structure determined after soaking crystals with (*R*)-mandelonitrile, are indicated by dashed lines. Residues important for the enzymatic activity are shown in stick representation and two water molecules as spheres. The bridging water is labeled W1. **c: Water channel** extending from the active site to the surface of the protein, as calculated using CAVER[34].

The observed polar interactions between (*R*)-mandelonitrile and *Dt*HNL 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 (pKa of tyrosine ~10) and there is no other amino acid residue in the 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, H_2O, Arg^0 and c) Tyr-O^-, H_2O, 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, H_2O, 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 the mechanistic proposal a number of amino acids were replaced and the enzymatic activity of these variants was determined (**Supplementary Table 8**). 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 8**).

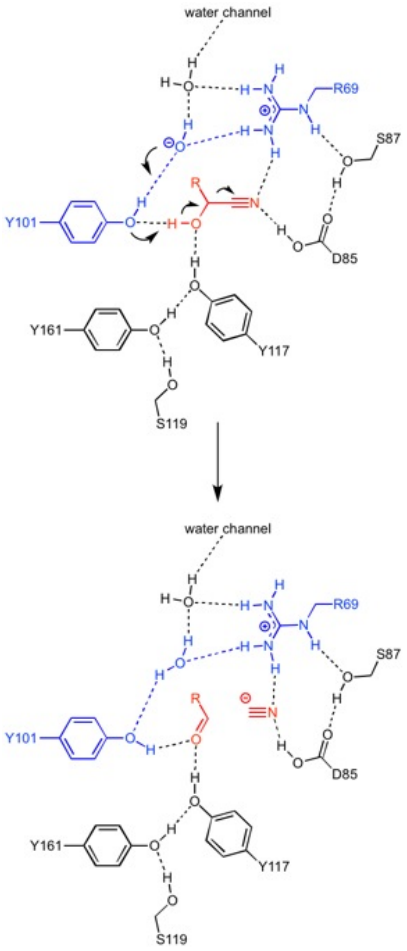


Figure 4. Proposed catalytic mechanism for the *Dt*HNL1 cyanohydrin cleavage of (*R*)-mandelonitrile based on complex-crystal structures and mutagenesis experiments.

Different HNL classes from ferns

Ferns are a group of sparsely studied vascular plants, which include different 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 genomes[35]. The sequence of *Dt*HNL is the first HNL sequence identified from ferns, however several fern species adopt cyanogenesis as chemical defense. For these reasons, we investigated whether cyanogenic ferns express homologous HNLs or they developed them independently, as often happens in seed plants.

The braken fern *Pteridium aquilinum* (L.) Kuhn is widespread throughout the northern hemisphere and Africa and belongs to the *Dennstaedtiaceae* family[35]. It is another well-known example of cyanogenic fern in addition to *D. tyermannii* and *P. aureum* but wide spread in nature and easily accessible and therefore, a preferred additional case for our studies. First, high quality mRNA was isolated from enzymatically active leaves and croziers. The transcriptome was obtained by sequencing the normalized cDNA library (Microsynth, CH; the quality of the transcriptome and assembling results are reported in Supplementary Results 2). Again the transcriptome did not show proteins with significant similarity to any known HNL sequences upon tblastn search (Supplementary Results 3). When *Dt*HNL1 was used as query for tblastn, we obtained 17 hits (Supplementary Table 7.1). We narrowed the list to two sequences with identity to *Dt*HNL1 above 35%. Isotig02775 and isotig02778 belong to the same isogroup and their ORFs encode for two proteins with 96% identity. (Supplementary Table 9). The identity between the two sequences and *Dt*HNL1 is 41%, but most interestingly, the catalytic residues Arg69 and Tyr101 are conserved, and the two Tyr in the active site 117 and 161 as well (Fig. 6). *Dt*HNL1 was subjected to a second tblastn search in a different translated transcriptome from *P. aquilinum* obtained during its gametophyte life stage (http://www.biomedcentral.com/1471-2164/12/99/additional) [35]. The protein coded by the ORF of contig4149 is 99% identical to isotig02775, which strengthens our result (Fig. 6). However, when we subjected the three amino acid sequences, isotig02275,

isotig02778 and contig4149 to a tblastn search in *D. tyermannii* transcriptome, isotig04300 was obtained as the best hit instead of expected *Dt*HNL1.

Isotig04300 belongs to the Bet v 1 protein superfamily, however, it was not identified by mass spectrometry in fractions of *D. tyermannii* with HNL activity (**Appendix, Supplementary Dataset 1**) which indicates that these sequences do not contribute to the HNL activity.

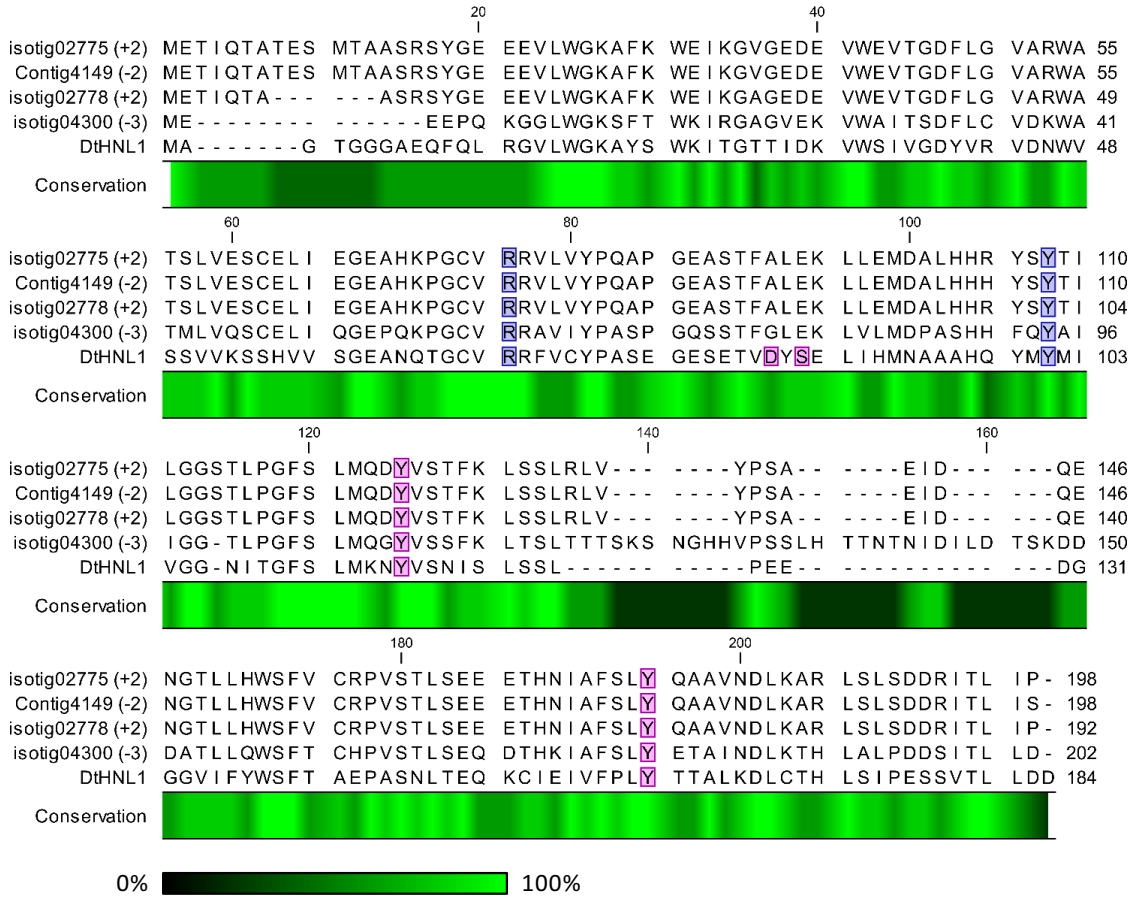


Figure 5. Multiple sequence alignment. Isotig02775 and isotig02778 are proteins with similarity to *Dt*HNL1 from the transcriptome of *P. aquilinum* leaves. Contig4149 was found in the transcriptome obtained from a *P. aquilinum* gametophyte[35]. Isotig04300 is the sequence most similar to isotig02775 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.

To test this hypothesis, isotig02775 was expressed as soluble protein in *E. coli* and purified by affinity chromatography. The enzymatic activity of isotig02775 was zero for racemic mandelonitrile degradation, which confirmed the theory. The protein was also inactive after mutation of Ala92 for Ser (The position corresponds to Asp85 in *Dt*HNL1). A double mutation of isotig02775 Ala92Asp and Glu94Ser simulating *Dt*HNL at this site led to inclusion body formation. Therefore, it was not possible to determine the enzymatic activity.

*Dt*HNL catalyzes the synthesis of different cyanohydrins

To test the biocatalytic ability of *Dt*HNL, we studied the activity and stereo selectivity for the synthesis of different cyanohydrins. In HNL-catalyzed reactions, high yields of an enantiomerically pure product are hampered by spontaneous non-enzymatic formation of racemic cyanohydrin and racemization of the product due to equilibration of the reaction. Therefore, it is especially important to suppress the chemical condensation and racemization of cyanohydrins and set up conditions for a fast and efficient reaction that is able to compete with the non-enzymatic transformations. The decrease of the water content by the use of a biphasic system and low pH conditions are two solutions extensively reported in HNL literature. The choice of pH and the organic solvent-water phase ratio depend also on the activity and stability of the biocatalyst. Here, we chose pH 4.0 and ratio 2:1 v/v organic phase/aqueous phase. Water is necessary for the enzymatic activity, as described above. *Dt*HNL was relatively stable at pH 4.0 for 24 hours (**Fig. 2d**) and ca. 80% of its maximum activity was recovered (**Fig. 2a**). Finally, we used 3 mg of *Dt*HNL1 (equal to 0.026 mol% for aldehyde substrates or 0.043 mol% for the ketone substrate). A time course study was carried out for 24 hours (**Table 1**).

Table 1. Synthesis of cyanohydrins with *Dt*HNL1. The synthesis of cyanohydrins was performed with 3 mg of *Dt*HNL1. 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, 10° C and 1000 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 was detected. ND, not determined.

Entry	Substrate → Product		0.5 hours		2 hours		6 hours		24 hours	
			conv %	ee %	conv %	ee %	conv %	ee %	conv %	ee %
1	benzaldehyde → (*2R*)-2-hydroxy-2-phenylacetonitrile	***Dt*HNL1**	95.1	99.0	97.4	94.5	97.9	85.9	99.5	61.4
		Control	-4.9	0.0	11.7	0.8	18.5	0.3	39.7	0.5
2	2-chlorobenzaldehyde → (*2R*)-2-(2-chlorophenyl)-2-hydroxyacetonitrile	***Dt*HNL1**	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	***Dt*HNL1**	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	***Dt*HNL1**	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-hydroxyacetonitrile	***Dt*HNL1**	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	***Dt*HNL1**	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	---

*Dt*HNL1 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 ≥99% enantiomeric excess were obtained in less than 2 hours for the synthesis of (*R*)-mandelonitrile and (*2S*)-furan-2-yl-hydroxyacetonitrile (**Table 1, entries 1 and 5**). As expected for such 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 ≥ 99% ee in 24 hours, when 0.0244e-2 mol% of *Dt*HNL1 were employed; about hundred times less catalyst. Depending on the substrate, the enzyme competes with the chemical condensation reaction for substrate availability. Exceptions are 1-phenylethanone and 3-phenylprop-2-enal, where no product was detected in the non-enzymatic transformation (**Table 1, entries 4 and 6**).

Comparably low enantiomeric excess was obtained for (*2R*)-2-hydroxy-4-phenylbutanenitrile (**Table 1, entry 3**) as a consequence of slow reaction combined with fast unselective background. Consistent results were observed with less biocatalyst amounts as well: 90.4% conversion, 43.5% ee and 97.0% conversion, 50.4% ee were retrieved with 0.017e-1 mol% and 0.087e-1 mol% of *Dt*HNL1 respectively, after 24 hours. These results indicate that *Dt*HNL1 displays a lower stereo selectivity 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-hydroxyacetonitrile, (*2R*)-2-hydroxy-4-phenylbut-3-enenitrile and (*2S*)-furan-2-yl-hydroxyacetonitrile (**Table 1**, **entries 1,2,4,5**) is due to the cyanohydrin racemization in buffer[36]. Finally, *Dt*HNL1 can convert aromatic ketones such as 1-phenylethanone as well, however, in low analytic yield (<24%), which reveals ketones as more difficult substrates for this enzyme (**Table 1, entry 6**), as observed for all other known HNLs.

Discussion

Several bioinformatic tools have been developed for enzyme discovery, especially for wide screening of metagenomic libraries and sequenced genomes of bacteria and fungi. They are mainly based on similarities between sequences or common features, such as characteristic protein motif. Therefore, only homologous genes can be identified, even if they are distantly related. However, species varieties evolved different solutions to address the same issue. For example, plants adopt several defense tactics against herbivory, chemical or mechanical defenses. This is valid at the molecular level as well which is exemplified by hydroxynitrile lyases. HNLs reveal that nature could develop numerous different strategies to address the same chemical reaction and no single common feature between the different classes has been identified yet.

The discovery of new HNLs (*Dt*HNL and *Chua*HNL[6]) further proves that a search only based on sequence similarity or structure based features is limited and prevents the discovery of truly new enzymes. This fact might be underestimated for other enzymes where there was less commercial interest so far to identify alternative enzymes catalyzing similar reactions. However, some other known cases such as alkane hydroxylation catalyzed by completely different enzymes (for example methane monooxygenases[37], cytochrome P450s (such as CYP153AA6[38] or CYP52[39] from *Candida*), alkB from *Pseudomonas*[40], or fungal peroxygenases[41]) indicate that this is well represented in nature for other biochemical reactions. Starting from scratch, we discovered a protein unique within the Bet v 1 superfamily. The most similar characterized protein is the lachrymatory factor synthase from onion and the sequence identity is less than 28%. Other family members are polyketide cyclases2 and abscisic acid receptor; most other sequences are annotated as unknown proteins. Norcoclaurine synthases or proteins with RNA activity belong to the Bet v 1 protein superfamily as well, however they are significantly distant from *Dt*HNL and they are classified differently in Pfam (pf00407, *Dt*HNL pf10604).

Based on structural analysis 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 relevant for the enzymatic activity as well. We further increased our knowledge on Bet v 1 fold HNLs by enzymatic assays on protein Isotig02275 and its mutants A93S and A92D-E94S. The catalytic residues, Tyr117 and Tyr161 are conserved in Isotig02775, however, they are not sufficient for the activity. Positions 85 and 87 are critical as well (Isotig02775 numbering: 92 and 94). Just restoring an Asp and a Ser as in *Dt*HNL did not lead to a stable expressed protein, therefore, it is not possible to define if the six residues all together are sufficient for HNL activity in Bet v 1 fold proteins. Probably the whole structural environment has to be engineered as well in order to enable correct folding. Interestingly, *Dt*HNL is the only sequence in the entire Bet v 1 superfamily which exhibits the six residues simultaneously. A Glu in position 87 (*Dt*HNL numbering) is strongly conserved in a protein subset created with 3DM[42], where Arg69, Tyr101 and Tyr117 are fixed (**Supplementary Fig. 10g**), and it is likely not compatible with HNL activity in a bet v 1 fold protein. Surprisingly, it appears to be unlikely to identify another protein in the Bet v 1 protein superfamily with HNL activity, based on today's knowledge (**Supplementary Results 10**).

P. aquilinum expresses an HNL (*Pta*HNL), which likely belongs to yet another protein family, as confirmed by the list of putative HNL sequences obtained by our combined activity screening/proteomics/transcriptomics approach (**Supplementary Results 9**; **Appendix, Supplementary Dataset 2**). However, no functionally expressed recombinant enzyme with HNL activity from this fern is available so far.

From the biocatalytic point of view, we discovered promising enzymes for cyanohydrin synthesis. *Dt*HNL isoenzymes are active at low pH conditions, the optimal environment for cyanohydrins. *Dt*HNL1 efficiently converts different aldehydes into the respective cyanohydrins, allowing for short reaction times. Excellent ee can be achieved by keeping reaction time 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, 5 (**Table 1**); whereas engineering of the protein would be necessary for improving the stereoselectivity of *Dt*HNL1 for (*R*)-3-phenylpropionaldehyde cyanohydrin formation.

Finally, the approach for *Dt*HNL discovery can speed up the process to identify additional enzymes with hydroxynitrile lyase function. Several purification steps are avoided, and there is no need of degenerate primers for the amplification of the genes from gDNA or cDNA, which can be an issue for organisms with big genomes.

In summary, *Dt*HNL's novelty and properties open perspectives for the development of a new competitive class of biocatalysts. With our work, we extend the knowledge about HCN release strategies in nature and the diversity of natural functions in the Bet v 1 superfamily. The discovery of such an absolutely new HNL and its new catalytic site for the cleavage and synthesis of cyanohydrins as well as the surprising and seemingly uniqueness in nature raises the question how many further strategies exist for this widespread cyanide release reaction.

Material and Methods

General

The fern *Davallia tyermannii* was purchased in a local shop. *Pteridium aquilinum* leaves and croziers were harvested from local forest (Styria, Austria). All chemicals were purchased from Sigma-Aldrich or Carl Roth GmbH, if not stated otherwise. Racemic mandelonitrile was purchased from abcr GmbH & Co. KG. (*R*)-mandelonitrile was purchased from Sigma-Aldrich or kindly donated by DSM Fine Chemicals Austria. 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 if not specifically reported. 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 in this paper have been submitted to GenBank® database (Supplementary Table 13).

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: 80V, 40 min). Normalized transcriptome sequencing was obtained by the commercial service from Microsynth AG as following: 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 paper[43] in a 384-well plate in 100 mM citrate buffer pH 4.0 and 3 mM racemic mandelonitrile as substrate. The mixture was incubated for 20 minutes. 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 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 protrein gel

and HNL activity assay was performed after the electrophoretic run. The procedure was performed similar to that described previously[32]. Afterwards, the gel was stained by silver staining as described[44] with the following modifications: fix step over-night and 30 incubation in water after the first wash ethanol 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. Peptide extracts were 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. It was analyzed in a Thermo LTQ-FT mass spectrometer operated in positive ion mode, applying alternating full scan MS (m/z 400 to 2000, 50000 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 max. 10 ppm precursor mass error required for protein identification). Hits were subjected to BLAST alignment against NCBI non-redundant public protein database.

Isoenzyme 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 10) and genes were amplified by PCR. PCR products were isolated from a 1% agarose gel after electrophoresis and sequenced (LGC Genomics GmbH). Results obtained by Sanger and Roche/454 sequencing were compared and corrected in case of inconsistency.

Cloning

The entire list of primers, gBlocks® Gene Fragments and synthetic genes is reported in Supplementary Tables 10-12. Different electrocompetent *E. coli* strains were transformed by electroporation and positive clones were selected on LB agar plates with the appropriate antibiotic reported below. *E. coli* TOP 10 F' 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 *Dt*HNL1-4.

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 (*Dt*HNL), restriction sites NdeI/HindIII. Clones were selected on ampicillin (100 mg/L).

DtHNL1 expression. A Synthetic gene of isotig 02643 was cloned into the pEHISTEV[45] vector utilizing restriction sites NcoI/HindIII, in order to add the His-TEV tag at the N-term 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-term of the protein. Clones were selected on kanamycin (50 mg/L). Sequences were verified by Sanger sequencing (LGC Genomics).

DtHNL1 and PtaIso02775 mutants. Optimized nucleotide sequences coding for parts of the protein and containing the desired mutation(s) was purchased as gBlocks® Gene Fragments (Integrated DNA Technologies). pEHISTEV containing *Dt*HNL1 or *Pta*Iso02775 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 verified 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 OD_{600} 0.7. Afterwards the culture was incubated 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) disrupted by sonication (80% duty cycle, 7 output, 6 minutes).

Protein purification

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 *Dt*HNL 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 150mM NaCl, 10 mM Tris-HCl pH 8. The protein (0.5 ml; 1mg/ml in 10mM Tris-HCl pH8) 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 *Dt*HNL activity

Standard assay. The enzymatic activity was quantified in a similar manner as described[46]. 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 substrate, previously dissolved in 3 mM sodium citrate-phosphate buffer pH 3.5 (60 mM fresh stock solution). Control reactions contained storage buffer instead of protein. Benzaldehyde formation was detected at 280 nm for 10 minutes. One unit is defined as

the amount of enzyme that catalyzes the formation of 1 µmol of benzaldehyde in 1 minute. Each experiment was performed in biological and technical triplicate unless otherwise stated.

Activity at different pH values was performed in 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 as final volume and 0.01 µg *Dt*HNL. SigmaPlot™ was used for calculation of K_m and $V_{max.}$

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 minutes.

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 through a Feigl–Anger test paper[43]. 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-*Dt*HNL1 expression and purification

E. coli BL21 (DE3) Star expressing tagged *Dt*HNL1 were cultivated in a shake flask in minimal medium (M9 salts 5X, 2% (w/v) glucose, 2 mM $MgSO_4$, 0.01 mg/mL thiamine, 0.01 mg/mL $FeCl_3$) supplemented with 50 mg/L kanamycin and 50 mg/L selenomethionine. Induction was performed at OD_{600} 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-*Dt*HNL1 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-*Dt*HNL1 were combined and desalted (PD10 Desalting columns, GE Healthcare). Protein was stored at -20°C in 50 mM potassium phosphate buffer pH 6.0 prior to crystallization.

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 crystal conditions was performed manually by the sitting drop vapor-diffusion method in Crystal Clear Duo crystallization frames at 16° C.

Native crystals of *Dt*HNL1 were obtained by mixing 0.5µl 4 mg/mL protein sample (in 10 mM Tris-HCl pH 8.0) with 1 µl reservoir solution (0.9 M $NaNO_3$; Na_2HPO_4; $(NH_4)_2SO_4$ 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 0.5µl 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-*Dt*HNL1 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 *Dt*HNL1 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 package[47] or iMosflm[48] / SCALA[49]. The AutoSol Program[50,51] –and the AutoBuild Program[52] from the PHENIX software suit[53] were used to define the selenium heavy metal-atom sites using a SeMet-*Dt*HNL1 SAD data set, as well as to build an initial model. The resulting model was completed manually in Coot[54] 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-MR[50]. The previously obtained SeMet-*Dt*HNL 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 Molprobity[55]. Detailed data processing and structure refinement statistics are summarized in Supplementary Table 7.1. The atomic coordinates and structure factors have been deposited in the Protein Data Bank under the accession codes 5E46 (*Dt*HNL1 SeMet), 5E4B (DtHNL1-MXN), 5E4D (DtHNL1–BEZ), 5E4M (*Dt*HNL1-HBA).

Cyanohydrin Synthesis

Synthesis of different cyanohydrins was carried out in a biphasic system as described in detail by Wiedner *et al.*[46] except for 1-phenylethanone which was used in 0.3 M concentration. The aqueous phase contained 3 mg of purified *Dt*HNL1. After acetylation, samples were analyzed by isothermal GC at 110°C for 20 minutes. 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 transformation) was set up in the same conditions, with buffer instead of the enzyme solution. Apparent kinetic parameters were determined with 0.05 mg of *Dt*HNL1 (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 minutes. 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-H_2O of benzaldehyde. Each reaction was performed in duplicate as two different independent experiments. SigmaPlot™ was employed for determination of K_m and v_{max}.

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 (BMWFW) and the Federal Ministry of Science, Research and Economy (BMWFW), 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.

Authors Contribution

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; B.D. designed and performed mass spectrometry analysis; B.D., R.B.G and collected proteomics data; E.L. B.D. and R.B.G. analyzed proteomics data, G.G.T. designed transcriptome assembling, J.H. performed transcriptome assembling, H.J.J. and T.V.D.B. built 3DM, E.L. collected and analyzed bioinformatics 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., M.W., T.P.K., K.G. and A.G. wrote the manuscript.

References

1. 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).
2. Dadashipour, M. & Asano, Y. Hydroxynitrile lyases: insights into biochemistry, discovery, and engineering. *ACS Catal.* **1,** 1121–1149 (2011).
3. Sharma, M., Sharma, N. N. & Bhalla, T. C. Hydroxynitrile lyases: at the interface of biology and chemistry. *Enzyme Microb. Technol.* **37,** 279–294 (2005).
4. Hussain, Z. *et al.* Characterization of two bacterial hydroxynitrile lyases with high similarity to cupin superfamily proteins. *Appl. Environ. Microbiol.* **78,** 2053–2055 (2012).
5. 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).
6. 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.* 201508311 (2015).
7. Andexer, J. *et al.* An R-selective hydroxynitrile lyase from *Arabidopsis thaliana* with an α/β-hydrolase fold. *Angew. Chemie - Int. Ed.* **46,** 8679–8681 (2007).
8. Gruber-Khadjawi, M., Fechter, M. H. & Griengl, H. *Enzyme Catalysis in Organic Synthesis*. *Enzyme Catalysis in Organic Synthesis, Third Edition* **2,** (Wiley-VCH Verlag GmbH & Co. KGaA, 2012).
9. Lanfranchi, E. *et al.* Mini-review: recent developments in hydroxynitrile lyases for industrial biotechnology. *Recent Pat. Biotechnol.* **7,** 197–206 (2013).
10. Purkarthofer, 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).
11. Winkler, M., Glieder, A. & Steiner, K. in *Comprehensive Chirality* (eds. Carreira, E. M. & Yamamoto, H.) **7,** 350–371 (2012).
12. Glieder, A. *et al.* Comprehensive step-by-step engineering of an (*R*)-hydroxynitrile lyase for large-scale asymmetric synthesis. *Angew. Chem. Int. Ed. Engl.* **42,** 4815–8

(2003).

13. Hanefeld, U. Immobilisation of hydroxynitrile lyases. *Chem. Soc. Rev.* **42,** 6308–21 (2013).
14. Pscheidt, B. *et al.* Efficient biocatalytic synthesis of (*R*)-pantolactone. *Adv. Synth. Catal.* **350,** 1943–1948 (2008).
15. Sano, Y. A. *et al.* Screening for New Hydroxynitrilases from Plants. **69,** 2349–2357 (2005).
16. 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).
17. 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).
18. Hickel, A., Heinrich, G., Schwab, H. & Griengl, H. Screening for hydroxynitrile lyases in plants. **11,** 55–58 (1997).
19. 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).
20. Hajnal, I. *et al.* Biochemical and structural characterization of a novel bacterial manganese-dependent hydroxynitrile lyase. *FEBS J.* **280,** 5815–5828 (2013).
21. Dreveny, I., Gruber, K., Glieder, A., Thompson, A. & Kratky, C. The hydroxynitrile lyase from almond: a lyase that looks like an oxidoreductase. *Structure* **9,** 803–15 (2001).
22. 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).
23. Guterl, J. K. *et al.* Uneven twins: Comparison of two enantiocomplementary hydroxynitrile lyases with α/β-hydrolase fold. *J. Biotechnol.* **141,** 166–173 (2009).
24. 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).

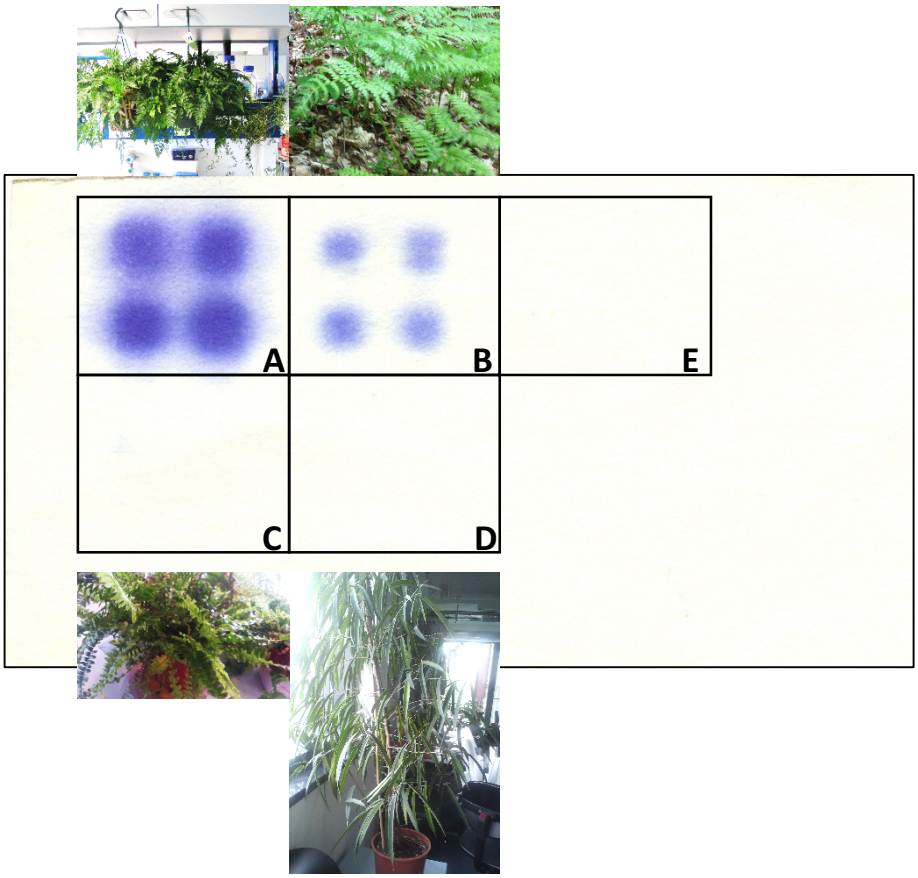
25. Trummler, K., Fo, S., Pfizenmaier, K. & Wajant, H. Expression of the Zn^{2+} - containing hydroxynitrile lyase from flax (*Linum usitatissimum*) in *Pichia pastoris* — utilization of the recombinant enzyme for enzymatic analysis and site-directed mutagenesis. **139,** 19–27 (1998).
26. Alagöz, D., 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).
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. 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).
29. Kofod, H. & Fson, R. E. Y. J. Cyanogenesis in Species of the Fern Genera *Cystopteris* and *Davallia*. **8,** 1509–1511 (1969).
30. 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).
31. Altschul, S. Basic Local Alignment Search Tool. *J. Mol. Biol.* **215,** 403–410 (1990).
32. Lanfranchi, E. *et al.* Bioprospecting for hydroxynitrile lyases by Blue Native PAGE Coupled HCN Detection. *Current Biotech*. **4**,111–117 (2015).
33. Radauer, C., Lackner, P. & Breiteneder, H. The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol. Biol.* **8,** 286 (2008).
34. 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).
35. 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).

36. 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).
37. 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**, 1611-1615 (2015).
38. 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**, 5220-5227 (2006).
39. Zimmer, T., Ohkuma, M., Ohta, A., Takagi, M. & Schunck, W. H. The CYP52 multigene family of *Candida maltosa* encodes functionally diverse n-alkane-inducibile cytochromes P450. *Biochm. Biophys. Res. Commun*. **224**, 784-789 (1996).
40. Alonso, H. et al. Structural and mechanistic insight into alkane hydroxylation by *Pseudomonas putida* AlkB. *Biochem. J*. **460**, 283-293 (2014).
41. Hofrichter, M., Kellner, H, Pecyna, M.J. & Ullrich, R. Fungal unspecific peroxygenases: heme-thiolate proteins that combine peroxidase and cytocrome p450 properties. *Adv. Exp. Med*. *Biol*. **851**, 341-368 (2015).
42. Kuipers, R. K. *et al.* 3DM: systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins* **78,** 2101–13 (2010).
43. Krammer, B., Rumbold, K., Tschemmernegg, M., Pöchlauer, P. & Schwab, H. A novel screening assay for hydroxynitrile lyases suitable for high-throughput screening. *J. Biotechnol.* **129,** 151–61 (2007).
44. Blum, H. *et al.* Improved silver staining of plant-proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* **8,** 93–99 (1987).
45. 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).
46. Wiedner, R. *et al.* Improving the properties of bacterial *R*-selective hydroxynitrile lyases for industrial applications. *ChemCatChem* **7,** 325–332 (2015).
47. Kabsch, W. Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 125–132 (2010).
48. 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).

49. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62,** 72–82 (2006).
50. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
51. 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).
52. 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).
53. 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).
54. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).
55. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 12–21 (2010).

Supplementary Information

Supplementary Result 1. Screening for HNL activity in cyanogenic ferns



Supplementary Figure 1a. **HNL activity screening in different plants.**

Two cyanogenic ferns *Davallia tyermannii* and *Pteridium aquilinum* were screened for hydroxynitrile lyase (HNL) activity. Fresh leaves were disrupted and the resultant protein extract was mixed with 100 mM citrate buffer pH 4.5 containing racemic mandelonitrile. Cyanide release was detected via a Feigl–Anger test paper[1].

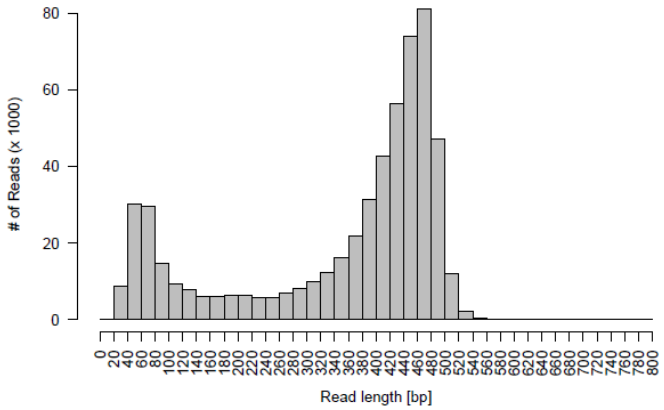
Davallia tyermannii (A); *Pteridium aquilinum* (B). Negative controls: Non cyanogenic fern from *Nephrolepsis* genus (C); Non cyanogenic plant from *Ficus* genus (D). Blank: 100 mM citrate-phosphate buffer pH 4.5 (E).

Supplementary Result 2. Transcriptome sequencing and assembling

Poly A mRNA was isolated from fresh leaves of the two ferns species *P. aquilinum* and *D. tyermannii*. The major results are summarized in the following tables and figures:

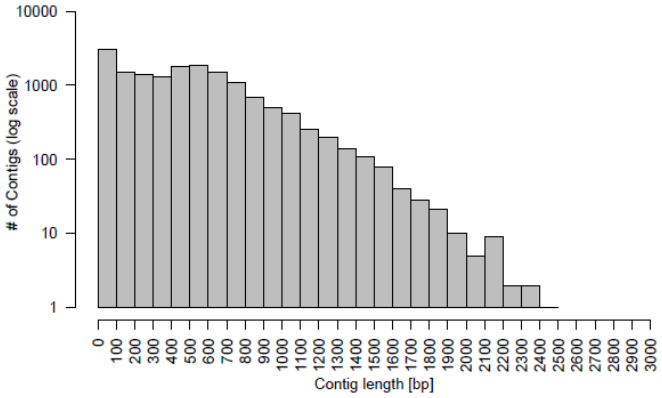
Summary of transcriptome sequencing and assembly.

	P. aquilinum		*D. tyermannii*	
	Number	**% of total**	**Number**	**% of total**
Total number of reads	834,642	-	560,161	-
Reads after Newbler quality control	828,772	100.0	557,396	100.0
Reads aligned	657,753	79.4	447,265	80.2
Reads assembled	526,742	53.6	291,066	52.2
Reads partially assembled	101,168	12.2	155,950	28.0
Singletons	69,936	8.4	48,729	8.7
Contigs	78,726	-	15,964	-
Isogroups	18,357	-	7,792	-
Isotigs	48,207	-	11,497	-
Average isotig length	568.6		808.5	-
Median isotig length	514.0	-	737.0	-
Largest isotig length	2,019	-	2,772	-



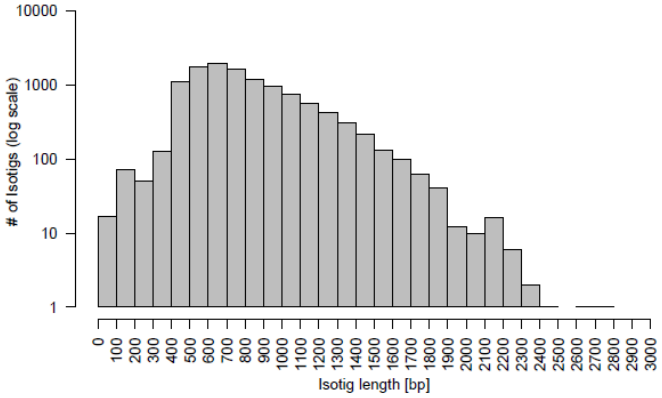
Supplementary Figure 2a. *D. tyermannii* transcriptome read length distribution.

Total reads: 560161; average/median length: 350.5/418.0.



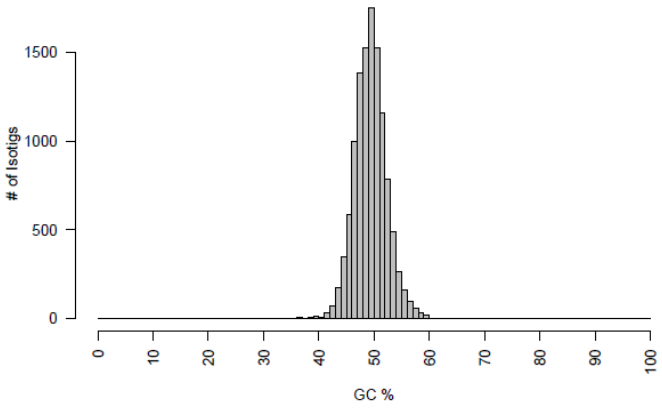
Supplementary Figure 2b. *D. tyermannii* transcriptome contig length distribution.

Total contigs: 15964; average/median length: 466.0/454.0.



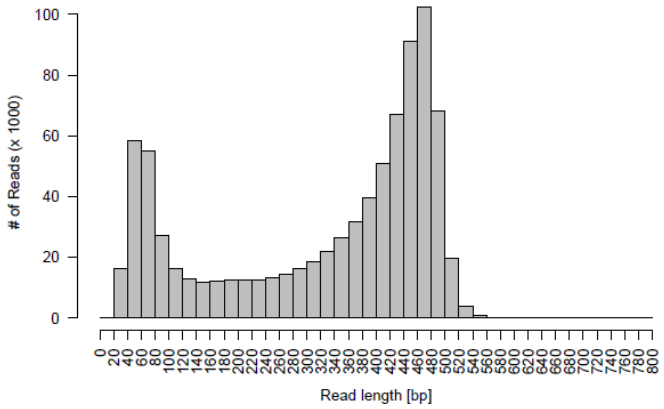
Supplementary Figure 2c. *D. tyermannii* transcriptome isotig length distribution.

Total isotigs 11497; average/median length: 808.5/737.0.



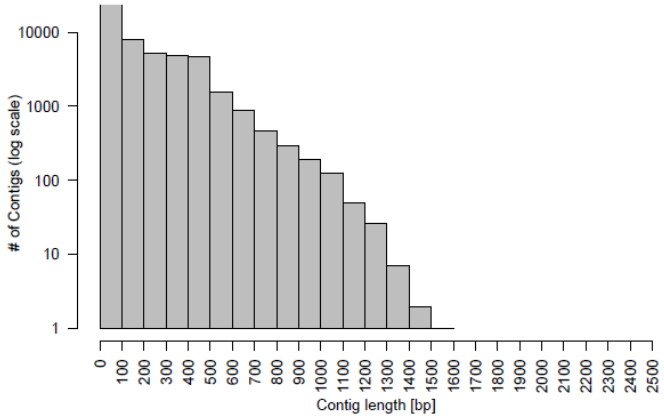
Supplementary Figure 2d. *D. tyermannii* transcriptome isotig GC content distribution.

Total isotigs: 11497; average/median: %: 49.4/49.4.



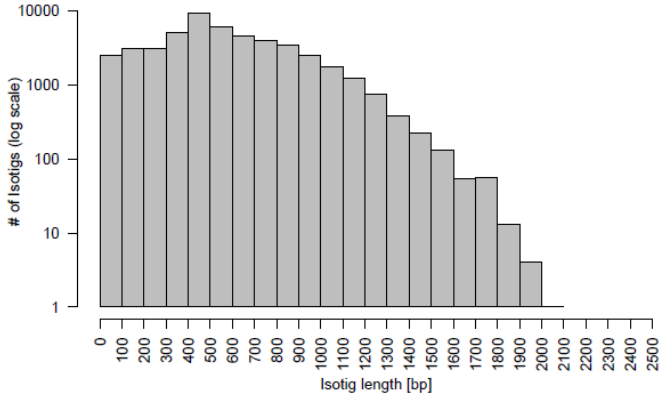
Supplementary Figure 2e. *P. aquilinum* transcriptome read length distribution.

Total reads: 834642: average/median length: 325.6/395.0.



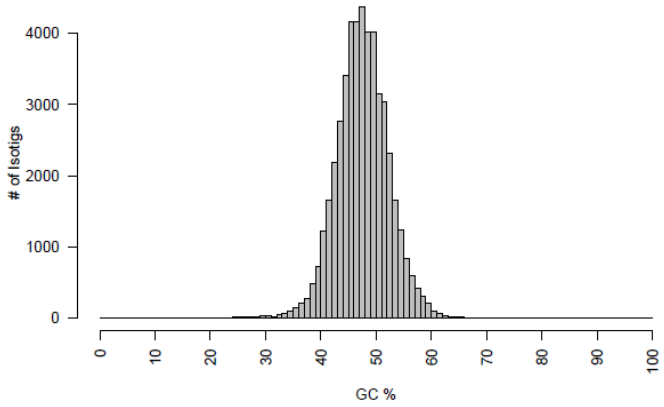
Supplementary Figure 2f. *P. aquilinum* transcriptome contig length distribution.

Total contigs 78726; average/median length: 129.8/40.0.



Supplementary Figure 2g. *P. aquilinum* transcriptome isotig length distribution.

Total isotigs: 48207: average/median: 568.6/514.0.



Supplementary Figure 2h. ***P. aquilinum*** **transcriptome isotig GC content distribution.**

Total isotigs: 48207; average/median %: 47.6/47.5.

Supplementary Result 3. *In silico* HNL search in fern transcriptomes

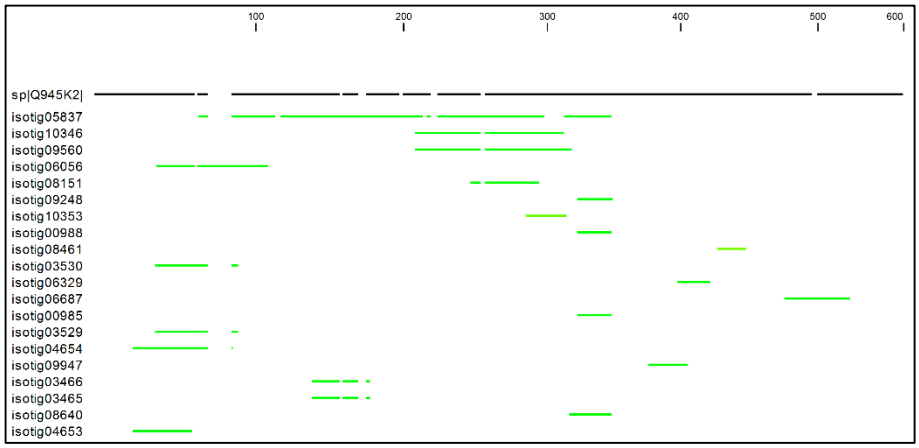
The hydroxynitrile lyase from *Phelobodium aureum* was described as an (*R*)-selective non-flavin protein and not related to any known HNL, when the publication was released two decades ago[2]. These results were based on biochemical properties of the protein and up to today no indication of the protein sequence has been reported. The number of known HNLs has been increasing in the last years and new HNL classes and sequences were published. Therefore, we subjected 10 known HNL sequences to a tblastn search in both transcriptomes of *D. tyermannii* and *P. aquilinum* using CLC Main Workbench 7.6 (QIAGEN Aarhus A/S), with the default parameters reported by the software (program: tblastn; expectation value: 100; word size: 3; mask lower case: no; filter low complexity: yes; maximum number of hits: 500; protein matrix and gap costs: BLOSUM62 existence 11 extension 1; number of threads: 1; genetic code: 1). Sequence queries are summarized in Supplementary Table 1).

All hits show low sequence identity, as indicated by the color code (green is less than 50%) and/or low sequence coverage and no similar sequence can be found (Supplementary Fig. 3a-3t). Results indicate that indeed a new HNL type(s) exist in ferns, which confirms the previous claims of Wajant and coworkerkers[2].

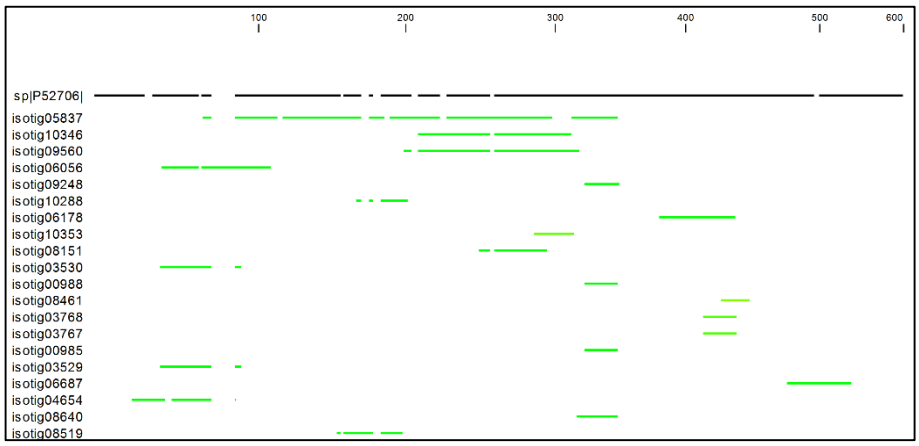
Supplementary Table 1. List of HNLs employed as sequence queries.

Entry	UniProtKB	ID	Organism	Protein Family (pfam database)
1	Q95K2	*Pa*HNL	*Prunus amygdalus*	GMC oxidoreductase
2	P52706	*Ps*HNL	*Prunus serotina*	GMC oxidoreductase
3	B7YF77	*Ej*HNL	*Eriobotrya japonica*	GMC oxidoreductase
4	Q9LFT6	*At*HNL	*Arabidopsis thaliana*	Alpha/beta hydrolase
5	P52704	*Hb*HNL	*Hevea brasiliensis*	Alpha/beta hydrolase
6	P52705	*Me*HNL	*Manihot esculenta*	Alpha/beta hydrolase
7	D1MX73	*Bm*HNL	*Baliospermum montanum*	Alpha/beta hydrolase
8	E8WN5	*Gt*HNL	*Granulicella tundricola*	Cupin
9	P52708	*Sb*HNL	*Sorghum bicolor*	Peptidase S10
10	P93243	*Lu*HNL	*Linum usitatissimum*	Zinc binding alcohol dehydrogenase

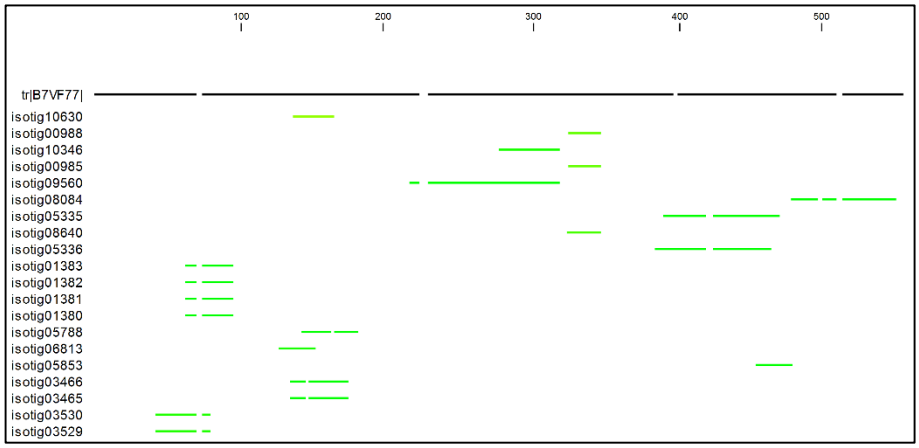
Supplementary Figures 3a–3t. Qualitative visualization of the tblatn results for each query in *D. tyermannii* (**Supplementary Fig. 3a–3j**) and *P. aquilinum* (**Supplementary Fig. 3k–3t**) transcriptomes. First 20 hits are depicted. **0%**  **100% Identity**



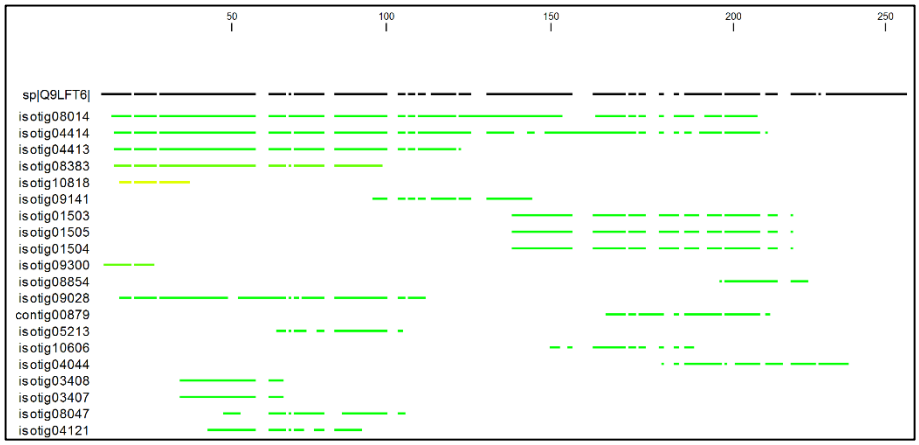
Supplementary Figure 3a. **tblastn *Pa*HNL – *D. tyermannii* transcriptome.**



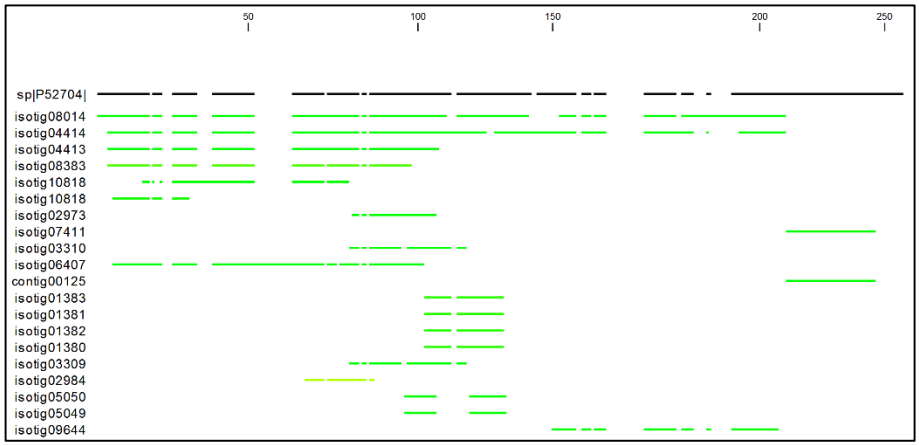
Supplementary Figure 3b. **tblastn *Ps*HNL – *D. tyermannii* transcriptome.**



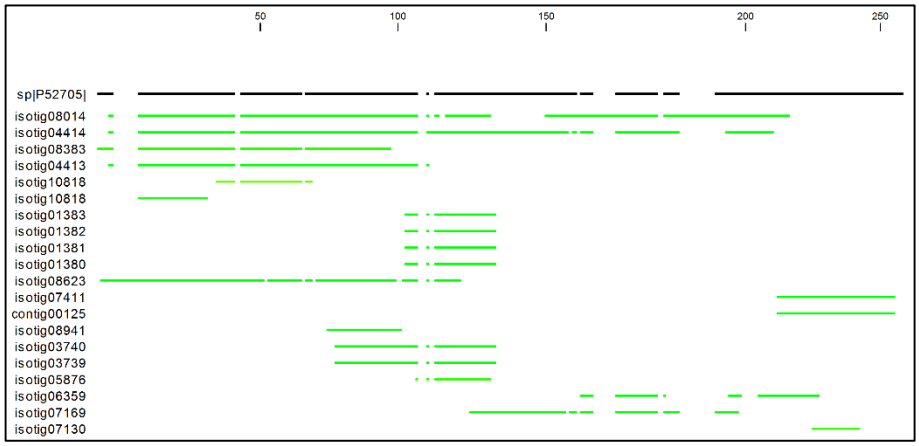
Supplementary Figure 3c. **tblastn *Ej*HNL – *D. tyermannii* transcriptome.**



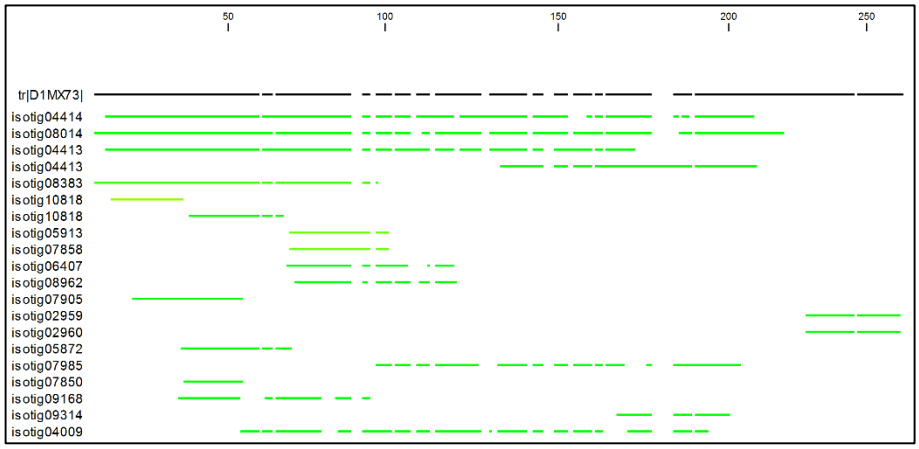
Supplementary Figure 3d. **tblastn *At*HNL – *D. tyermannii* transcriptome.**



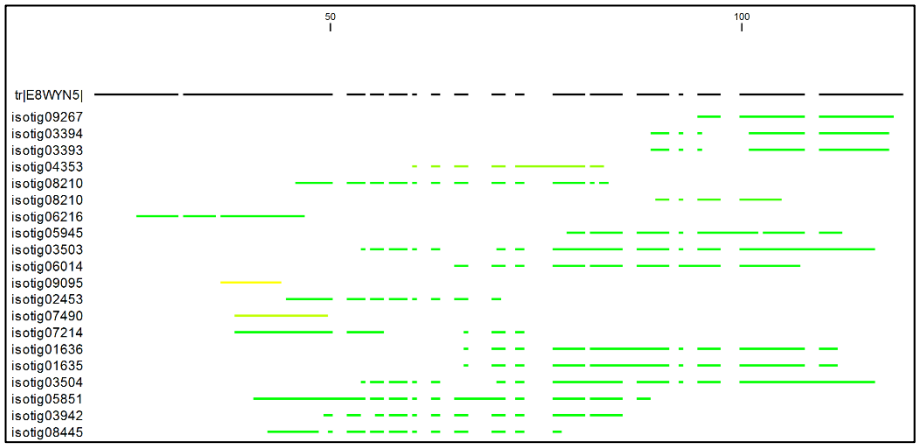
Supplementary Figure 3e. tblastn *Hb*HNL – *D. tyermannii* transcriptome.



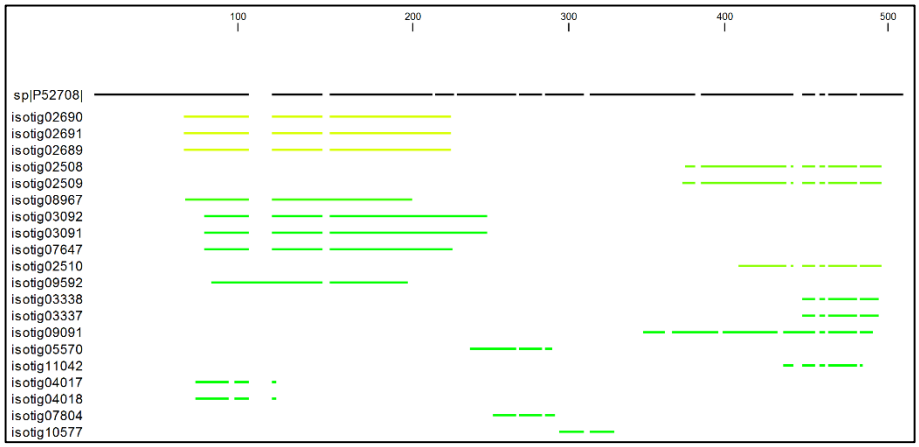
Supplementary Figure 3f. tblastn *Me*HNL – *D. tyermannii* transcriptome.



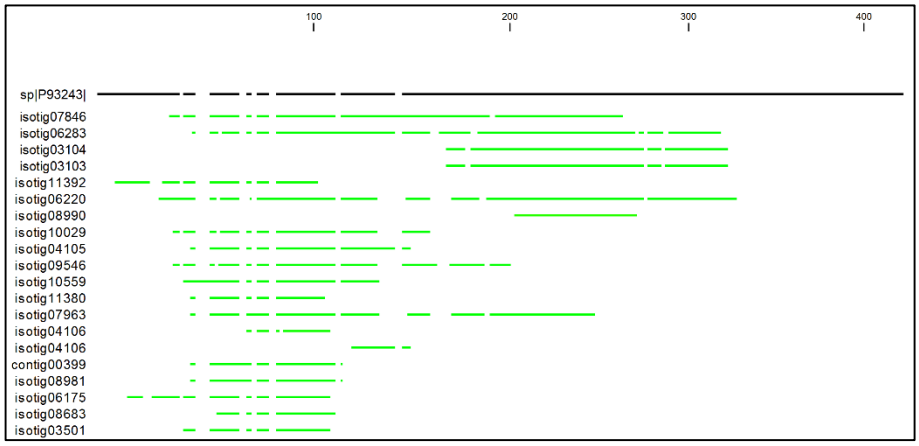
Supplementary Figure 3g. **tblastn *Bm*HNL – *D. tyermannii* transcriptome.**



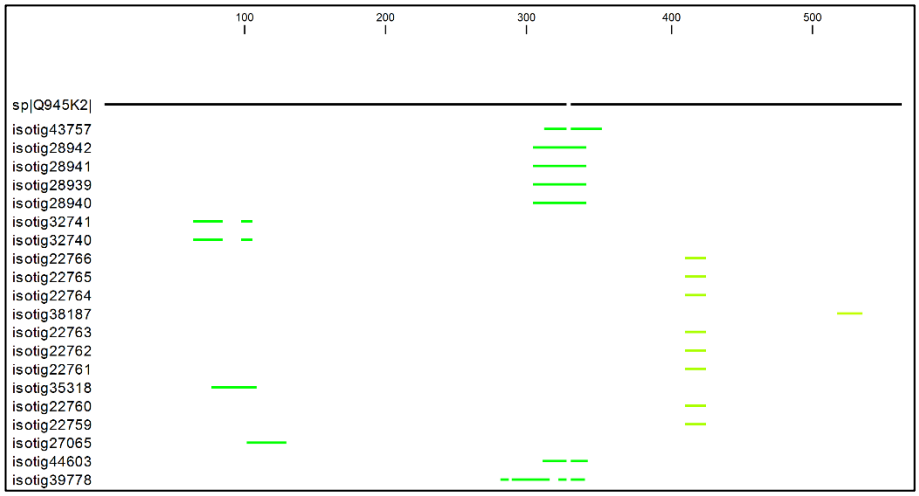
Supplementary Figure 3h. **tblastn *Gt*HNL – *D. tyermannii* transcriptome.**



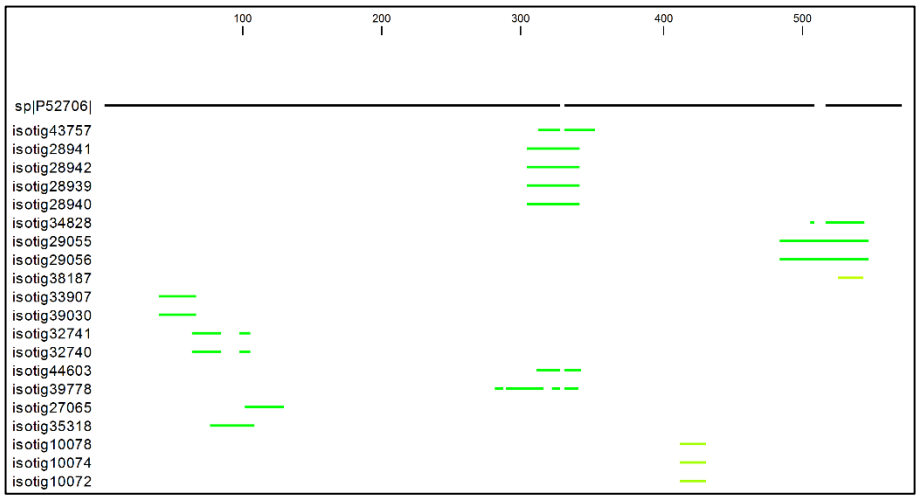
Supplementary Figure 3i. **tblastn *Sb*HNL – *D. tyermannii* transcriptome.**



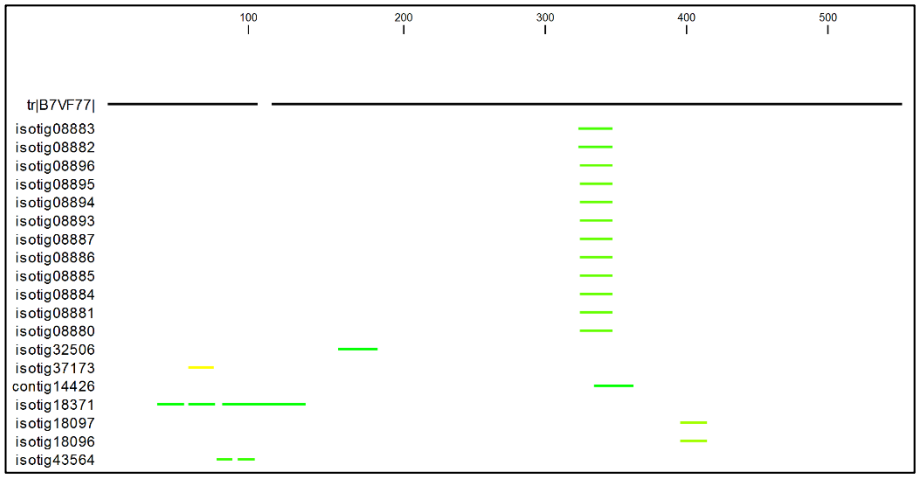
Supplementary Figure 3j. tblastn *Lu*HNL – *D. tyermannii* transcriptome.



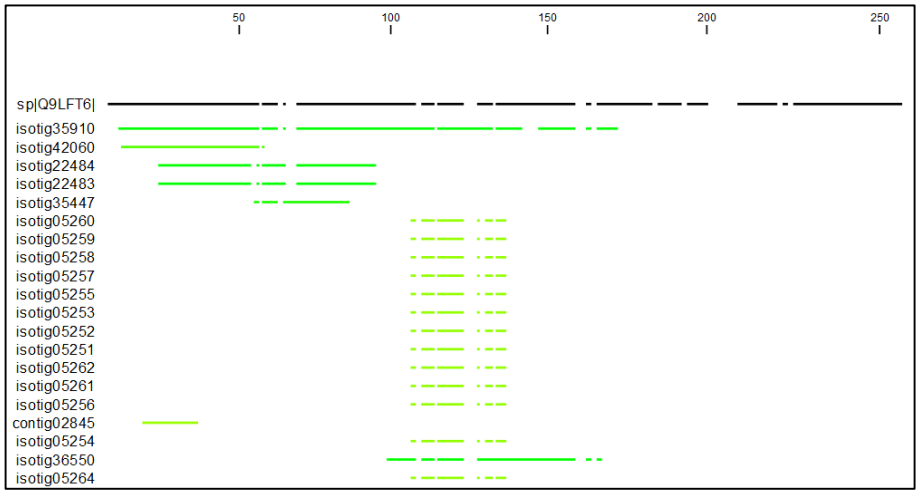
Supplementary Figure 3k. tblastn *Pa*HNL – *P. aquilinum* transcriptome.



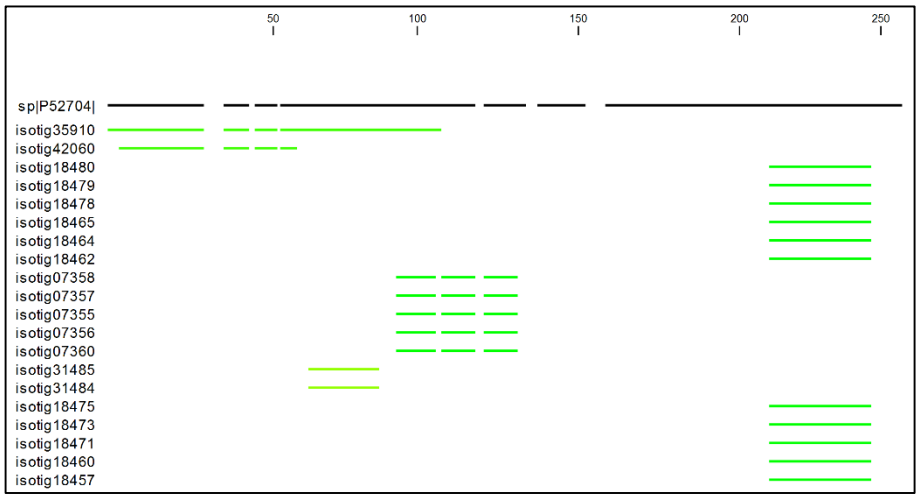
Supplementary Figure 3l. tblastn *Ps*HNL – *P. aquilinum* transcriptome.



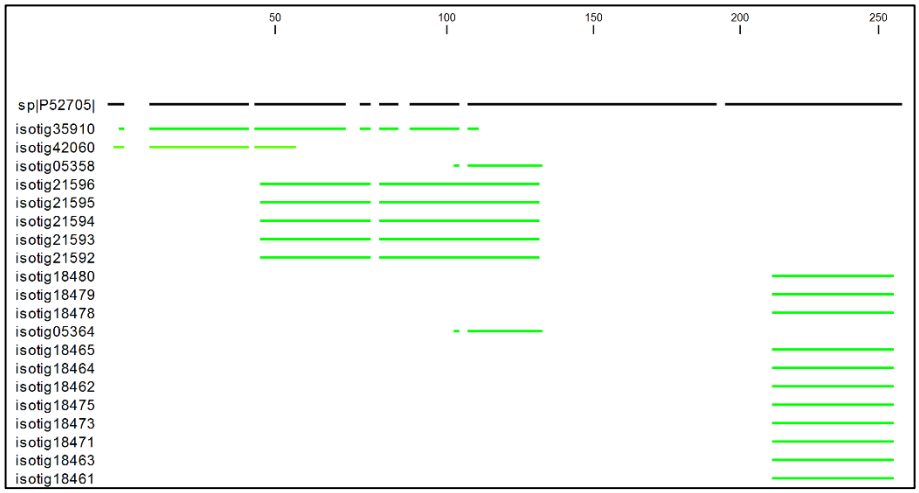
Supplementary Figure 3m. **tblastn *Ej*HNL – *P. aquilinum* transcriptome.**



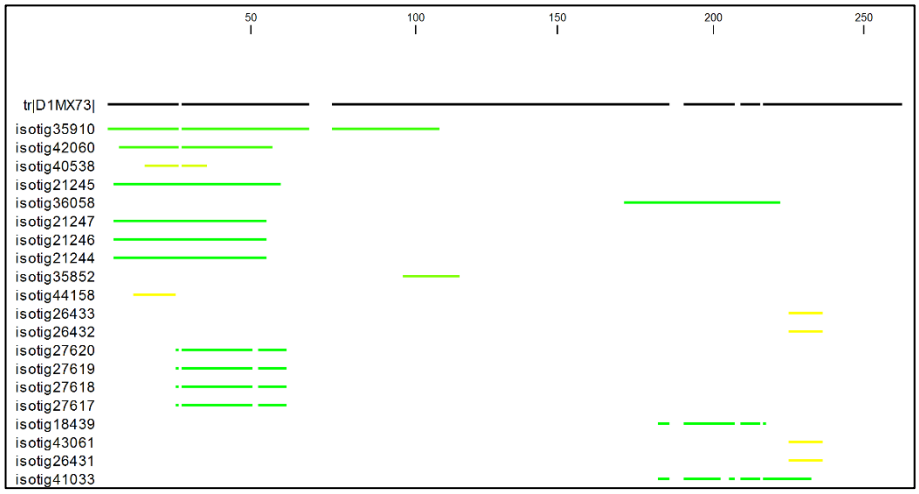
Supplementary Figure 3n. **tblastn AtHNL – *P. aquilinum* transcriptome.**



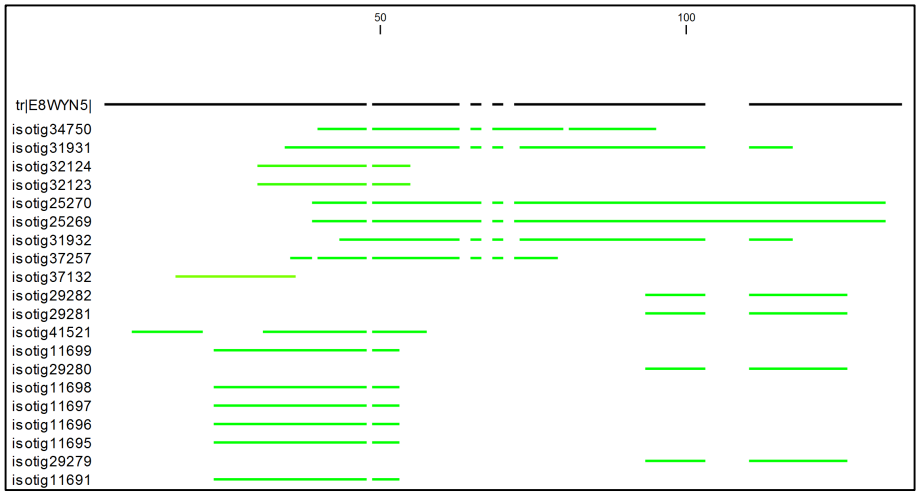
Supplementary Figure 3o. **tblastn *Hb*HNL – *P. aquilinum* transcriptome.**



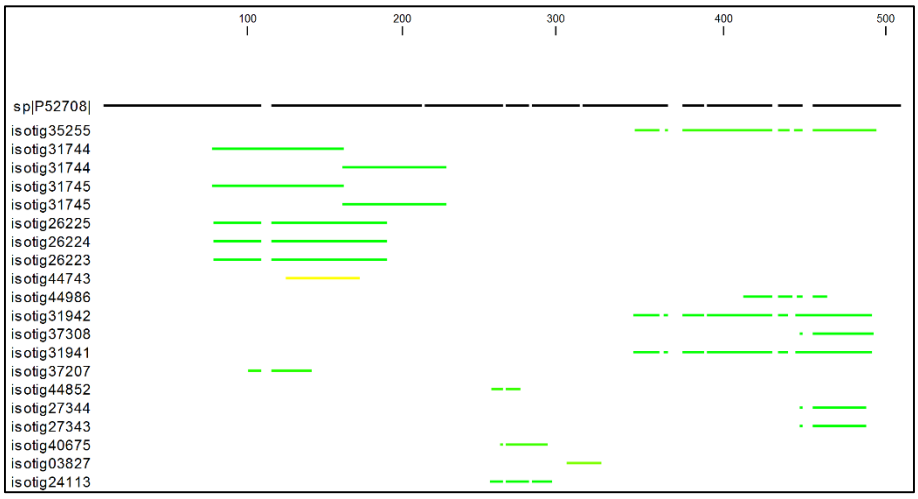
Supplementary Figure 3p. **tblastn *Me*HNL – *P. aquilinum* transcriptome.**



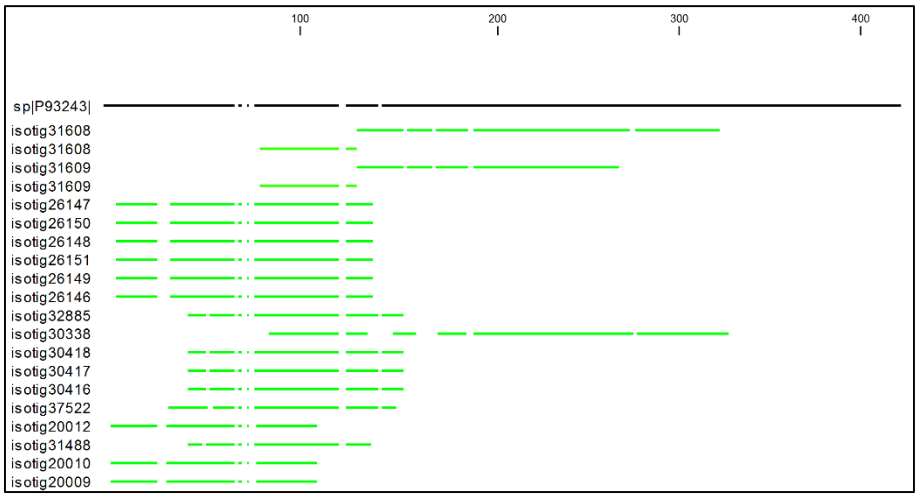
Supplementary Figure 3q. **tblastn *Bm*HNL – *P. aquilinum* transcriptome.**



Supplementary Figure 3r. tblastn *Gt*HNL – *P. aquilinum* transcriptome.



Supplementary Figure 3s. tblastn *Sb*HNL – *P. aquilinum* transcriptome.



Supplementary Figure 3t. tblastn *Lu*HNL – *P. aquilinum* transcriptome.

Supplementary Result 4. Mass Spectrometry

Supplementary Table 2. Summary of the proteins screened for HNL activity

Entry		1	2	3	4	5	6
ID		**Isotig 02643**	**Isotig 06604**	**Isotig 07200**	**Contig 00505**	**Isotig 04065**	**Isotig 04379**
Family members		Isotig 02641 Isotig 07602 Contig 00751			Contig 00644 Contig 00096	Isotig 04066 Isotig 07043	Isotig 04380
MS/MS	Σ# Unique Peptides	5	8	3	1	2	2
	Σ# PSMs	32	13	3	8	4	3
Blastp (ncbi)	Best Hit	XP_ 009405224.1 PREDICTED: Lachrymatory factor synthase like	NP_ 001051733.1 Os03g0822200	XP_ 0068559057.2 PREDICTED: plasma membrane associated cation binding protein 1	XP_ 010555234.1 PREDICTED: thaumatin-like protein 1b	XP_ 002313728.1 Disease resistance responsive family protein	XP_ 002971933.1 Hypothetical protein SELMODRAFT 270941
	Identity	28	77	43	35	48	49
	Query coverage	72	100	100	96	55	93
	E value	8e-12	6e-141	1e-35	1e-20	6e-19	2e-68
pfam		Polyketide Cyc2	NAD binding 10	DREPP	Thaumatin	Dirigent	Thioredoxin 4

Supplementary Result 5. *Dt*HNL isoenzymes

Based on the mRNA sequence derived from the transcriptome data genomic DNA was amplified from isolated fern DNA and the PCR product was directly sequenced by Sanger sequencing. Except for *Dt*HNL2 the derived amino acid sequences were identical. The observed difference in case of *Dt*HNL2 might be a mistake originating from error prone reverse transcription or sequencing mistakes during transcriptome sequencing.

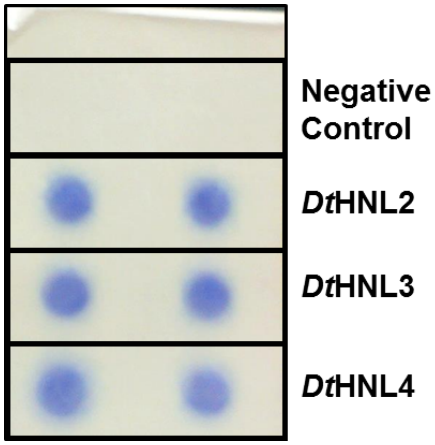
Comparison of the *Dt*HNL isoenzyme transcriptome and genome sequences.

***Dt*HNL**	**Transcriptome**	**Genome**
	aa	**aa**
*Dt*HNL1	Identical	
*Dt*HNL2	GGVSIF (130-136)	GGGVIF (130-136)
*Dt*HNL3	Identical	
*Dt*HNL4	Identical	

```
DtHNL1          MAGTGGGAEQFQLRGVLWGKAYSWKITGTTIDKVWSIVGDYVRVDNWVSSVVKSSHVVSG 60
DtHNL2          MAGTRGGAEEFQLRGVLWGKAYSWKITGTTIDKVWSIVGDYVRVDNWVSSVVKSSHVVSG 60
DtHNL3          MAGTGGGAEEFQLRGVLWGKAYSWKIRGTTIDKVWAIVGDYVRVDNWVSSVVKSSHVVSG 60
DtHNL4          MAGTGGGAEEFQLRGVLWGKAYSWKITGTTIDKVWSIVGDYVRVDNWVSSVVKSSHVVSG 60
                **** ****:****************:********:************************

DtHNL1          EANCTGCVRRFVCYPASEGESETVDYSELIHMNAAAHQYMYMIVGGNITGFSLMKNYVSN 120
DtHNL2          DANCTGCVRRFVCYPASKGESETVDYSELIHMNAAAHQYMYMIVGGNITGFSLMKNYVSN 120
DtHNL3          DANKTGCVRRFVCYPASEGESETVDYSELIHMNAAAHQYMYMIVGGNITGFSLMKNYVSN 120
DtHNL4          DANKTGCVRRFVCYPASEGESETVDYSELIHMNAAAHQYMYMIVGGNITGFSLMKNYVSN 120
                :**:*************:****************************************

DtHNL1          ISLSSLPEEDGGGVIFYWSFTAEPASNLTEQKCIEIVFPLYTTALKDLCTHLSIPESSVT 180
DtHNL2          ISLSSLPEEDGGGVIFYWSFTAEPASNLTEQKCIEIVFPLYTTALKDLCTHLSIPESSVT 180
DtHNL3          ISLNSLPEADGGGVILHWSFTAEPASNLTEQKCIEIVFPLYTTALKDLCTHLSIPESSVT 180
DtHNL4          ISLNSLPEADGGGVIFHWSFTAEPASNLTEQKCIEIVFPLYTTALKDLCTHLSIPESSVT 180
                ***.**** ******::*******************************************

DtHNL1          LLDD 184
DtHNL2          LLDD 184
DtHNL3          LLGD 184
DtHNL4          LLGD 184
                  **.*
```

Supplementary Figure 5a. **Multiple sequence alignment of *Dt*HNL isoenzymes.**

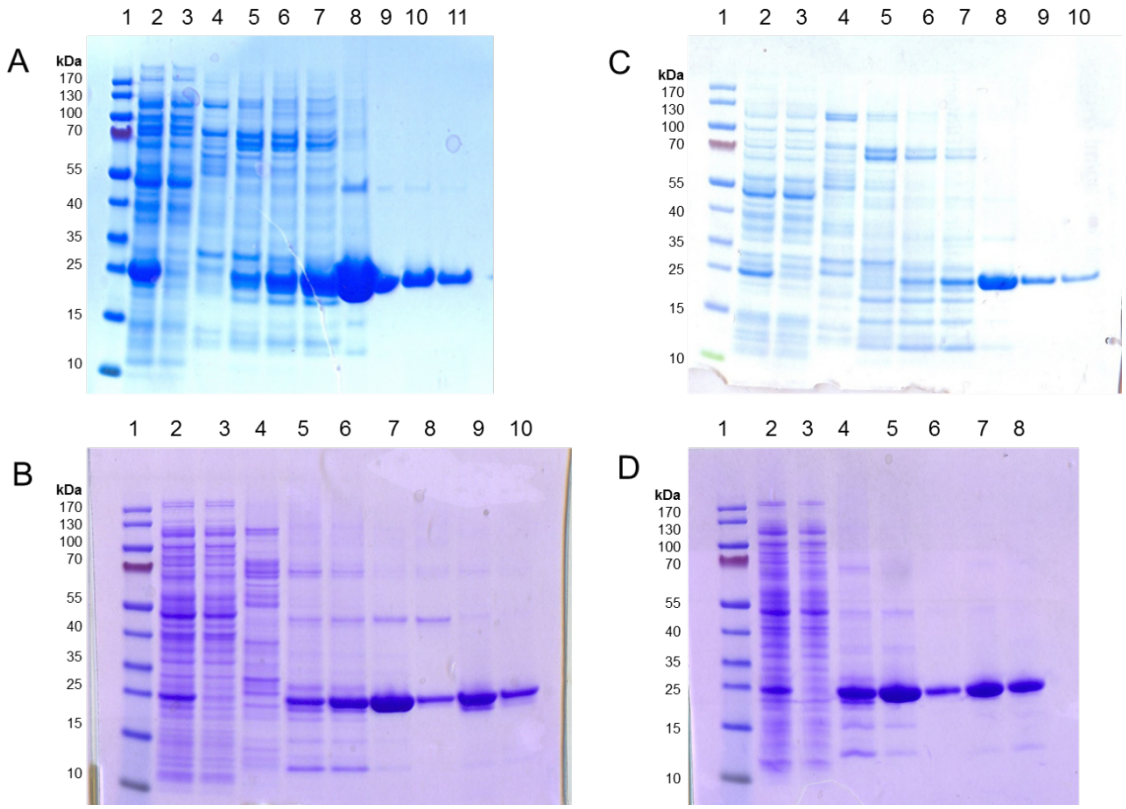
Highlighted residues: differences (red); conserved residues in *Dt*HNL1 and *Dt*HNL2 only (magenta); conserved residues in *Dt*HNL3 and *Dt*HNL4 only (green). Protein sequence alignment was performed with Clustal Omega[3].



Supplementary Figure 5b. HNL activity assay for *Dt*HNL2, 3 and 4.

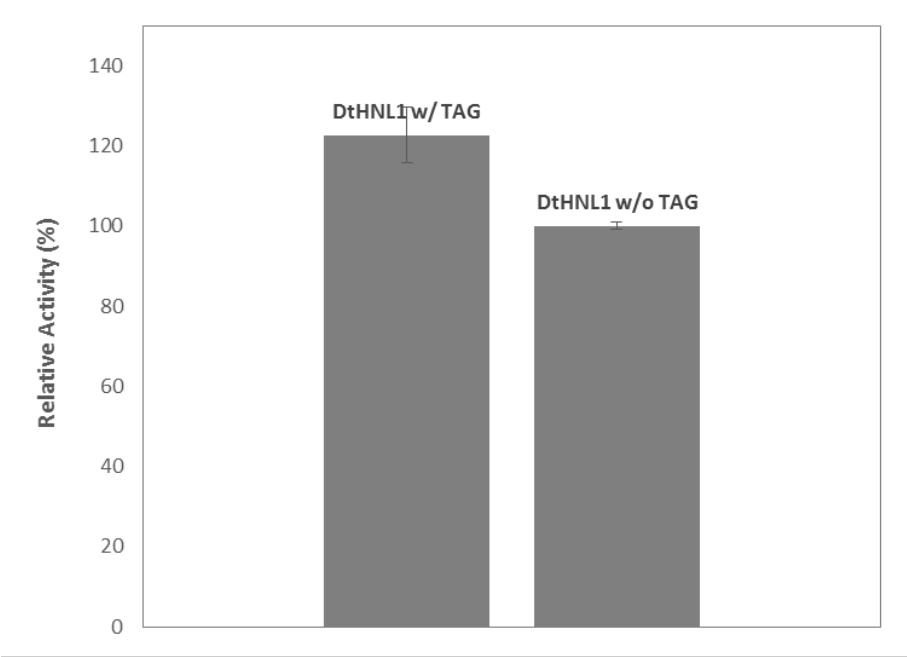
Each protein was recombinant expressed in *E. coli* with standard cultivation protocol as described in Material and Methods. Cells were disrupted with BugBuster™ Protein extraction reagent (Novagen) according to the provided manual. 50 µL of clear protein lysate was mixed with 100 µL 50 mM sodium citrate – phosphate buffer pH 5.0 and 50 µL of racemic mandelonitrile, previously dissolved in 3 mM sodium citrate – phosphate buffer pH 3.5 (8 µL/mL). Negative control: 50 µL BugBuster™ Protein extraction reagent (Novagen). Cyanide release was detected throughout a Feigl–Anger test paper[43]. All three isoenzymes are similar active.

Supplementary Result 6. *Dt*HNLs characterization



Supplementary Figure 6a. **SDS PAGE of *Dt*HNL purification.**

*Dt*HNL isoenzymes with N-terminal HIS-tags were purified by affinity chromatography as described in **Material and Methods**. Molecular weight standard: PageRuler™ Prestained Protein Ladder (Thermo Fisher Scientific) (1). **Purification of *Dt*HNL1 (A):** cell free lysate (2); flow through (3); elution fractions (4-11). **Purification of *Dt*HNL2 (B):** cell free lysate (2); flow through (3); best elution fractions (4-8); desalted fractions (9, 10). **Purification of *Dt*HNL3 (C):** cell free lysate (2); flow through (3); elution fractions (4-10). **Purification of *Dt*HNL4 (D):** cell free lysate (2); flow through (3); best elution fractions (4-6); desalted fractions (7, 8). Tagged Expected molecular weight of *Dt*HNL isoenzymes MW = 23 kDa.



Supplementary Figure 6b. Specific activity of HisTEV-*Dt*HNL1 *vs* untagged *Dt*HNL1.

0.5 mg of purified HisTEV-*Dt*HNL1 were incubated with TEV protease (His tagged protein, recombinantly expressed and purified by Ni-affinity chromatography). The reaction was carried out over night at 4°C. The obtained untagged *Dt*HNL1 was separated from the remaining uncleaved protein and TEV protease by Ni-affinity chromatography. Specifically, pure *Dt*HNL1 eluted in the flow through, whereas HisTEV-*Dt*HNL1 and TEV protease bind to the nickel resin. Specific activity of tagged and untagged pure *Dt*HNL1 were determined by the standard assay. The tag in *N*-terminal position does not negatively affect *Dt*HNL1 activity (*Dt*HNL1 w/TAG). Therefore, all other experiments reported in this work were performed with the purified HisTEV tagged enzymes, which have been named *Dt*HNL1-4 for convenience.

Supplementary Table 3. *Dt*HNL pH stability

Enzyme stability at different pH conditions was performed as described in **Material and Methods**. Additionally, pH 5.0 (50 mM sodium citrate phosphate buffer) and pH 6.5 (50 mM sodium phosphate buffer) were tested. Standard deviations refer to three different independent experiments.

***Dt*HNL1**				
Time of incubation	**pH 2.5**	**pH 4.0**	**pH 5.0**	**pH 6.5**
[h]	**Relative [%]**	**Relative [%]**	**Relative [%]**	**Relative [%]**
0	100±31	100±0	100±31	100±31
2	112±17	92±18	98±17	106±19
4	142±12	111±18	133±26	151±37
8	125±2	92±9	113±15	113±37
24	94±2	76±12	113±16	111±24
48	56±31	38±18	64±17	64±20
72	63±16	28±5	91±2	85±17
***Dt*HNL2**				
Time of incubation	**pH 2.5**	**pH 4.0**	**pH 5.0**	**pH 6.5**
[h]	Relative [%]	Relative [%]	Relative [%]	Relative [%]
0	100±13	100±13	100±13	100±13
2	112±24	101±11	121±23	119±22
4	127±36	120±18	122±27	123±26
8	110±19	98±20	95±19	96±12
24	83±19	96±16	95±7	101±7
48	64±15	103±18	91±25	110±17
72	28±11	102±19	104±14	104±11
*Dt*HNL3				
Time of incubation	pH 2.5	**pH 4.0**	**pH 5.0**	**pH 6.5**
[h]	Relative [%]	Relative [%]	Relative [%]	Relative [%]
0	100±4	100±4	100±19	100±19
4	91±1	95±13	101±11	117±15
8	88±11	101±5	109±8	113±20
24	80±0	86±6	98±13	78±14
48	71±1	103±2	111±13	86±15

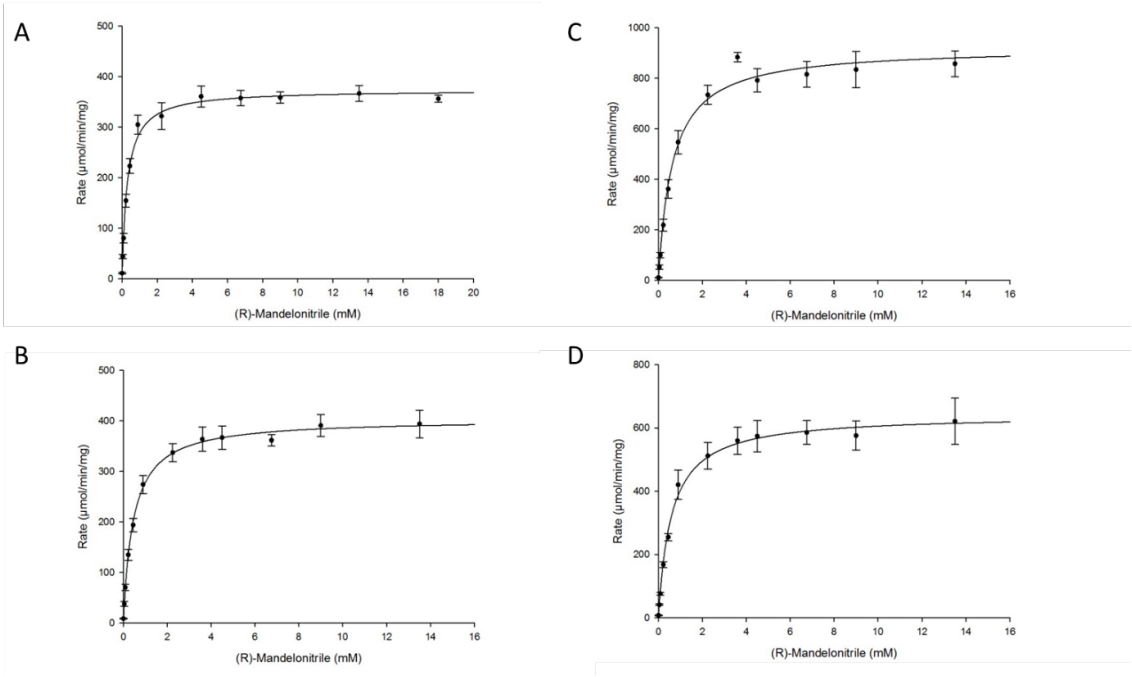
*Dt*HNL4				
Time of incubation	**pH 2.5**	**pH 4.0**	**pH 5.0**	**pH 6.5**
[h]	Relative [%]	Relative [%]	Relative [%]	Relative [%]
0	100±5	100±5	100±5	100±5
2	94±12	95±6	95±9	112±6
4	101±9	102±4	101±2	116±5
8	88±9	78±13	85±1	100±13
24	53±2	88±10	86±14	105±6
48	29±0	82±4	82±10	94±4
72	7±2	64±1	64±3	74±2

Supplementary Table 4. *Dt*HNL temperature stability.

1 mg/mL of purified protein was incubated in 50 mM sodium phosphate buffer pH 6.5 at 8°C, 20°C, 30°C and 40°C. At certain time-points an aliquot of enzyme was diluted to 0.01 mg/mL and HNL activity was measured at standard conditions. Standard deviations are reported (SD) and refer to three independent experiments. *Dt*HNL1 and *Dt*HNL2 are not stable at 40°C and loose activity already after 2 hours, whereas, *Dt*HNL3 and *Dt*HNL4 are more stable at 40°C. After 48 hours, *Dt*HNL3 and *Dt*HNL4 show 65% and 50% of their original activity, respectively. The four isoenzymes are stable at 20°C and 8°C, conditions typically compatibly to cyanohydrins stability.

***Dt*HNL1**								
Time of incubation	**8°C**		**20°C**		**30°C**		**40°C**	
[h]	**Relative [%]**	**SD**	**Relative [%]**	**SD**	**Relative [%]**	**SD**	**Relative [%]**	**SD**
0	100	5	100	5	100	5	100	5
2	118	14	114	10	113	5	2	2
4	110	25	106	6	107	4	1	1
8	123	11	120	21	126	13	1	0
24	119	10	107	8	119	2	0	0
96	122	1	145	12	113	3	0	0
***Dt*HNL2**								
Time of incubation	**8°C**		**20°C**		**30°C**		**40°C**	
[h]	**Relative [%]**	**SD**	**Relative [%]**	**SD**	**Relative [%]**	**SD**	**Relative [%]**	**SD**
0	100	2	100	2	100	2	100	2
4	67	14	99	19	68	11	0	0
8	67	15	102	22	75	2	1	0
24	56	17	83	29	69	3	0	0
48	53	15	108	20	64	5	0	0
72	61	17	115	29	38	5	0	1

***Dt*HNL3**								
Time of incubation	**8°C**		**20°C**		**30°C**		**40°C**	
[h]	**Relative [%]**	**SD**	**Relative [%]**	**SD**	**Relative [%]**	**SD**	**Relative [%]**	**SD**
0	100	23	100	25	100	30	100	26
4	108	18	93	23	96	29	94	32
8	99	3	91	16	96	5	95	15
24	88	34	98	46	123	50	106	87
48	95	30	107	25	96	39	68	24
72	94	25	110	12	53	2	67	18
***Dt*HNL4**								
Time of incubation	**8°C**		**20°C**		**30°C**		**40°C**	
[h]	Relative [%]	SD	Relative [%]	SD	Relative [%]	SD	Relative [%]	SD
0	100	18	100	18	100	18	100	18
4	96	26	93	20	107	30	98	5
8	87	23	82	19	94	23	79	5
24	77	27	69	20	105	21	53	21
48	73	24	85	32	111	27	51	15
72	64	26	97	37	102	30	41	8



Supplementary Figure 6c. **Michaelis Menten curves of *Dt*HNL isoenzymes.**

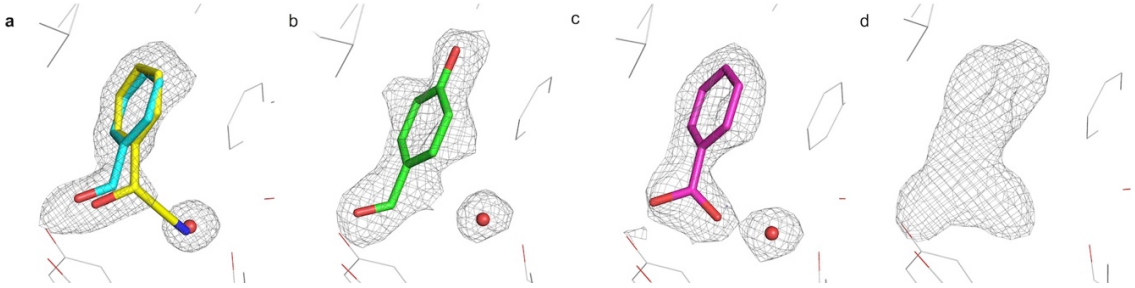
*Dt*HNL1 (A): v_{max} 373 ± 7 µmol min^{-1} mg^{-1}, K_m 0.297 ± 0.029 mM. *Dt*HNL2 (B): v_{max} 404 ± 8 µmol min^{-1} mg^{-1}, K_m 0.451 ± 0.047 mM. *Dt*HNL3 (C): v_{max} 922 ± 24 µmol min^{-1} mg^{-1}, K_m 0.639 ± 0.079 mM. *Dt*HNL4 (D): v_{max} 641 ± 20 µmol min^{-1} mg^{-1}, K_m 0.591 ± 0.089 mM. Error bars indicate the standard deviation obtained from three independent experiments.

Supplementary Result 7. *Dt*HNL1 Structure

Supplementary Table 5. Data-collection and processing statistics.

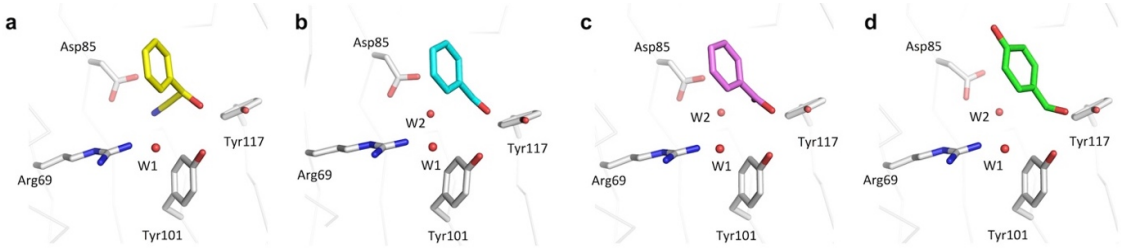
	***Dt*HNL1-SeMet**	***Dt*HNL1-MXN (10 sec)**	***Dt*HNL1-BEZ (1 min)**	***Dt*HNL1-HBA (15 min)**
Wavelength (Å)	0.9790	0.9184	0.9184	0.9184
Resolution range (Å)	57.97-1.85 (1.92-1.85)	35.8-1.50 (1.55-1.50)	57.93-1.85 (1.90-1.85)	36.76-1.80 (1.86-1.80)
Space group	*I*222	*I*222	*I*222	*I*222
Unit cell parameters (Å, °)	73.63, 94.02, 117.05 90, 90, 90	73.36, 94.14, 116.14 90, 90, 90	73.62, 93.87, 116.17 90, 90, 90	73.51, 93.47, 117.86 90, 90, 90
Total reflections	462708 (43863)	260107 (15412)	240033 (17990)	265397 (18210)
Unique reflections	34750 (3382)	63070 (5168)	34309 (3230)	37654 (3567)
Multiplicity	13.3 (13.0)	4.1 (3.0)	7.0 (5.6)	7.0 (5.1)
Completeness (%)	99.84 (98.60)	97.48 (80.88)	98.55 (93.69)	99.27 (95.25)
Mean I/σ(I)	9.97 (2.57)	13.74 (1.86)	13.72 (2.35)	16.66 (1.84)
Wilson B-factor	20.10	14.55	13.27	17.12
R-merge	0.188 (0.796)	0.066 (0.611)	0.150 (0.862)	0.108 (0.782)
R-meas	0.195	0.075	0.162	0.116
CC1/2	0.995 (0.825)	0.998 (0.713)	0.995 (0.925)	0.998 (0.634)
CC*	0.999 (0.951)	1 (0.912)	0.999 (0.98)	0.999 (0.881)
R-work	0.155 (0.212)	0.157 (0.270)	0.180 (0.383)	0.165 (0.294)
R-free	0.192 (0.270)	0.180 (0.277)	0.235 (0.470)	0.198 (0.345)
Number of non-hydrogen atoms	3399	3500	3287	3294
macromolecules	2932	2924	2830	2856
ligands	-	36	45	54
water	467	540	412	384
Protein residues	351	352	351	354
RMS(bonds)	0.007	0.007	0.007	0.007
RMS(angles)	1.02	1.03	1.01	1.00
Ramachandran favored (%)	97	98	98	97
Ramachandran allowed (%)	3	2	2	3
Ramachandran outliers (%)	0	0	0	0
Average B-factor	21.00	19.80	17.60	21.10
macromolecules	19.20	17.30	16.20	19.60
ligands	-	18.30	22.70	32.10
solvent	32.20	33.20	26.40	31.10
PDB code				

Statistics for the highest-resolution shell are shown in parentheses.



Supplementary Figure 7a. **Electron density maps.**

Fo-Fc omit density within the active site (contoured at 2σ) of the *Dt*HNL1 complexes with *(R)*-mandelonitrile (yellow) / benzaldehyde (cyan) (**a**), 4-hydroxybenzaldehyde (green) (**b**) and benzoic acid (magenta) (**c**); and for the native structure (d). Amino acid residues are shown as grey lines, the bound ligands are shown as sticks and water molecule as red sphere. The figure was prepared using the program PyMOL (Schrodinger Inc.).



Supplementary Figure 7b. **Ligand binding site of all determined *Dt*HNL1 complex structures.**

Amino acid residues are shown in grey sticks, the bound ligands in yellow (*(R)*-mandelonitrile), cyan (benzaldehyde), magenta (benzoic acid) and green (4-hydroxybenzaldehyde) sticks, and the water molecules as red spheres. The figure was prepared using the program PyMOL (Schrodinger Inc.).

Supplementary Result 8. Reaction mechanism and *Dt*HNL1 mutants.

Supplementary Table 6. Summary of specific activity of *Dt*HNL1 mutants.

n.d.: activity was not determined, due to insoluble expression of the protein.

Residue	Mutation	Activity	Activity Loss
		U/mg	%
R69	A69	n.d.	n.d.
R69	K69	n.d.	n.d.
D85	A85	1.5 ± 0.1	≥99
S87	A87	29 ± 1	91
Y101	A101	n.d.	n.d.
Y101	F101	0	100
Y117	A117	n.d.	n.d.
Y117	F117	≤ 0.5	≥99
Y161	A161	n.d.	n.d.
Y161	F161	26 ± 3.8	92
D85-S87	S85-D87	≤ 0.5	≥99

Supplementary Result 9. HNL from *P. aquilinum*

Supplementary Table 7. *Dt*HNL1 tblastn in *P. aquilinum* transcriptome (Accession n. PRJEB10897)

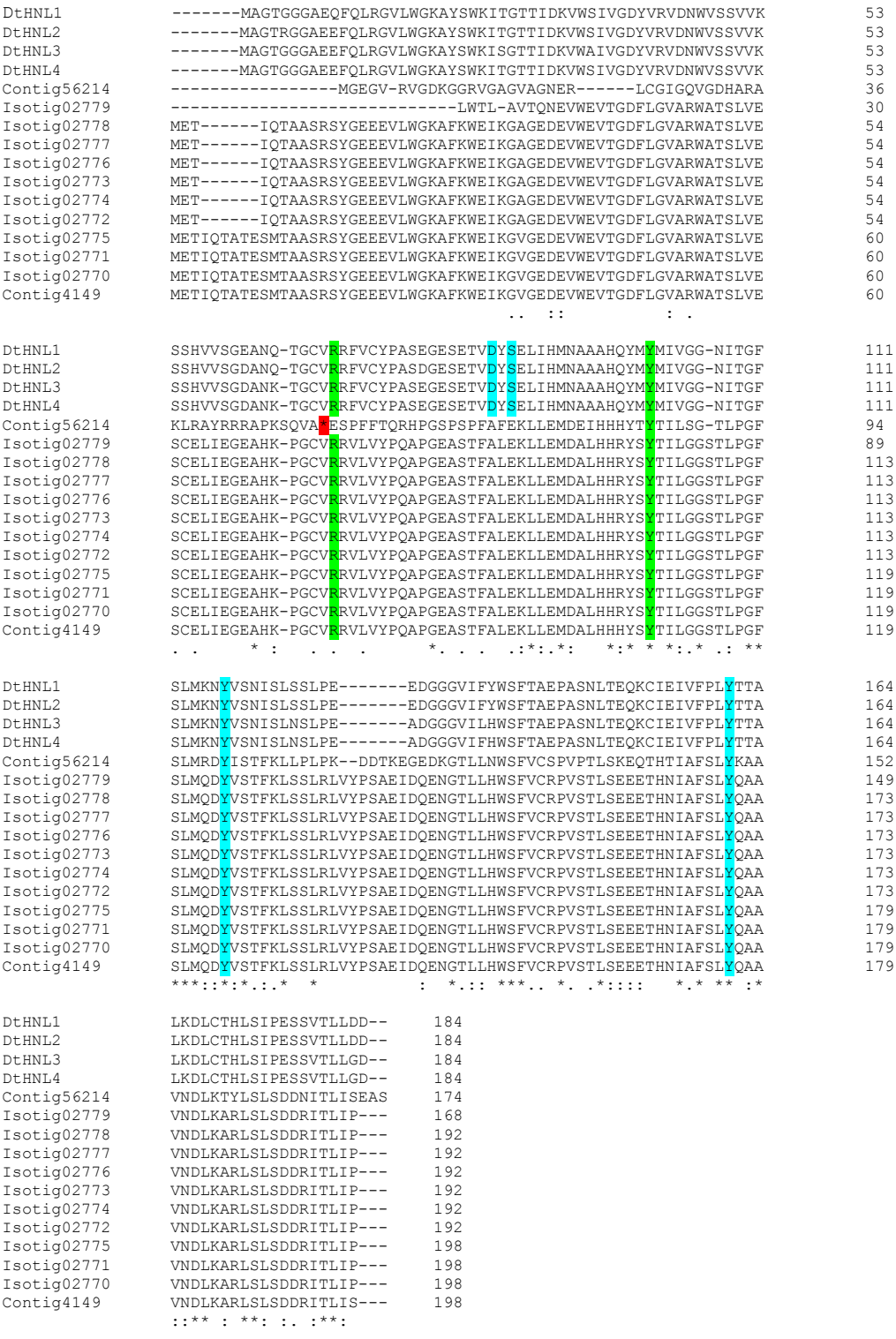
A tblastn search was performed with CLC Main Workbench 7.6.2 (QIAGEN Aarhus A/S), with the default parameters reported by the software (program: tblastn; expectation value: 100; word size: 3; mask lower case: no; filter low complexity: yes; maximum number of hits: 50; protein matrix and gap costs: BLOSUM62 existence 11 extension 1; number of threads: 1; genetic code: 1).

Entry	Hit	E-value	Score	Overlap	Identity	Positive	Gaps	Note
					%	%	%	
1	isotig02778	1.39E-39	350	285.3261	41	62.28571	4.571429	
2	isotig02777	1.41E-39	350	285.3261	41	62.28571	4.571429	Identical to entry 1
3	isotig02776	1.58E-39	350	285.3261	41	62.28571	4.571429	Identical to entry 1
4	isotig02775	1.58E-39	351	285.3261	41	62.28571	4.571429	
5	isotig02773	1.69E-39	350	285.3261	41	62.28571	4.571429	Identical to entry 1
6	isotig02771	1.71E-39	351	285.3261	41	62.28571	4.571429	Identical to entry 4
7	isotig02774	1.78E-39	350	285.3261	41	62.28571	4.571429	Identical to entry 1
8	isotig02772	1.81E-39	350	285.3261	41	62.28571	4.571429	Identical to entry 1
9	isotig02770	2.49E-39	351	285.3261	41	62.28571	4.571429	Identical to entry 4
10	isotig02779	3.38E-32	296	265.7609	39	60.7362	4.907975	Not full protein
11	contig56214	7.2E-22	191	180.9783	33	54.95495	4.504505	Mistakes in the sequence
12	contig56214	7.2E-22	96	107.6087	28	59.09091	0	
13	isotig32801	3.83E-08	118	210.3261	30	45.65217	12.31884	Low sequence query coverage
14	isotig32800	7.55E-08	117	203.8043	30	45.52239	12.68657	
15	isotig32290	3.29E-07	112	143.4783	27	52.17391	4.347826	
16	isotig32289	5.03E-07	112	143.4783	27	52.17391	4.347826	
17	isotig35067	6.26E-07	112	171.1957	26	47.16981	3.773585	

Supplementary Table 8. *Dt*HNL1 pblastn in gametophyte transcriptome of *P. aquilinum*[4].

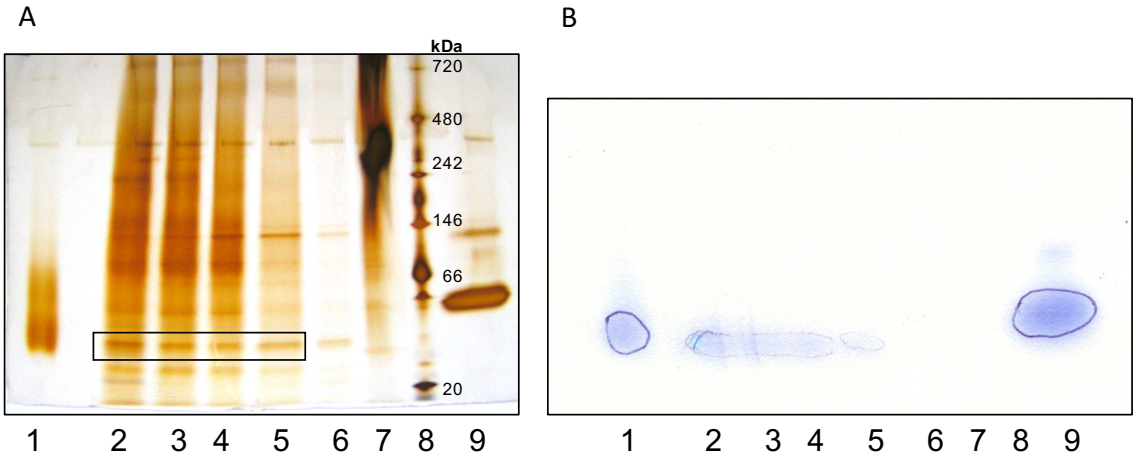
A tblastn search was performed with CLC Main Workbench 7.6.2 (QIAGEN Aarhus A/S), with the default parameters reported by the software (program: tblastn; expectation value: 100; word size: 3; mask lower case: no; filter low complexity: yes; maximum number of hits: 50; protein matrix and gap costs: BLOSUM62 existence 11 extension 1; number of threads: 1; genetic code: 1).

Entry	Hit	E-value	Score	Overlap	Identity	Positive	Gaps
					%	%	%
1	Contig4149	$1.41e^{-39}$	350	285.3261	41.71429	61.71429	4.571429
2	PtaqEST_c54074	$3.23e^{-06}$	103	146.7391	31.63265	51.02041	8.163265
3	PtaqEST_s93201	0.000521	88	127.1739	28.04878	51.21951	4.878049
4	Contig5171	3.52346	60	79.8913	32.65306	46.93878	4.081633
5	PtaqEST_c12935	6.24495	57	35.86957	50	72.72727	0
6	Contig2423	8.93556	57	35.86957	50	72.72727	0
7	Contig21095	9.31704	57	44.02174	40.74074	55.55556	0



Supplementary Figure 9a. **Multiple sequence alignment of *Dt*HNLs and putative *Pta*HNLs.**

Color code: stop codon (red); catalytic residues (green); important residues for HNL activity (cyan).



Supplementary Figure 9b. **BLUE NATIVE PAGE followed by HNL activity assay.**

In order to identify the HNL from *P. aquilinum*, the protein preparation from disrupted leaves was subjected to an anion exchange chromatography as described in **Online Method**. Elution fractions were concentrated and applied on a BN PAGE and then assayed for HNL activity[5]. A weak signal appears after 10 minutes at the same height as *Dt*HNL1 (dimer). The corresponding bands were analyzed by mass spectrometry and obtained peptides were matched against the *P. aquilinum* transcriptome. A list of isotigs is reported in the **Appendix, Supplementary Dataset 2**. None of the sequences found by blast were among the hits.

A: Different elution fractions after anion exchange purification and protein concentration were applied separately on BN PAGE (lane 2-6); flow through (7); positive control: purified *Dt*HNL1 (1); positive control *Pa*HNL (9); NativeMark™ Unstained Protein Standard (Thermo Fisher Scientific) (8). **B**: HNL activity is depicted by the blue spot in correspondence to the different purification fractions and the two positive controls *Dt*HNL (1) and *Pa*HNL (9).

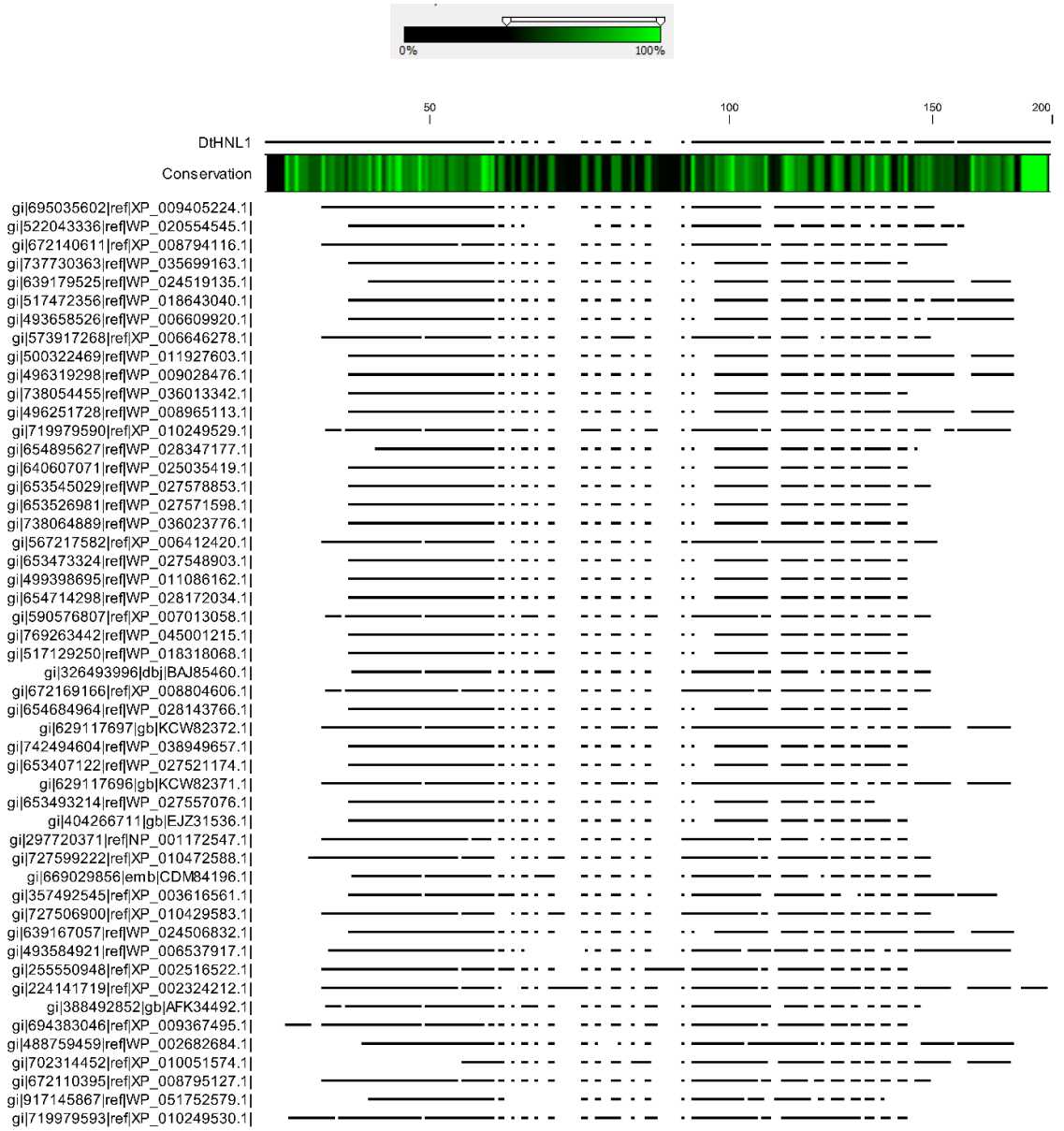
Supplementary Result 10. *Dt*HNL: a unique sequence within Bet v 1 superfamily

Bet v 1 protein superfamily is characterized by small acidic proteins moderately conserved in their tertiary structure, but definitely diverse at sequence level[6]. *Dt*HNL belongs to polyketide_cyc2 pf10604 protein family in release 28.0 of the pfam database[7].

Herein, we compared *Dt*HNL with other members of Bet v 1 superfamily, in order to find other similar proteins with HNL activity. *Dt*HNL was subjected to blastp search (**Supplementary Fig. 10a**), several unknown proteins with low similarity and sequence coverage were obtained. The sequence alignment between *Dt*HNL1 and the closest protein found (XP_009405224) is reported (**Supplementary Fig. 10b**). The six residues, important for HNL activity, are marked and Arg and two Tyr are conserved only. Furthermore, a Glu is present instead of Ser, which is unlikely compatible with HNL activity (**Results**). The closest related characterized protein is the lachrymatory factor synthase (LFS) from *Allium cepa* (Uniprot P59082)[8]. The two superimposed structures are visualized in **Supplementary Fig. 10e**. The architecture is similar, however sequence identity is less than 30% and catalytic and binding residues are different (**Supplementary Fig. 10d**). Second representatives of the protein family are phytohormone abscisic acid (ABA) receptors. They are characterized proteins and tertiary structured is solved[9]. They play a role in different biological functions including plant defense response from pathogen attack[10]. Again, important binding residues between *Dt*HNL and *At*PYL (Q8VZS8) are not conserved.

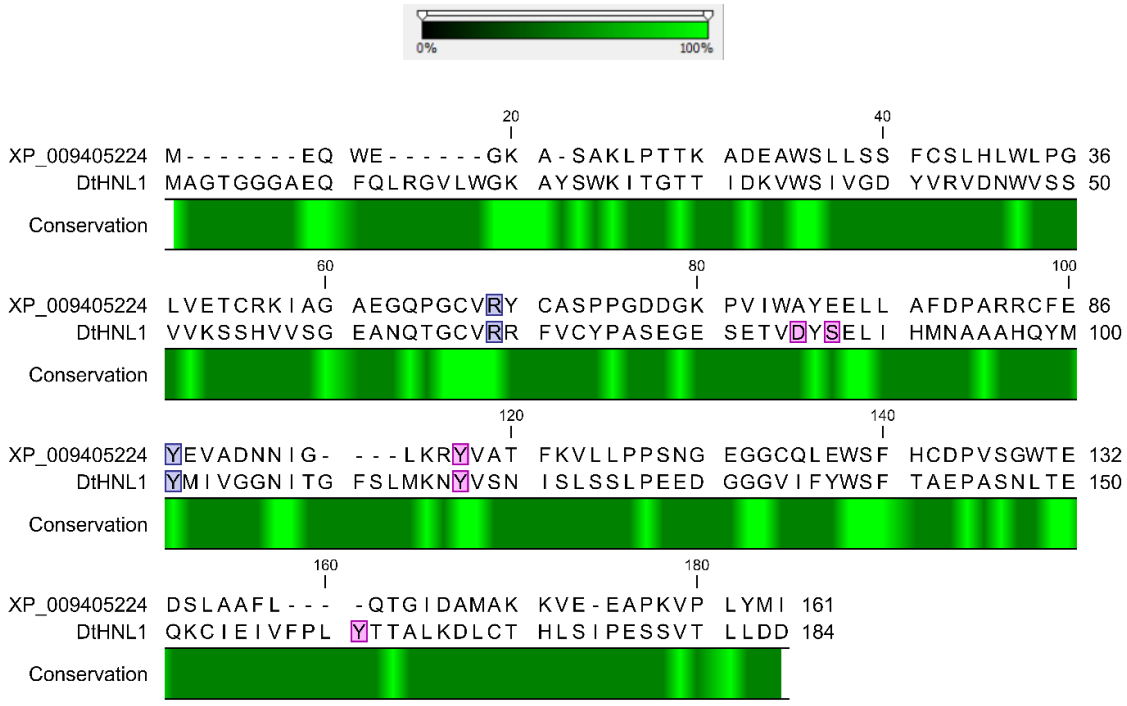
Due to the low similarity between the amino acid sequences, we focused on the comparison between *Dt*HNL and other protein structures of the same superfamily. However, there is a limited number of structures deposited in the PDB, which belong to the Bet v 1 protein superfamily. 3DM database overcome this problem[11], therefore a specific 3DM for Bet v 1 superfamily was developed. Specifically, the database includes 264 structured and 13,904 sequences (October 2014).The difference between DtHNL and the other superfamily's member is again remarked. Even if each subfamily includes

protein with very low sequence identity (up to 30%), *Dt*HNLs constitute a new subfamily named 3NEWA (Supplementary Fig. 10f). 3D numbers indicate the position of a specific residue within the protein structure, therefore we show residues Arg69, Tyr101, Tyr117, Tyr161, Asp85 and Ser87 and their 3D number (Supplementary Table. 11). We created a subset by picking only sequences having two Tyr at 3D position 79 and 91 respectively. Moreover, we set 3D position 50 as a basic residue (Arg or Lys). According to the described parameters, we obtained a subset of 491 sequences, 3.5% of the entire database. Then, we looked at the amino acid occurrence at 3D positions 63 and 65 within the created subset (Supplementary Fig.10g). Only *Dt*HNL isoenzymes show the desired residues Asp and Ser. Glu is relatively conserved at the 3D position 65, but unlikely compatible with HNL activity (Results). Therefore, we ruled out these related sequences. Based on these results, we can define *Dt*HNL as a truly unique enzyme within the whole protein superfamily and it is not possible to find another protein with the same substrate binding and catalysis features.



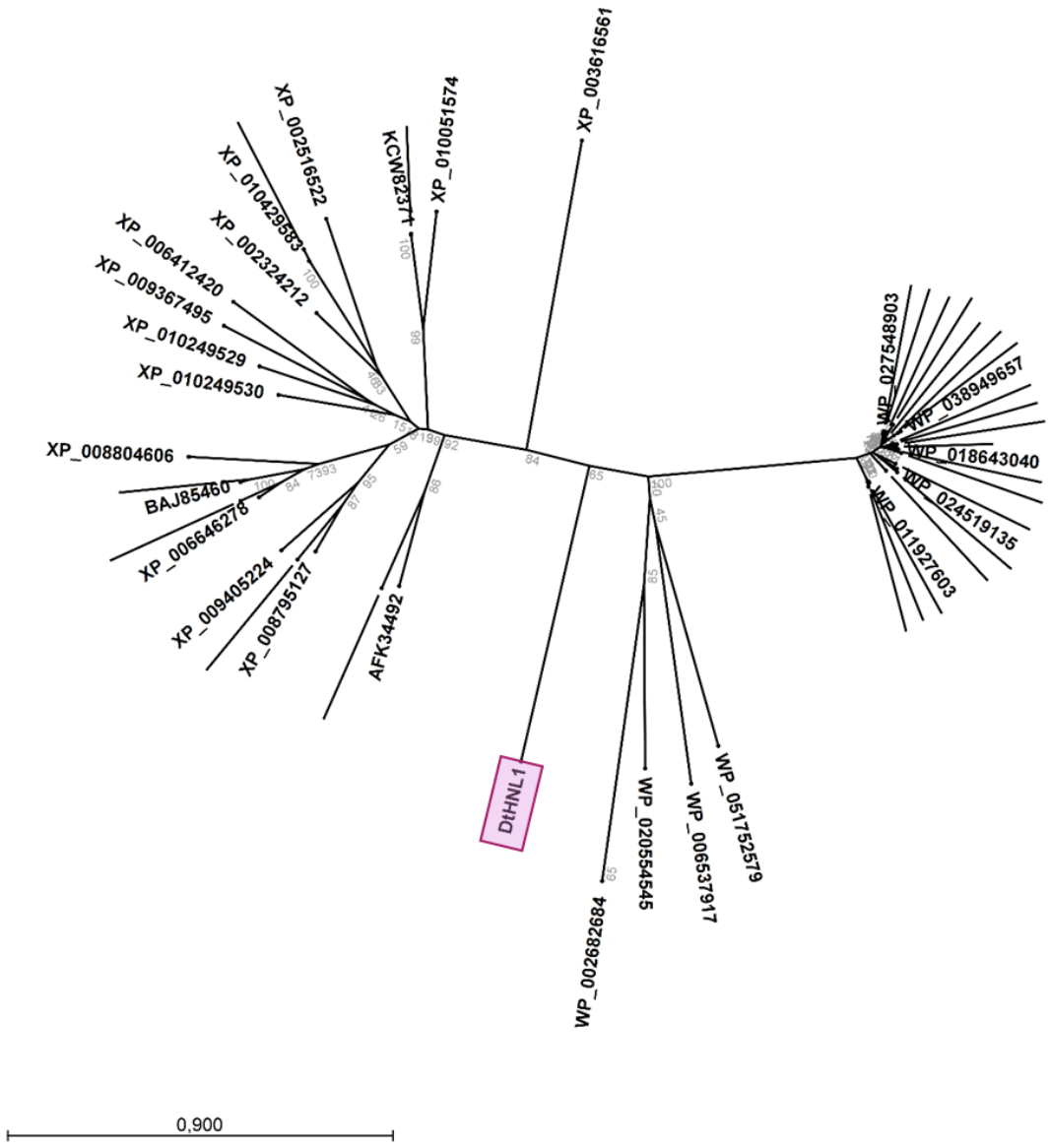
Supplementary Figure 10a. blastp of *Dt*HNL1 in NCBI.

A blastp search was performed with CLC Main Workbench 7.6.2 (QIAGEN Aarhus A/S), with the default parameters reported by the software (Protein matrix and gap costs: BLOSUM62 existence 11 extension 1; expectation value: 10.0; word size: 3; mask lower case: no; filter low complexity: yes; maximum number of hits: 50; limit by entry query: all organisms; database: nr).



Supplementary Figure 10b. **Sequence alignment of *Dt*HNL1 and XP_009405224.**

Sequence alignment was built with CLC Main Workbench 7.6.2 (QUIAGEN Aarhus A/S).



Supplementary Figure 10c. **Phylogenetic tree.**

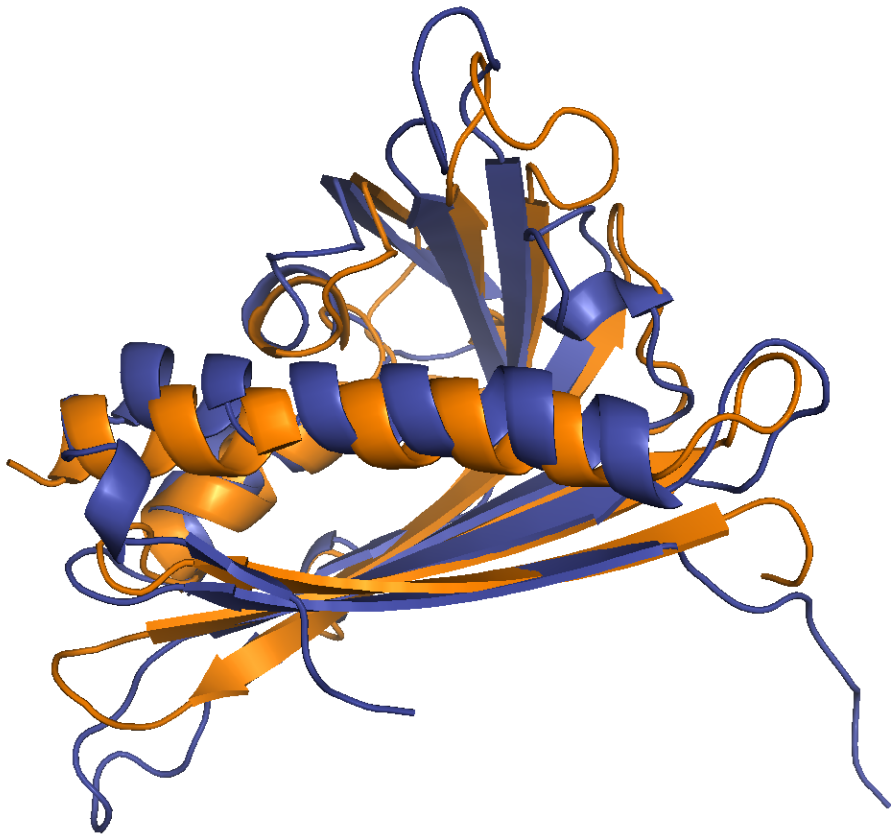
Thirty members of pf10604 are verified at the protein level sequences and they are currently deposited in the UniProtKB database[12] (Complete protein list **Appendix, Supplementary Dataset 3**). Most are not characterized proteins. A phylogenetic tree of the thirty sequences and *Dt*HNL1 is depicted. We can distinguish three main clusters, however *Dt*HNL does not belong to any of them.

Phylogenetic tree was built with CLC Main Workbench 7.6.2 (QIAGEN Aarhus A/S), with the default parameters reported by the software (Algorithm: neighbor Joining; distance measure: Jukes-Cantor; boostrap: 100 replicates).

```
DtHNL1     MAGTGGGAEQFQ-----------------------------------LRGVLWGKAY---SWKITGTTIDKVWSIVGDY 41

P59082     MELNPGAPAVVADSANGARK---------------------------WSGKVH---ALLP-NTKPEQAWTLLKDF 44
Q8VZS8     MANSESSSSPVNEEENSQRISTLHHQTMPSDLTQDEFTQLSQSIAEFHTYQLGNGRCSSLLAQRI-HAPPETVWSVVRRF 79
           *  . ...  .                                          *:     :     :  : .*:::  :

DtHNL1     VRVDNWVSSVVKSSHVVSGEANQTGCVRRFVCYPA-SEGESETVDYSELIHMNAAAHQYMYMIVGGNITGFSLMKNYVSN 120
P59082     INLHKVM-PSLSVCELVEGEANVVGCVRYV-KGIM-HPIEEEFWAKEKLVALDNKNMSYSYIFTE----CFTGYEDYTAT 117
Q8VZS8     DRPQIYK-HFIKSCNVSEDFEMRVGCTRDV-NVISGLPAN---TSRERLDLLDDDRRVTGFSITG----GEHRLRNYKSV 150
            . .      :. ..: ..    .**.* .         :      ..*  ::       : :.          .:* :

DtHNL1     ISLSSLPEEDGGGVIFYWSFTAE-PASNLTEQKCIEIVFPLYTT----ALKDLCTHLSIPESSVTLLDD------- 184
P59082     MQIVEGPEHKGS--RFDWSFQCK-YIEGMTESAFTEILQHWATE-IGQKIEEVC------S--------------A 169
Q8VZS8     TTVHRFEKEEEEERI--WTVVLESYVVDVPEGNSEEDTRLFADTVIRLNLQKLA---SITEAMNRNNNNNNSSQVR 221
             :    :..       *:.  :    .:.*    *             ::.:.      .
```

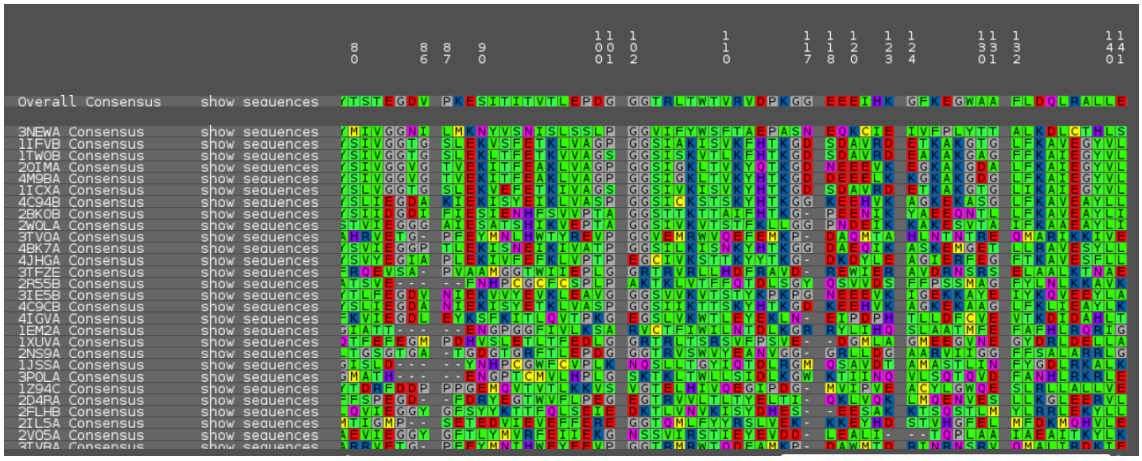
Supplementary Figure 10d. **Multiple sequence alignment of *Dt*HNL1, lachrymatory factor synthase and abscisic acid receptor.**

Important residues for substrate binding and activity are highlighted: *Dt*HNL1 (red); *Ac*LFS (cyan)[13]; *At*PYL1 (green)[9]. Multiple sequence alignment was performed with clustal Omega[3].



Supplementary Figure 10e. **Superimposed structures of *Ac*LFS and *Dt*HNL1.**

The *Ac*LFS structure was determined by homology modeling (Phyre2[14]).

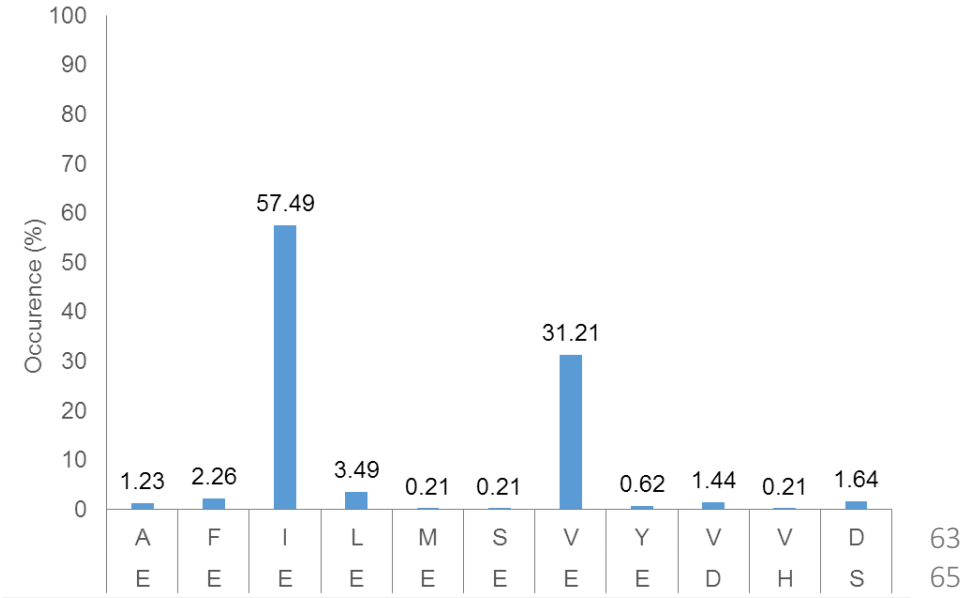


Supplementary Figure 10f. **Snapshot of 3DM[11] of the Bet v 1 protein superfamily.**

Part of the consensus alignment is represented. Specifically, the database includes 264 structures and 13,904 sequences. Subfamily 3NEWA includes *Dt*HNLs only. *Dt*HNL1 is the top entry (3NEWA).

Supplementary Table 9. 3D Number of *Dt*HNL relevant residues.

***Dt*HNL1 Residue**	**3D Number**
R69	50
D85	63
S87	65
Y101	79
Y117	91
Y161	129



Supplementary Figure 10g. **Residue occurrence in 3D positions 63 and 65.**

We created a subset selecting only sequences that contain a tyrosine at 3D position 79 and 91. Moreover, we kept a basic residue (arginine or lysine) at position 50. The subset consists of 493 sequences, 3.5% of the total database. In a second step, we observed for the amino acid occurrence at positions 63 and 65 within the latter subset. Only the four *Dt*HNL isoenzymes show the desired residues at the two positions (1.64% occurrence).

Appendix

Supplementary Table 10. Protein sequence list

Entry	Sequence
1	>*Dt*Isotig06604 MADSARTTVLVTGAGGRTGHIVYEKLKHKADKFHVRGLVRSEPSKAKIGGGEDVYIGDITKAESLGPAFAGVDVLIILT SAAPQMKPGFDPSKGGRPEFYYEEGAYPEQVDWIGQKNQIDAAKEAGVKHIILVGSMGGTNPNHPLNSFGNGKILIWKR KSEQYLADSDTTYTLIRPGGLLDKEGGLRELLIGNNDELLATDTKTVPRADVAEVCVQAIVHDAVKNKAFDLTSKPEGE GTPTTDFKILFSQVTASF
2	>*Dt*Isotig07200 MSYWKSKVVPKIKKFFDKGKKKGAAEFSKNFDSSKESLDKEIGEKSSDLSPKVVEIYRSSSTFIAKKLLKEPNEATVKE NSDATQGVLQELATAGFPGAQGIADAGKKYGPALLPGPVVYLFEKASVFLAEEPLPEEPKAETREVSAEDVKPAEAPAT TSETPPPPPVADVPPPAVVEEEKKEAEPIVAAPPPEAAPPAAVEIPTSVDPTPPPPAPPADKPE
3	>*Dt*Isotig04065 MAKVHIMLLCTLCALSLSSLSAPQPALAASGPDHLDFYMYIAVQNNSNLDNPNVTFTAVQSAQPLSTQPNSFGIIHTFD NPLTSAADLNSTQLGHVQGWYGDVGQNLLTLFLAQTFTWNDGTYNGTFSLLGVDVASDAVKFAPIVGGTGDFAYVRGVA QQSLVSTATVNMETVSWFFYAIDFVY
4	>*Dt*Isotig04379 MGTSTWVVWSILLLAVAQVAGSIPIPRRYDGFVFNASSSSSPVLLEAFFDPLCPDSADAWPVVKKIAQYFQDDLLLIVH PFPLPYHHNAYFASRALHIINNLNSSLTYPLLELFFENQDSFSTSETLAEAPSSVVDRIVQLAADSLNELVSSDFESQF KAGFSDTGTDLITRVSFKFGCSRVVVGTPYFFVNGIPLYDADSAWTFSEWAEIIEPLTGAQKIALA
5	>*Dt*Contig00505 MPFAQSLIVLFLAASALSYGGVLATTITAVNNCGTSGPLEFTGTSANGMNLAPAQSSGPIGVPDGWSGRVSLDPSPSTL AEFSIVQNNKNTMDISLVDGFNVALGISYTGGNCIRNGEAAASNVACHISIDQCPASYRQGDRCVNPNKDAQTDYSATV KGICPDAYSWSKDDATSTFTCDVGGDFTVTFCPP
6	>*Dt*HNL1 MAGTGGGAEQFQLRGVLWGKAYSWKITGTTIDKVWSIVGDYVRVDNWVSSVVKSSHVVSGEANQTGCVRRFVCYPASEG ESETVDYSELIHMNAAAHQYMYMIVGGNITGFSLMKNYVSNISLSSLPEEDGGGVIFYWSFTAEPASNLTEQKCIEIVF PLYTTALKDLCTHLSIPESSVTLLDD
7	>*Dt*HNL2 MAGTRGGAEEFQLRGVLWGKAYSWKITGTTIDKVWSIVGDYVRVDNWVSSVVKSSHVVSGDANQTGCVRRFVCYPASDG ESETVDYSELIHMNAAAHQYMYMIVGGNITGFSLMKNYVSNISLSSLPEEDGGGVIFYWSFTAEPASNLTEQKCIEIVF PLYTTALKDLCTHLSIPESSVTLLDD
8	>*Dt*HNL3 MAGTGGGAEEFQLRGVLWGKAYSWKISGTTIDKVWAIVGDYVRVDNWVSSVVKSSHVVSGDANKTGCVRRFVCYPASEG ESETVDYSELIHMNAAAHQYMYMIVGGNITGFSLMKNYVSNISLNSLPEADGGGVILHWSFTAEPASNLTEQKCIEIVF PLYTTALKDLCTHLSIPESSVTLLGD
9	>*Dt*HNL4 MAGTGGGAEEFQLRGVLWGKAYSWKITGTTIDKVWSIVGDYVRVDNWVSSVVKSSHVVSGDANKTGCVRRFVCYPASEG ESETVDYSELIHMNAAAHQYMYMIVGGNITGFSLMKNYVSNISLNSLPEADGGGVIFHWSFTAEPASNLTEQKCIEIVF PLYTTALKDLCTHLSIPESSVTLLGD
10	>*Pta*Isotig02775 METIQTATESMTAASRSYGEEEVLWGKAFKWEIKGVGEDEVWEVTGDFLGVARWATSLVESCELIEGEAHKPGCVRRVL VYPQAPGEASTFALEKLLEMDALHHRYSYTILGGSTLPGFSLMQDYVSTFKLSSLRLVYPSAEIDQENGTLLHWSFVCR PVSTLSEEETHNIAFSLYQAAVNDLKARLSLSDDRITLIP
11	>*Dt*HNL1_R69A MAGTGGGAEQFQLRGVLWGKAYSWKITGTTIDKVWSIVGDYVRVDNWVSSVVKSSHVVSGEANQTGCVARFVCYPASEG ESETVDYSELIHMNAAAHQYMYMIVGGNITGFSLMKNYVSNISLSSLPEEDGGGVIFYWSFTAEPASNLTEQKCIEIVF PLYTTALKDLCTHLSIPESSVTLLDD
12	>*Dt*HNL1_S87A MAGTGGGAEQFQLRGVLWGKAYSWKITGTTIDKVWSIVGDYVRVDNWVSSVVKSSHVVSGEANQTGCVRRFVCYPASEG ESETVDYAELIHMNAAAHQYMYMIVGGNITGFSLMKNYVSNISLSSLPEEDGGGVIFYWSFTAEPASNLTEQKCIEIVF PLYTTALKDLCTHLSIPESSVTLLDD
13	>*Dt*HNL1_D85A MAGTGGGAEQFQLRGVLWGKAYSWKITGTTIDKVWSIVGDYVRVDNWVSSVVKSSHVVSGEANQTGCVRRFVCYPASEG ESETVAYSELIHMNAAAHQYMYMIVGGNITGFSLMKNYVSNISLSSLPEEDGGGVIFYWSFTAEPASNLTEQKCIEIVF PLYTTALKDLCTHLSIPESSVTLLDD
14	>*Dt*HNL1_Y101A MAGTGGGAEQFQLRGVLWGKAYSWKITGTTIDKVWSIVGDYVRVDNWVSSVVKSSHVVSGEANQTGCVRRFVCYPASEG ESETVDYSELIHMNAAAHQYMAMIVGGNITGFSLMKNYVSNISLSSLPEEDGGGVIFYWSFTAEPASNLTEQKCIEIVF PLYTTALKDLCTHLSIPESSVTLLDD

Entry	Sequence
15	>*Dt*HNL1_Y117A MAGTGGGAEQFQLRGVLWGKAYSWKITGTTIDKVWSIVGDYVRVDNWVSSVVKSSHVVSGEANQTGCVRRFVCYPASEG ESETVDYSELIHMNAAAHQYMYMIVGGNITGFSLMKNAVSNISLSSLPEEDGGGVIFYWSFTAEPASNLTEQKCIEIVF PLYTTALKDLCTHLSIPESSVTLLDD
16	>*Dt*HNL1_Y161A MAGTGGGAEQFQLRGVLWGKAYSWKITGTTIDKVWSIVGDYVRVDNWVSSVVKSSHVVSGEANQTGCVRRFVCYPASEG ESETVDYSELIHMNAAAHQYMYMIVGGNITGFSLMKNYVSNISLSSLPEEDGGGVIFYWSFTAEPASNLTEQKCIEIVF PLATTALKDLCTHLSIPESSVTLLDD
17	>*Pta*Isotig02775_A92S METIQTATESMTAASRSYGEEEVLWGKAFKWEIKGVGEDEVWEVTGDFLGVARWATSLVESCELIEGEAHKPGCVRRVL VYPQAPGEASTFSLEKLLEMDALHHRYSYTILGGSTLPGFSLMQDYVSTFKLSSLRLVYPSAEIDQENGTLLHWSFVCR PVSTLSEEETHNIAFSLYQAAVNDLKARLSLSDDRITLIP
18	>*Pta*Isotig02775_A92D_E94S METIQTATESMTAASRSYGEEEVLWGKAFKWEIKGVGEDEVWEVTGDFLGVARWATSLVESCELIEGEAHKPGCVRRVL VYPQAPGEASTFDLSKLLEMDALHHRYSYTILGGSTLPGFSLMQDYVSTFKLSSLRLVYPSAEIDQENGTLLHWSFVCR PVSTLSEEETHNIAFSLYQAAVNDLKARLSLSDDRITLIP

Supplementary Table 11. Genes Deposited to GenBank® Database

Entry	GenBanK Accession Number	Description	Ref. Protein (Supplementary Table 12)
1	KT804569	*Davallia tyermannii* Hydroxnyitrile lyase Isoform 1	DtHNL1
	Sequence		
	ATGGCGGGAACGGGAGGGGGCGCAGAACAGTTCCAGCTCCGGGGAGTGCTGTGGGGGAAAGCCTACTCTTGGAAGATAAC CGGAACGACAATCGACAAGGTGTGGTCGATTGTGGGCGATTATGTGCGCGTCGACAACTGGGTCTCTTCCGTCGTGAAGA GCTCGCACGTCGTGTCTGGCGAGGCCAACCAGACGGGGTGCGTGAGGAGGTTCGTCTGCTACCCAGCCTCCGAGGGAGAG TCGGAGACTGTGGACTACTCGGAGCTCATCCACATGAACGCTGCCGCGCACCAGTACATGTACATGATCGTGGGAGGTAA CATCACTGGCTTCTCTCTCATGAAGAACTATGTGAGCAATATCTCGCTGTCTTCTCTTCCTGAGGAGGACGGTGGTGGTG TAATCTTTTACTGGAGCTTCACAGCCGAGCCTGCCTCTAACCTCACGGAACAAAAATGCATAGAAATTGTGTTCCCTCTC TATACCACTGCCCTGAAGGATTTATGCACTCACCTTTCCATACCCGAAAGCTCTGTTACACTTCTCGATGATTAA		
	GenBanK Accession Number	**Description**	**Ref. Protein** (Supplementary Table 12)
2	KT805919	*Davallia tyermannii* Hydroxnyitrile lyase Isoform 2	DtHNL2
	Sequence		
	ATGGCGGGAACGAGAGGAGGCGCTGAAGAGTTCCAGCTCCGGGGAGTGCTGTGGGGGAAAGCCTACTCTTGGAAGATAAC GGGAACGACAATCGACAAGGTGTGGTCGATTGTGGGTGATTATGTGCGCGTCGACAACTGGGTCTCTTCCGTCGTGAAGA GCTCGCACGTCGTGTCCGGCGATGCCAACCAGACGGGGTGCGTGAGGAGGTTCGTCTGCTACCCAGCCTCCGATGGAGAG TCGGAGACTGTGGACTACTCGGAGCTCATCCACATGAACGCCGCCGCTCACCAATACATGTACATGATTGTGGGAGGTAA CATCACTGGCTTCTCTCTCATGAAGAACTATGTGAGCAATATCTCGCTGTCTTCTCTTCCTGAGGAGGACGGTGGTGGTG TCATCTTTTACTGGAGCTTCACAGCCGAGCCTGCCTCTAACCTCACGGAACAAAAATGCATAGAAATTGTGTTTCCTCTC TACACCACTGCCCTGAAGGATTTATGCACTCACCTTTCCATACCCGAAAGCTCTGTTACACTTCTCGATGATTAA		
	GenBanK Accession Number	**Description**	**Ref. Protein** (Supplementary Table 12)
3	KT805921	*Davallia tyermannii* Hydroxnyitrile lyase Isoform 3	DtHNL3
	Sequence		
	ATGGCAGGAACGGGAGGGGGCGCAGAAGAGTTCCAGCTGCGGGGAGTGCTGTGGGGGAAAGCCTACTCGTGGAAGATATC GGGAACGACAATTGACAAGGTGTGGGCGATTGTGGGCGACTATGTGCGCGTCGACAACTGGGTCTCTTCTGTAGTGAAGA GCTCGCACGTCGTGTCTGGCGACGCTAACAAGACGGGGTGCGTGAGGAGGTTCGTCTGCTACCCAGCCTCCGAGGGAGAG TCGGAGACTGTGGACTACTCGGAGCTCATCCACATGAACGCGGCCGCGCACCAGTACATGTACATGATTGTGGGAGGTAA CATCACTGGCTTCTCTCTCATGAAGAACTATGTGAGCAATATATCGCTCAATTCTCTTCCTGAGGCGGACGGTGGTGGTG TCATCCTCCACTGGAGCTTCACAGCCGAGCCTGCTTCTAACCTCACCGAACAAAAATGTATAGAAATTGTGTTCCCTCTC TATACCACTGCCTTGAAGGATTTATGCACTCACCTTTCTATTCCGGAAAGCTCTGTTACACTCCTCGGTGATTAA		
	GenBanK Accession Number	**Description**	**Ref. Protein** (Supplementary Table 12)
4		*Davallia tyermannii* Hydroxnyitrile lyase Isoform 4	DtHNL4
	Sequence		
	ATGGCAGGAACGGGAGGGGGCGCAGAAGAGTTCCAGCTGCGGGGAGTGCTGTGGGGGAAAGCCTACTCTTGGAAGATAAC GGGAACGACAATCGACAAGGTGTGGTCGATTGTGGGCGACTATGTCCGCGTCGATAACTGGGTCTCTTCCGTAGTGAAGA GCTCGCACGTCGTGTCTGGCGATGCGAACAAGACGGGGTGCGTGAGGAGGTTCGTCTGCTACCCAGCCTCCGAGGGAGAG TCGGAGACTGTGGACTACTCGGAGCTCATCCACATGAATGCGGCCGCGCACCAGTACATGTACATGATCGTGGGAGGTAA CATCACTGGCTTCTCTCTCATGAAGAACTATGTGAGCAATATATCGCTCAATTCTCTTCCTGAGGCGGACGGAGGTGGTG TCATCTTCCACTGGAGCTTCACAGCCGAGCCTGCTTCTAACCTCACCGAACAAAAATGCATCGAAATTGTGTTCCCTCTC TATACCACTGCCTTGAAGGATTTATGCACTCACCTTTCTATCCCGGAAAGCTCTGTTACACTCCTCGGTGATTAA		
	GenBanK Accession Number	**Description**	**Ref. Protein** (Supplementary Table 12)
5	KT818577	*Davallia tyermannii* Unknown Protein	DtIsotig07200
	Sequence		
	ATGAGTTATTGGAAGAGCAAGGTTGTGCCCAAAATCAAGAAGTTTTTTGACAAGGGGAAGAAGAAAGGAGCTGCTGAGTT CTCCAAAAACTTTGATTCGTCCAAGGAGTCTTTGGACAAGGAGATAGGAGAGAAGAGTTCAGATCTGAGCCCCAAGGTTG TGGAGATATACAGATCCTCTTCCACCTTCATTGCCAAGAAGTTGCTGAAGGAACCCAATGAGGCAACAGTGAAGGAAAAT TCCGATGCAACGCAAGGCGTGCTTCAGGAGCTGGCAACAGCAGGCTTTCCTGGAGCGCAGGGCATCGCTGATGCAGGCAA AAAGTATGGACCGGCGCTTTTACCGGGGCCGGTCGTGTACTTGTTCGAGAAAGCATCCGTGTTTTTGGCGGAGGAGCCTT TGCCAGAGGAGCCCAAGGCAGAGACTAGAGAGGTGAGTGCGGAGGACGTCAAGCCAGCAGAAGCGCCAGCGACGACTTCG GAAACTCCGCCGCCGCCACCTGTAGCAGACGTGCCTCCCCCAGCTGTCGTCGAGGAAGAAAAGAAGGAAGCAGAGCCCAT TGTTGCTGCCCCGCCGCCCGAAGCTGCCCCCCCTGCAGCTGTTGAGATCCCCACCTCCGTTGACCCTACGCCCCCGCCTC CTGCTCCGCCCGCCGACAAACCTGAGTAG		

	GenBanK Accession Number	**Description**	**Ref. Protein** (Supplementary Table 12)
6	KT818578	*Davallia tyermannii* Unknown Protein	DtIsotig06604
	Sequence		
	ATGGCAGACTCTGCTCGCACAACCGTACTCGTAACTGGTGCTGGTGGAAGAACGGGGCACATTGTATATGAGAAGCTGAA ACATAAGGCAGACAAGTTTCATGTGAGAGGTCTTGTGAGGTCAGAGCCAAGCAAGGCAAAGATTGGAGGGGGTGAGGATG TGTACATAGGTGACATAACAAAGGCAGAGAGCCTGGGCCCAGCATTTGCAGGGGTGGATGTGCTCATCATCCTCACAAGT GCTGCTCCTCAAATGAAGCCAGGGTTTGATCCAAGCAAAGGAGGGCGCCCCGAGTTCTACTACGAAGAAGGCGCATATCC TGAGCAGGTTGATTGGATCGGGCAAAAGAATCAAATCGATGCAGCCAAAGAAGCGGGGGTGAAGCATATCATTTTGGTTG GCTCGATGGGTGGCACAAATCCGAATCATCCTCTGAACTCTTTCGGAAATGGGAAGATTTTGATTTGGAAAAGGAAGTCG GAGCAGTATTTGGCAGACTCTGATACAACCTACACCTTAATCAGGCCGGGTGGGCTGCTAGACAAAGAAGGTGGGTTGCG AGAGCTCTTGATTGGAAACAACGACGAACTTCTCGCCACAGATACCAAAACGGTTCCACGAGCAGATGTTGCGGAAGTCT GTGTACAGGCAATTGTTCATGATGCAGTAAAGAATAAGGCTTTTGACTTGACCTCGAAACCGGAGGGTGAAGGTACGCCG ACGACAGATTTCAAGATTCTCTTCAGTCAAGTAACTGCAAGCTTTTAG		
	GenBanK Accession Number	**Description**	**Ref. Protein** (Supplementary Table 12)
7	KT818579	*Davallia tyermannii* Unknown Protein	DtIsotig04379
	Sequence		
	ATGGGCACCTCAACGTGGGTGGTATGGAGCATTCTGTTGCTAGCAGTGGCGCAAGTAGCGGGGAGCATCCCCATCCCAAG ACGCTACGATGGCTTCGTCTTCAATGCTTCTTCATCCTCGTCGCCTGTGTTGCTGGAGGCCTTCTTCGATCCCCTCTGCC CGGATAGCGCAGACGCTTGGCCTGTTGTCAAGAAAATCGCCCAATACTTCCAGGACGATCTGCTCCTCATTGTCCACCCC TTCCCTCTCCCGTACCATCACAATGCATATTTTGCAAGTAGAGCATTGCACATCATCAATAACCTGAACAGTTCTCTCAC TTATCCATTGCTTGAGTTGTTTTTTGAAAACCAGGATAGCTTTTCAACGAGTGAAACGCTAGCGGAGGCACCATCCTCCG TCGTAGACAGAATCGTTCAACTGGCAGCAGATAGCTTGAATGAACTCGTGTCTTCCGATTTTGAGAGCCAGTTCAAAGCA GGATTTTCTGATACAGGAACGGATCTGATCACTCGTGTTTCGTTCAAGTTTGGGTGCTCGCGCGTTGTGGTCGGTACGCC GTACTTTTTTGTCAATGGCATACCTCTTTATGATCGGATTCGGCATGGACCTTCTCTGAATGGGCAGAGATTATTGAGCC ATTAACAGGGGCGCAAAAAATCGCGCTTGCATAA		
	GenBanK Accession Number	**Description**	**Ref. Protein** (Supplementary Table 12)
8	KT818580	*Davallia tyermannii* Unknown Protein	DtIsotig04065
	Sequence		
	ATGGCCAAGGTGCACATTATGCTCCTATGTACATTATGCGCTCTCTCCCTCTCCTCCCTCTCTGCCCCACAGCCAGCTTT GGCTGCCTCGGGTCCTGACCATCTTGACTTCTACATGTACATTGCTGTTCAGAATAACAGCAATCTCGACAACCCCAATG TCACCTTCACAGCCGTGCAGTCCGCGCAGCCGCTTTCGACGCAACCCAACTCATTCGGCATCATCCACACCTTCGACAAC CCACTCACCAGTGCCGCAGATCTCAACTCTACGCAGCTCGGGCACGTGCAGGGCTGGTATGGTGATGTGGGGCAGAATCT TTTGACGTTGTTCCTGGCGCAGACCTTTACCTGGAACGACGGCACCTACAATGGCACATTCAGCCTGTTGGGGGTGGACG TCGCGTCAGATGCGGTAAAGTTTGCACCCATTGTTGGCGGCACGGGCGACTTTGCGTATGTGCGCGGAGTGGCTCAACAG TCCCTTGTCTCAACCGCTACTGTAAACATGGAAACGGTCTCATGGTTCTTTTATGCCATTGACTTTGTATACTAG		
	GenBanK Accession Number	**Description**	**Ref. Protein** (Supplementary Table 12)
9	KT818581	*Davallia tyermannii* Unknown Protein	DtContig00505
	Sequence		
	ATGCCCTTTGCTCAATCCTTGATAGTGCTTTTCCTTGCGGCTTCAGCACTCAGCTATGGAGGAGTGTTGGCAACAACCAT AACGGCTGTGAACAACTGTGGAACAAGCGGCCCACTTGAGTTCACAGGCACTAGTGCCAACGGCATGAACTTGGCACCTG CACAATCGTCTGGCCCCATCGGTGTACCTGACGGATGGTCGGGCCGAGTTTCGCTGGACCCTTCGCCGTCCACTTTAGCA GAGTTCAGCATCGTCCAAAACAACAAGAATACCATGGATATTAGTCTGGTGGATGGCTTCAACGTTGCTCTGGGAATCTC ATACACCGGTGGTAATTGCATAAGGAATGGTGAAGCTGCAGCTAGCAACGTGGCATGCCACATTTCTATCGACCAGTGTC CTGCAAGCTACAGACAAGGCGACCGATGCGTCAACCCTAACAAAGACGCCCAGACTGACTACTCCGCTACTGTGAAGGGG ATATGTCCGGACGCCTATAGCTGGTCCAAGGATGATGCAACTAGCACGTTCACGTGCGATGTTGGTGGTGACTTTACCGT CACATTCTGCCCTCCATGA		
	GenBanK Accession Number	**Description**	**Ref. Protein** (Supplementary Table 12)
10	KT818582	*Pteridium aquilinum* Unknown Protein	PtaIsotig02775
	Sequence		
	ATGGAGACGATTCAAACAGCGACGGAGTCGATGACAGCGGCTAGCAGGAGCTATGGAGAGGAGGAGGTATTATGGGGGAA GGCGTTCAAGTGGGAGATAAAGGGTGTAGGGGAGGACGAGGTGTGGGAGGTAACCGGAGACTTTCTGGGAGTGGCCAGGT GGGCAACCTCGCTGGTGGAGAGCTGTGAGCTTATAGAAGGAGAGGCCCATAAGCCAGGCTGCGTGAGAAGGGTCCTTGTT TATCCCCAGGCTCCTGGGGAGGCCTCCACTTTTGCCCTTGAAAAGCTCTTAGAAATGGACGCGCTACACCACCGTTACTC TTACACTATCCTTGGCGGAAGCACCTTGCCTGGCTTCTCTCTCATGCAGGACTATGTCAGTACCTTCAAGCTCTCTTCCC TACGCCTGGTGTACCCCTCTGCAGAAATTGACCAAGAAAATGGTACCCTCCTCCATTGGAGCTTTGTTTGTCGCCCAGTC TCTACCTTGTCTGAGGAGGAAACCCACAACATTGCCTTCTCTCTCTACCAGGCTGCAGTCAACGATCTCAAAGCTCGCCT CTCCTTGTCTGACGACCGCATTACTCTCATCCCGTAA		

Supplementary Table 12. Primer List

Entry	Primer Name	Sequence	Purpose
1	DtIsotig02643_fw	AGCTCCCTAGCAAGTCATG	amplification from gDNA
2	DtIsotig02641_fw	AGAGAGTGAGGCGAGGTAG	amplification from gDNA
3	DtIsotig02641/3_rev	GGAGGATGAAAAGCTTAATC	amplification from gDNA
3	DtIsotig07602_fw	TAGAAAATGTAATTAGGGGGGTGAGATAAAG	amplification from gDNA
4	DtIsotig07602_rev	GAGAGTAAGGAGCAGTAGGCAAGC	amplification from gDNA
5	DtContig00751_fw	TATAATTAGGGAGGGGTGAGATAAAGC	amplification from gDNA
6	DtContig00751_rev	GTAAGGGGCAGTAGGCAAGC	amplification f rom gDNA
7	DtHNL1_Ec_NcoI_fw	AATGCCCATGGCAGGCACCGGTGGTG	Cloning pEHisTEV
8	DtHNL1_Ec_HindIII_rev	AATGCAAGCTTTTAATCATCCAGCAGGGTAACGC	Cloning pEHisTEV
9	DtHNL2_Dt_NcoI_fw	AATGCCCATGGCGGGAACGAGAGGAGGCG	Cloning pEHisTEV
10	DtHNL2_Dt_HindIII_rev	AATGCAAGCTTTTAATCATCGAGAAGTGTAACAGAGC	Cloning pEHisTEV
11	DtHNL3/4_Dt_NcoI_fw	AATGCCCATGGCAGGAACGGGAGGGGGC	Cloning pEHisTEV
12	DtHNL3/4_Dt_HindIII_rev	AATGCAAGCTTTTAATCACCGAGGAGTGTAACAG	Cloning pEHisTEV
13	PtaIsotig02775_Ec_NcoI_fw	AATGCCCATGGAAACCATTCAGACC	Cloning pEHisTEV
14	PtaIsotig02775_Ec_HindIII_rev	AATGCAAGCTTTTACGGAATCAGGGTAATAC	Cloning pEHisTEV
15	pEHisTEV_DtHNL1_gibson_fw	TACCCATCTGAGCATTCCGGAAAG	Gibson cloning pEHisTEV
16	pEHisTEV_DtHNL1_gibson_rev	CAGCTATATGCTTTACCCCACAGAAC	Gibson cloning pEHisTEV
17	pEHisTEV_PtaIsotig02775_gibson_fw	ATATTGCATTTAGCCTGTATCAGGCAG	Gibson cloning pEHisTEV
18	pEHisTEV_PtaIsotig02775_gibson_rev	AGGCTTTACCCCAAAGAACCTCTTC	Gibson cloning pEHisTEV
19	pEHisTEVseq1	CTTTAATAGTGGACTCTTGTTC	Sequencing
20	pEHisTEVseq2	GTTTATGCATTTCTTTCCAGAC	Sequencing
21	pEHisTEVseq3	GCAAGACGTTTCCCGTTGAATATG	Sequencing
22	pEHisTEVseq4	AGATACCTACAGCGTGAGCTATG	Sequencing
23	pEHisTEVseq5	GTGACTGGGTCATGGCTGCG	Sequencing
24	pEHisTEVseq6	TTCCACAGGGTAGCCAGCAGCATC	Sequencing
25	pEHisTEVseq7	TTGAAGGCTCTCAAGGGCATCG	Sequencing
26	pEHisTEVseq8	CGGCTGAATTTGATTGCGAGTG	Sequencing
27	pEHisTEVseq9	CACTTTTTCCCGCGTTTTCGCAG	Sequencing
28	pEHisTEVseq10	GGAATTGTGAGCGGATAACAATTC	Sequencing

Standard T7 forward and reverse primers provided by LGC Genomics or Microsynth AG were employed for the sequencing of CDS in pEHisTEV plasmid.

Supplementary Table 13. gBLock® Gene fragments list

Entry	Name	Sequence
1	DtHNL1_R69A	GTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCACCATTGATAAAGTTTGGA GCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCA TGTTGTTAGCGGTGAAGCAAATCAGACCGGTTGTGTTGCACGTTTTGTTTGTTATCCGGCA AGCGAAGGTGAAAGCGAAACCGTTGATTATAGCGAACTGATTCACATGAATGCAGCAGCAC ATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTA CGTGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTAT TGGAGCTTTACCGCAGAACCGGCAAGCAATCTGACCGAACAGAAATGTATTGAAATTGTGT TTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAAAG
2	DtHNL1_D85A	GTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCACCATTGATAAAGTTTGGA GCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCA TGTTGTTAGCGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCA AGCGAAGGTGAAAGCGAAACCGTTGCATATAGCGAACTGATTCACATGAATGCAGCAGCAC ATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTA CGTGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTAT TGGAGCTTTACCGCAGAACCGGCAAGCAATCTGACCGAACAGAAATGTATTGAAATTGTGT TTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAAAG
3	DtHNL1_S87A	GTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCACCATTGATAAAGTTTGGA GCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCA TGTTGTTAGCGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCA AGCGAAGGTGAAAGCGAAACCGTTGATTATGCAGAACTGATTCACATGAATGCAGCAGCAC ATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTA CGTGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTAT TGGAGCTTTACCGCAGAACCGGCAAGCAATCTGACCGAACAGAAATGTATTGAAATTGTGT TTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAAAG
4	DtHNL1_Y101A	GTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCACCATTGATAAAGTTTGGA GCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCA TGTTGTTAGCGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCA AGCGAAGGTGAAAGCGAAACCGTTGATTATAGCGAACTGATTCACATGAATGCAGCAGCAC ATCAGTATATGGCAATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTA CGTGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTAT TGGAGCTTTACCGCAGAACCGGCAAGCAATCTGACCGAACAGAAATGTATTGAAATTGTGT TTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAAAG
5	DtHNL1_Y117A	GTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCACCATTGATAAAGTTTGGA GCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCA TGTTGTTAGCGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCA AGCGAAGGTGAAAGCGAAACCGTTGATTATAGCGAACTGATTCACATGAATGCAGCAGCAC ATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACGC AGTGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTAT TGGAGCTTTACCGCAGAACCGGCAAGCAATCTGACCGAACAGAAATGTATTGAAATTGTGT TTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAAAG
6	DtHNL1_Y161A	GTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCACCATTGATAAAGTTTGGA GCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCA TGTTGTTAGCGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCA AGCGAAGGTGAAAGCGAAACCGTTGATTATAGCGAACTGATTCACATGAATGCAGCAGCAC ATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTA CGTGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTAT TGGAGCTTTACCGCAGAACCGGCAAGCAATCTGACCGAACAGAAATGTATTGAAATTGTGT TTCCGCTGGCAACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAAAG
7	DtHNL1_Y101F	GTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCACCATTGATAAAGTTTGGA GCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCA TGTTGTTAGCGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCA AGCGAAGGTGAAAGCGAAACCGTTGATTATAGCGAACTGATTCACATGAATGCAGCAGCAC ATCAGTATATGTTTATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTA CGTGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTAT TGGAGCTTTACCGCAGAACCGGCAAGCAATCTGACCGAACAGAAATGTATTGAAATTGTGT TTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAAAG
8	DtHNL1_Y117F	GTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCACCATTGATAAAGTTTGGA GCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCA TGTTGTTAGCGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCA AGCGAAGGTGAAAGCGAAACCGTTGATTATAGCGAACTGATTCACATGAATGCAGCAGCAC ATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTT TGTGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTAT TGGAGCTTTACCGCAGAACCGGCAAGCAATCTGACCGAACAGAAATGTATTGAAATTGTGT TTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAAAG

Entry	Name	Sequence
9	DtHNL1_Y161F	GTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCACCATTGATAAAGTTTGGA GCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCA TGTTGTTAGCGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCA AGCGAAGGTGAAAGCGAAACCGTTGATTATAGCGAACTGATTCACATGAATGCAGCAGCAC ATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTA CGTGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTAT TGGAGCTTTACCGCAGAACCGGCAAGCAATCTGACCGAACAGAAATGTATTGAAATTGTGT TTCCGCTGTTTACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAAAG
10	DtHNL1_ D85S_S87D	GTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCACCATTGATAAAGTTTGGA GCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCA TGTTGTTAGCGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCA AGCGAAGGTGAAAGCGAAACCGTTAGCTATGATGAACTGATTCACATGAATGCAGCAGCAC ATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTA CGTGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTAT TGGAGCTTTACCGCAGAACCGGCAAGCAATCTGACCGAACAGAAATGTATTGAAATTGTGT TTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAAAG
11	DtHNL1_R69K	GTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCACCATTGATAAAGTTTGGA GCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCA TGTTGTTAGCGGTGAAGCAAATCAGACCGGTTGTGTTAAACGTTTTGTTTGTTATCCGGCA AGCGAAGGTGAAAGCGAAACCGTTGATTATAGCGAACTGATTCACATGAATGCAGCAGCAC ATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTA CGTGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTAT TGGAGCTTTACCGCAGAACCGGCAAGCAATCTGACCGAACAGAAATGTATTGAAATTGTGT TTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAAAG
12	Ptaiso02775_ A92S	AGCAAGCCGTAGCTATGGTGAAGAAGAGGTTCTTTGGGGTAAAGCCTTTAAATGGGAAATT AAAGGTGTGGGCGAAGATGAAGTTTGGGAAGTTACCGGTGATTTTCTGGGTGTTGCACGTT GGGCAACCAGCCTGGTGGAAAGCTGTGAACTGATTGAAGGTGAAGCACATAAACCGGGTTG TGTTCGTCGTGTTCTGGTTTATCCGCAGGCACCGGGTGAAGCAAGCACCTTTAGCCTGGAA AAACTGCTGGAAATGGATGCACTGCATCATCGTTATAGTTATACCATTCTGGGTGGTAGCA CCCTGCCTGGTTTTAGCCTGATGCAGGATTATGTTAGCACCTTTAAACTGAGCAGCCTGCG TCTGGTGTATCCGAGCGCAGAAATTGATCAAGAAAATGGCACCCTGCTGCATTGGAGCTTT GTTTGTCGTCCGGTGAGCACCCTGAGCGAAGAAGAAACCCATAATATTGCATTTAGCCTGT ATCAGGCAGCCG
13	Ptaiso02775_ A92D_E94S	AGCAAGCCGTAGCTATGGTGAAGAAGAGGTTCTTTGGGGTAAAGCCTTTAAATGGGAAATT AAAGGTGTGGGCGAAGATGAAGTTTGGGAAGTTACCGGTGATTTTCTGGGTGTTGCACGTT GGGCAACCAGCCTGGTGGAAAGCTGTGAACTGATTGAAGGTGAAGCACATAAACCGGGTTG TGTTCGTCGTGTTCTGGTTTATCCGCAGGCACCGGGTGAAGCAAGCACCTTTGATCTGAGC AAACTGCTGGAAATGGATGCACTGCATCATCGTTATAGTTATACCATTCTGGGTGGTAGCA CCCTGCCTGGTTTTAGCCTGATGCAGGATTATGTTAGCACCTTTAAACTGAGCAGCCTGCG TCTGGTGTATCCGAGCGCAGAAATTGATCAAGAAAATGGCACCCTGCTGCATTGGAGCTTT GTTTGTCGTCCGGTGAGCACCCTGAGCGAAGAAGAAACCCATAATATTGCATTTAGCCTGT ATCAGGCAGCCG

Supplementary Table 14. Isotigs List

Entry	Sequence
1	>DtIsotig06604 TTTTAAACTTGTGTAGCAATAAATATACAAATATAAGTGATACACATCAACGAGTTTCGGTAGTAATTTATGTACAAA CTGCACATAAATTAATTATTGCACAGATAACAAAATGGGGAGAACCCCTACAGTATGCTGCAAAATTTGCAATATACA TCACATCGAGGGTACCCACCATCATCTAAAAGCTTGCAGTTACTTGACTGAAGAGAATCTTGAAATCTGTCGTCGGCG TACCTTCACCCTCCGGTTTCGAGGTCAAGTCAAAAGCCTTATTCTTTACTGCATCATGAACAATTGCCTGTACACAGA CTTCCGCAACATCTGCTCGTGGAACCGTTTTGGTATCTGTGGCGAGAAGTTCGTCGTTGTTTCCAATCAAGAGCTCTC GCAACCCACCTTCTTTGTCTAGCAGCCCACCCGGCCTGATTAAGGTGTAGGTTGTATCAGAGTCTGCCAAATACTGCT CCGACTTCCTTTTCCAAATCAAAATCTTCCCATTTCCGAAAGAGTTCAGAGGATGATTCGGATTTGTGCCACCCATCG AGCCAACCAAAATGATATGCTTCACCCCCGCTTCTTTGGCTGCATCGATTTGATTCTTTTGCCCGATCCAATCAACCT GCTCAGGATATGCGCCTTCTTCGTAGTAGAACTCGGGGCGCCCTCCTTTGCTTGGATCAAACCCTGGCTTCATTTGAG GAGCAGCACTTGTGAGGATGATGAGCACATCCACCCCTGCAAATGCTGGGCCCAGGCTCTCTGCCTTTGTTATGTCAC CTATGTACACATCCTCACCCCCTCCAATCTTTGCCTTGCTTGGCTCTGACCTCACAAGACCTCTCACATGAAACTTGT CTGCCTTATGTTTCAGCTTCTCATATACAATGTGCCCCGTTCTTCCACCAGCACCAGTTACGAGTACGGTTGTGCGAG CAGAGTCTGCCATTGAAGCTGCAAGAAATGAAGATGGCGCAAAATGCTGAGACGAGGAGATGGCGCGACAAACCCTAT ATAAGCGTAATCTCAAGGTCTTCCAGAAGCTGCCATTTGTCGGCGATTTTTGGGCGGTGGGCCTTTCCTAATTCTGTG TGTTCA
2	>DtIsotig07200 ATCTGCCTCCATTCAGAGCTATTCTTAGCCTGTGCTCCCCACGCTCCCTCCCTCCTCGGATTATCACCCCTCACTTAC GCAGTATTATTTGGCGTGTTTTTTCATATTTGATATCGAGAGCAGCCATGAGTTATTGGAAGAGCAAGGTTGTGCCCA AAATCAAGAAGTTTTTTGACAAGGGGAAGAAGAAAGGAGCTGCTGAGTTCTCCAAAAACTTTGATTCGTCCAAGGAGT CTTTGGACAAGGAGATAGGAGAGAAGAGTTCAGATCTGAGCCCCAAGGTTGTGGAGATATACAGATCCTCTTCCACCT TCATTGCCAAGAAGTTGCTGAAGGAACCCAATGAGGCAACAGTGAAGGAAAATTCCGATGCAACGCAAGGCGTGCTTC AGGAGCTGGCAACAGCAGGCTTTCCTGGAGCGCAGGGCATCGCTGATGCAGGCAAAAAGTATGGACCGGCGCTTTTAC CGGGGCCGGTCGTGTACTTGTTCGAGAAAGCATCCGTGTTTTTGGCGGAGGAGCCTTTGCCAGAGGAGCCCAAGGCAG AGACTAGAGAGGTGAGTGCGGAGGACGTCAAGCCAGCAGAAGCGCCAGCGACGACTTCGGAAACTCCGCCGCCGCCAC CTGTAGCAGACGTGCCTCCCCCAGCTGTCGTCGAGGAAGAAAAGAAGGAAGCAGAGCCCATTGTTGCTGCCCCGCCGC CCGAAGCTGCCCCCCCTGCAGCTGTTGAGATCCCCACCTCCGTTGACCCTACGCCCCCGCCTCCTGCTCCGCCCGCCG ACAAACCTGAGTAGTTTTCGCCTTCGATGCATAGTCGCACACCCATCCACACATCCATTGTCTTCTGTGCATTCATAA ATTTTAATTACAATGTCCTTGTTCATAGTTTAATTACAATTTCTTTATTTTCATATATAATCA
3	>DtIsotig04065 GTAAAAAGTTATGGTTTTTTCCATTGATACTTATCTAGAGACGTGTAAGAAGAAAAAAATCGCTTTAAACGGGTAATT TAAGAGTATCTTTTATTCAATGCATAACAACAATAGCTGCCTGACTAGCCAGGCAGCTGGGAAATTCATTTGGGAGAG AAGCACAAACACACTGAATTGTAGCAGCTAGCTAGTATACAAAGTCAATGGCATAAAAGAACCATGAGACCGTTTCCA TGTTTACAGTAGCGGTTGAGACAAGGGACTGTTGAGCCACTCCGCGCACATACGCAAAGTCGCCCGTGCCGCCAACAA TGGGTGCAAACTTTACCGCATCTGACGCGACGTCCACCCCCAACAGGCTGAATGTGCCATTGTAGGTGCCGTCGTTCC AGGTAAAGGTCTGCGCCAGGAACAACGTCAAAAGATTCTGCCCCACATCACCATACCAGCCCTGCACGTGCCCGAGCT GCGTAGAGTTGAGATCTGCGGCACTGGTGAGTGGGTTGTCGAAGGTGTGGATGATGCCGAATGAGTTGGGTTGCGTCG AAAGCGGCTGCGCGGACTGCACGGCTGTGAAGGTGACATTGGGGTTGTCGAGATTGCTGTTATTCTGAACAGCAATGT ACATGTAGAAGTCAAGATGGTCAGGACCCGAGGCAGCCAAAGCTGGCTGTGGGGCAGAGAGGGAGGAGAGGGAGAGAG CGCATAATGTACATAGGAGCATAATGTGCACCTTGGCCATGAGATCGACAGAGGAGGGGGAGGCTGAGAGTGAGCTTG TAGGGAGACGCGCGGACAATACGTCAAGTAGGAGTGTATGTAAGCTGTCGACAAAAGCTGGGCTTGTGTGTGTATGAG AGGTAGAGAAGAAGAGAAGACTGTTATAGGTGATGAATGAGTTCTTCCATTTATAACGTTCTGGCCCTTCTGAGTGGA TCAGACGTGCTTTGACAAGTCGCTCTACTGTGAGTGCCGGTGTGTGTGTGTGCGCCGGATCGGAGGTCAAAGTGAGAT TCCGGAAGCCTGATGTATTTCATAGAGAAAGCGATAAACGTTACAGACTGAA
4	>DtIsotig04379 CTCTGGACCTTGGCTGTCACTCAAATTTTtGTTTAGAGGAGGAGGAAATGGGCACCTCAACGTGGGTGGTATGGAGCA TTCTGTTGCTAGCAGTGGCGCAAGTAGCGGGGAGCATCCCCATCCCAAGACGCTACGATGGCTTCGTCTTCAATGCTT CTTCATCCTCGTCGCCTGTGTTGCTGGAGGCCTTCTTCGATCCCCTCTGCCCGGATAGCGCAGACGCTTGGCCTGTTG TCAAGAAAATCGCCCAATACTTCCAGGACGATCTGCTCCTCATTGTCCACCCCTTCCCTCTCCCGTACCATCACAATG CATATTTTGCAAGTAGAGCATTGCACATCATCAATAACCTGAACAGTTCTCTCACTTATCCATTGCTTGAGTTGTTTT TTGAAAACCAGGATAGCTTTTCAACGAGTGAAACGCTAGCGGAGGCACCATCCTCCGTCGTAGACAGAATCG5TTCAA CTGGCAGCAGATAGCTTGAATGAACTCGTGTCTTCCGATTTTGAGAGCCAGTTCAAAGCAG6GATTTTCTGATACAGG AACGGATCTGATCACTCGTGTTTCGTTCAAGTTTGGGTGCTCGCGCGTTGTGGTCGGTACGCCGTACTTTTTTGTCAA TGGCATACCTCTTTATGATGCGGATTCGGCATGGACCTTCTCTGAATGGGCAGAGATTATTGAGCCATTAACAGGGGC GCAAAAAATCGCGCTTGCATAACGACCTCGGCAACACCAAAGCTGTATATTTGATCATTTAAGGgTGCtaCATtGCAT GGAAGgTGCAGATACACTTTGCTTTtCTACTTGATTTtCctATCTCACAGTCTAATTTTAtACCaTAT

Entry	Sequence
5	>DtContig00505 CTCTGGACCTTGGCTGTCACTCAAATTTTTGTTTAGAGGAGGAGGAAATGGGCACCTCAACGTGGGTGGTATGGAGCATT CTGTTGCTAGCAGTGGCGCAAGTAGCGGGGAGCATCCCCATCCCAAGACGCTACGATGGCTTCGTCTTCAATGCTTCTTC ATCCTCGTCGCCTGTGTTGCTGGAGGCCTTCTTCGATCCCCTCTGCCCGGATAGCGCAGACGCTTGGCCTGTTGTCAAGA AAATCGCCCAATACTTCCAGGACGATCTGCTCCTCATTGTCCACCCCTTCCCTCTCCCGTACCATCACAATGCATATTTT GCAAGTAGAGCATTGCACATCATCAATAACCTGAACAGTTCTCTCACTTATCCATTGCTTGAGTTGTTTTTTGAAAACCA GGATAGCTTTTCAACGAGTGAAACGCTAGCGGAGGCACCATCCTCCGTCGTAGACAGAATCG5TTCAACTGGCAGCAGAT AGCTTGAATGAACTCGTGTCTTCCGATTTTGAGAGCCAGTTCAAAGCAG6GATTTTCTGATACAGGAACGGATCTGATCA CTCGTGTTTCGTTCAAGTTTGGGTGCTCGCGCGTTGTGGTCGGTACGCCGTACTTTTTTGTCAATGGCATACCTCTTTAT GATGCGGATTCGGCATGGACCTTCTCTGAATGGGCAGAGATTATTGAGCCATTAACAGGGGCGCAAAAAATCGCGCTTGC ATAACGACCTCGGCAACACCAAAGCTGTATATTTGATCATTTAAGGGTGCTACATTGCATGGAAGGTGCAGATACACTTT GCTTTTCTACTTGATTTTCCTATCTCACAGTCTAATTTTATACCATAT
6	>DtIsotig02643 CTCTGGACCTTGGCTGTCACTCAAAAGAGAAGGTAAGGAGAGGAAAGCCATTAGGGGTTTGAGGAGAGTGAGGCGAGGTA GGGATTATTAAGGGGGCAGCTCCCTAGCAAGTCATGGCGGGAACGGGAGGGGGCGCAGAACAGTTCCAGCTCCGGGGAGT GCTGTGGGGGAAAGCCTACTCTTGGAAGATAACCGGAACGACAATCGACAAGGTGTGGTCGATTGTGGGCGATTATGTGC GCGTCGACAACTGGGTCTCTTCCGTCGTGAAGAGCTCGCACGTCGTGTCTGGCGAGGCCAACCAGACGGGGTGCGTGAGG AGGTTCGTCTGCTACCCAGCCTCCGAGGGAGAGTCGGAGACTGTGGACTACTCGGAGCTCATCCACATGAACGCTGCCGC GCACCAGTACATGTACATGATCGTGGGAGGTAACATCACTGGCTTCTCTCTCATGAAGAACTATGTGAGCAATATCTCGC TGTCTTCTCTTCCTGAGGAGGACGGTGGTGGTGTAATCTTTTACTGGAGCTTCACAGCCGAGCCTGCCTCTAACCTCACG GAACAAAAATGCATAGAAATTGTGTTTCCTCTCTACACCACTGCCCTGAAGGATTTATGCACTCACCTTTCCATACCCGA AAGCTCTGTTACACTTCTCGATGATTAAGCTTTTCATCCTCCCCCCCCCCAAAAAAAAACAAAACAAAACAAAACAAACT TTCTGCGTTCCCTAATACTTGTGCTGCGTTTCCTATCTTTTCCTTAATGCTCTACATCTTGCGCTTCCTTACCCTAATAA TTGCCGCTTTACACTTTCTAATAGATTTTCTGCTTTCTTGACTTGCCAAGGACTAAATATAT
7	>DtIsotig02641 GTCTCTGGACCTTGGCTGTCACTCAAAAAAGAGAGGCGGAGGGAAGCCATTAGGAGTTGTGGGTTGAGAGAGTGAGGCGA GGTAGGGATTATTAAGGGGGCAAGCCATGGCGGGAACGAGAGGAGGCGCTGAAGAGTTCCAGCTCCGGGGAGTGCTGTGG GGGAAAGCCTACTCTTGGAAGATAACGGGAACGACAATCGACAAGGTGTGGTCGATTGTGGGTGATTATGTGCGCGTCGA CAACTGGGTCTCTTCCGTCGTGAAGAGCTCGCACGTCGTGTCCGGCGATGCCAACCAGACGGGGTGCGTGAGGAGGTTCG TCTGCTACCCAGCCTCCGATGGAGAGTCGGAGACTGTGGACTACTCGGAGCTCATCCACATGAACGCCGCCGCTCACCAA TACATGTACATGATTGTGGGAGGTAACATCACTGGCTTCTCTCTCATGAAGAACTATGTGAGCAATATCTCGCTGTCTTC TCTTCCTGAGGAGGACGGTGGTGTGTCAATCTTTTACTGGAGCTTCACAGCCGAGCCTGCCTCTAACCTCACGGAACAAA AATGCATAGAAATTGTGTTTCCTCTCTACACCACTGCCCTGAAGGATTTATGCACTCACCTTTCCATACCCGAAAGCTCT GTTACACTTCTCGATGATTAAGCTTTTCATCCTCCCCCCCCCCAAAAAAAAACAAAACAAAACAAAACAAACTTTCTGCG TTCCCTAATACTTGTGCTGCGTTTCCTATCTTTTCCTTAATGCTCTACATCTTGCGCTTCCTTACCCTAATAATTGCCGC TTTACACTTTCTAATAGATTTTCTGCTTTCTTGACTTGCCAAGGACTAAATATAT
8	>DtIsotig07602 AGAATTAATGGTAGTAGGAACGAGTTGGTCCTGCCAACTATGTGCATATATTTAATAAAACCTTGGCAAACAAAACATTA CTAACGGTATAGCAGCAATTATTAGGTGAGGTAGCTGCCGAGTTGGCTAAGGTAGAAAATGTAATTAGGGGGGTGAGATA AAGCTTAATCACCGAGGAGTGTAACAGAGCTTTCCGGAATAGAAAGGTGAGTGCATAAATCCTTCAAGGCAGTGGTATAG AGAGGGAACACAATTTCTATACATTTTTGTTCGGTGAGGTTAGAAGCAGGCTCGGCTGTGAAGCTCCAGTGGAGGATGAC ACCACCACCGTCCGCCTCAGGAAGAGAATTGAGCGATATATTGCTCACATAGTTCTTCATGAGAGAGAAGCCAGTGATGT TACCTCCCACAATCATGTACATGTACTGGTGCGCGGCCGCGTTCATGTGGATGAGCTCCGAGTAGTCCACAGTCTCCGAC TCTCCCTCGGAGGCTGGGTAGCAGACGAACCTCCTCACGCACCCCGTCTTGTTAGCGTCGCCAGACACGACGTGCGAGCT CTTCACTACAGAAGAGACCCAGTTGTCGACGCGCACATAGTCGCCCACAATCGCCCACACCTTGTCAATTGTCGTTCCCG ATATCTTCCACGAGTAGGCTTTCCCCCACAGCACTCCCCGCAGCTGGAACTCTTCTGCGCCCCCTCCCGTTCCTGCCATG GCTTGCAAGCTTGCCTACTGCTCCTTACTCTCTCTCACTCTCTACCCTCAGCTCCCTCCTCCGACTCTACTTTCCTTCTC TACTTCCCCAGGCTGATTGCAGCTGCTTTATAAAGTTT
9	>DtContig00751 ATTATTTAGCGGTGTAGCAGCAATTATTAGCTGAGGACATTAAGGGGTAGCTGCCGAGTTGGCTAAGGTAGAAAATATAA TTAGGGAGGGGTGAGATAAAGCTTAATCACCGAGGAGTGTAACAGAGCTTTCCGGGATAGAAAGGTGAGTGCATAAATCC TTCAAGGCAGTGGTATAGAGAGGGAACACAATTTCGATGCATTTTTGTTCGGTGAGGTTAGAAGCAGGCTCGGCTGTGAA GCTCCAGTGGAAGATGACACCACCTCCGTCCGCCTCAGGAAGAGAATTGAGCGATATATTGCTCACATAGTTCTTCATGA GAGAGAAGCCAGTGATGTTACCTCCCACGATCATGTACATGTACTGGTGCGCGGCCGCATTCATGTGGATGAGCTCCGAG TAGTCCACAGTCTCCGACTCTCCCTCGGAGGCTGGGTAGCAGACGAACCTCCTCACGCACCCCGTCTTGTTCGCATCGCC AGACACGACGTGCGAGCTCTTCACTACGGAAGAGACCCAGTTATCGACGCGGACATAGTCGCCCACAATCGACCACACCT TGTCGATTGTCGTTCCCGTTATCTTCCAAGAGTAGGCTTTCCCCCACAGCACTCCCCGCAGCTGGAACTCTTCTGCGCCC CCTCCCGTTCCTGCCATGGCTTGGAAGCTTGCCTACTGCCCCTTACTCTCTCTCACTCTCTCAATGCGCAGATCCCTCTT CCGGCTCTACTTACCTTATCTTTTGAGTGACAGCCAAGGTCCAGAGACGAGT

Entry	Sequence
10	>PtaIsotig02775 CTCTGGACCTTGGCTGTCACTCAAAGTGTGTGGCATTCACCGTAAGGAGGAGAGTTAAGGTGAGGCATCTGAGGTGCGGT TGGGGCAACTTGGAGGACAGCGAGCTGTGTAAGGGAAGGATATTACAGCACAGGTGATGGAGACGATTCAAACAGCGACG GAGTCGATGACAGCGGCTAGCAGGAGCTATGGAGAGGAGGAGGTATTATGGGGGAAGGCGTTCAAGTGGGAGATAAAGGG TGTAGGGGAGGACGAGGTGTGGGAGGTAACCGGAGACTTTCTGGGAGTGGCCAGGTGGGCAACCTCGCTGGTGGAGAGCT GTGAGCTTATAGAAGGAGAGGCCCATAAGCCAGGCTGCGTGAGAAGGGTCCTTGTTTATCCCCAGGCTCCTGGGGAGGCC TCCACTTTTGCCCTTGAAAAGCTCTTAGAAATGGACGCGCTACACCACCGTTACTCTTACACTATCCTTGGCGGAAGCAC CTTGCCTGGCTTCTCTCTCATGCAGGACTATGTCAGTACCTTCAAGCTCTCTTCCCTACGCCTGGTGTACCCCTCTGCAG AAATTGACCAAGAAAATGGTACCCTCCTCCATTGGAGCTTTGTTTGTCGCCCAGTCTCTACCTTGTCTGAGGAGGAAACC CACAACATTGCCTTCTCTCTCTACCAGGCTGCAGTCAACGATCTCAAAGCTCGCCTCTCCTTGTCTGACGACCGCATTAC TCTCATCCCGTAAACGTCTTTACCTAGCTAGGTGTGCGGTTGCAGGTAGTGGATTGCAAGCCCTAAATGCCGATAGTCCC CCGCCTCCCAACCGTGCGCAGTGCATACCTAGGTATGCAGGGTCTCCTCAACAAAAGCCCATGATGTTGTAGGCCACAAA GTGGAGATTACATTGGTAATTTGCAGGGTGAACTATGTATAAATGCTCTTCCTGGTCATCTTAAAAGGTACCATAGTTGA AGGAGTTGCAAAGGGAGTGGTTCCTCACAGTAAAGTAAAAGTGTTATGTGAGTAAAAGTGTATCAGATAACGTAAGCCTT GTTGTTAAGGTGGTTGCTTAGT
11	>PtaIsotig02778 CTCTGGACCTTGGCTGTCACTCAAAAAGGTGCGGTTGGGGCAACTTGGGGGGCAGCGAGCTGTGTAAGGGAAGGATATTT CAGCACAGGTGATGGAGACGATTCAAACAGCGGCTAGCAGGAGCTATGGAGAGGAGGAGGTATTATGGGGGAAGGCGTTC AAGTGGGAGATAAAGGGTGCAGGGGAGGACGAGGTGTGGGAGGTAACCGGAGACTTTCTGGGAGTGGCCAGGTGGGCAAC CTCGCTGGTGGAGAGCTGTGAGCTTATAGAAGGAGAGGCCCATAAGCCAGGCTGCGTGAGAAGGGTCCTTGTTTATCCCC AGGCTCCTGGGGAGGCCTCCACTTTTGCCCTTGAAAAGCTCTTAGAAATGGACGCGCTACACCACCGTTACTCTTACACT ATCCTTGGCGGAAGCACCTTGCCTGGCTTCTCTCTCATGCAGGACTATGTCAGTACCTTCAAGCTCTCTTCCCTACGCCT GGTGTACCCCTCTGCAGAAATTGACCAAGAAAATGGTACCCTCCTCCATTGGAGCTTTGTTTGTCGCCCAGTCTCTACCT TGTCTGAGGAGGAAACCCACAACATTGCCTTCTCTCTCTACCAGGCTGCAGTCAACGATCTCAAAGCTCGCCTCTCCTTG TCTGACGACCGCATTACTCTCATCCCGTAAACGTCTTTACCTAGCTAGGTGTGCGGTTGCAGGTAGTGGATTGCAAGCCC TAAATGCCGATAGTCCCCCGCCTCCCAACCGTGCGCAGTGCATACCTAGGTATGCAGGGTCTCCTCAACAAAAGCCCATG ATGTTGTAGGCCACAAAGTGGAGATTACATTGGTAATTTGCAGGGTGAACTATGTATAAATGCTCTTCCTGGTCATCTTA AAAGGTACCATAGTTGAAGGAGTTGCAAAGGGAGTGGTTCCTCACAGTAAAGTAAAAGTGTTATGTGAGTAAAAGTGTAT CAGATAACGTAAGCCTTGTTGTTAAGGTGGTTGCTTAGT
12	>PtaContig4149 AGACTCCTTCAACTATGGTACCTTTAAAGATGACCAGGAAGAGCATTTATACATAGTTCACCCTGCAAATTACCAATGTA ATCTCCACTTTGTGGCCTACAACATCATGGGCTTTTGTTGGGGAGACCCTGCATACCTAGGTATGCATTGCGCACTGTTG GGAGGCGGGGGACTATCGCCATTTAGGGCTTGCAATCCACTACCTGCAACCACACACCTAGCTAGGTAAAGATGCTTACG AGATGAGAGTAATGCGGTCGTCAGACAAGGAGAGGCGAGCTTTGAGATCGTTGACTGCAGCCTGGTAGAGAGAGAAGGCA ATGTTGTGGGTTTCCTCTTCAGACAAGGTAGAGACTGGGCGACAAACGAAGCTCCAATGGAGGAGGGTACCATTTTCTTG GTCAATTTCTGCAGAGGGGTACACCAGGCGTAGGGAAGAGAGCTTGAAGGTACTGACATAATCCTGCATGAGAGAGAAGC CAGGCAAGGTGCTTCCGCCAAGGATAGTGTAAGAGTAATGGTGGTGTAGGGCGTCCATTTCTAAGAGCTTTTCAAGGGCA AAAGTGGAGGCTTCCCCAGGAGCCTGGGGATAAACAAGGACCCTTCTCACGCAGCCTGGCTTATGGGCCTCTCCTTCTAT AAGCTCACAGCTCTCCACCAGCGAGGTTGCCCACCTGGCCACTCCCAGAAAGTCTCCGGTTACCTCCCACACCTCGTCCT CCCCTACACCCTTTATCTCCCACTTGAACGCCTTCCCCCATAATACCTCCTCCTCTCCATAGCTCCTGCTAGCCGCTGTC ATCGACTCCGTCGCTGTTTGAATCGTCTCCATCACCTGTGCTGTAATATCCTTCCCTTACACAGCTCGCTGTCCTCCAAG TTGCCCCAACCGCACCTCAGATGCCTCACCTGTGTTTACCTTAACTCTCCTCCTTACGGTGAATGCCACACCCCCGCGT

References

1. Krammer, B., Rumbold, K., Tschemmernegg, M., Pöchlauer, P. & Schwab, H. A novel screening assay for hydroxynitrile lyases suitable for high-throughput screening. *J. Biotechnol.* **129**, 151–61 (2007).
2. 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).
3. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst*. *Biol*. **7**, (2011).
4. 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).
5. Lanfranchi, E. et al. Bioprospecting for hydroxynitrile lyases by Blue Native PAGE coupled HCN detection. *Current Biotech*. **4**, 111–117 (2015).
6. Radauer, C., Lackner, P. & Breiteneder, H. The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol. Biol*. **8**, 286 (2008).
7. Finn, R. D. et al. Pfam: The protein families database. *Nucleic Acids Res*. **42**, 222–230 (2014).
8. Imai, S. et al. Plant biochemistry: an onion enzyme that makes the eyes water. *Nature* **419**, 685 (2002).
9. Yin, P. et al. Structural insights into the mechanism of abscisic acid signaling by PYL proteins. *Nat. Struct. Mol. Biol*. **16**, 1230–1236 (2009).
10. Lim, C. W. & Lee, S. C. *Arabidopsis* abscisic acid receptors play an important role in disease resistance. *Plant Mol. Biol*. **88**, 313–324 (2015).
11. Kuipers, R. K. et al. 3DM: systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins* **78**, 2101–13 (2010).
12. Consortium, T. U. UniProt: a hub for protein information. *Nucleic Acids Res*. **43**, D204–D212 (2014).
13. Masamura, N. et al. Identification of amino acid residues essential for onion lachrymatory factor synthase activity. *Biosci. Biotechnol. Biochem*. **76**, 447–453 (2012).

14. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc*. **10**, 845–858 (2015).

Chapter 4

Production and Immobilization of the Hydroxynitrile Lyase from White Rabbit's Foot Fern (*Davallia tyermannii*) for a robust biocatalyst generation

Elisa Lanfranchi[1], Zainab Raghoebar[2], Sander Van Pelt[2], Roger A. Sheldon[2], Margit Winkler[1,3]*, Anton Glieder[1,3]

[1]Austrian Centre Industrial Biotechnology, Petersgasse 14, 8010 Graz, Austria
[2]CLEA Technologies B.V. Delftechpark 34 2628XH Delft, The Netherlands
[3]Graz University of Technology, Institute of Molecular Biotechnology, NAWI Graz, Petersgasse 14, 8010 Graz, Austria

Corresponding Author:
Margit Winkler
Austrian Centre Industrial Biotechnology, Petersgasse 14/III, 8010 Graz Austria
Email: margit.winkler@acib.at; Tel: +43-316-879333; Fax: +43-316-87393

Abstract

Hydroxynitrile lyase from white rabbit's foot fern *Davallia tyermannii* (*Dt*HNLs) catalyze the enantioselective synthesis of different α-cyanohydrins, key building blocks for pharmaceutical and agrochemical industries. An efficient and competitive process has to take the availability and robustness of the biocatalyst into account. Herein, the recombinant production of *Dt*HNL1 in *Pichia pastoris* is described. *Dt*HNL1 constitutes approximately 80% of the total protein content, therefore no expensive protein purification is necessary for subsequent biocatalytic applications. After *Dt*HNL production, the crude enzyme can immediately be immobilized in the form of a cross-linked enzyme aggregate (CLEA), which resulted in a drastic enhancement of the biocatalyst stability under acidic conditions. *Dt*HNL1-CLEA retained its activity at pH 2.4 after 168 hours. The *Dt*HNL1-CLEA was employed for the synthesis of (*R*)-mandelonitrile in a biphasic system: 99% conversion and 98% enantiomeric excess were achieved. Furthermore, 18% enantiomeric excess in the synthesis of (*R*)-hydroxypivaldehyde cyanohydrin was obtained with *Dt*HNL1-CLEA whereas a very low enzymatic stereoselectivite conversion was obtained with the soluble crude enzyme. The results show *Dt*HNL1-CLEA to be a powerful and robust biocatalyst for the synthesis of enantiomerically pure cyanohydrins under acidic conditions.

Keywords

Hydroxynitrile lyase; cyanohydrin; cross-linked enzyme aggregate (CLEA); *Pichia pastoris*; *Davallia tyermannii*; protein production; mandelonitrile

Introduction

Hydroxynitrile lyases (HNLs, EC 4.1.2.10, EC 4.1.2.11, EC 4.1.2.46, and EC 4.1.2.47: alternative names: oxynitrilase, hydroxynitrilase, mandelonitrile lyase, and hydroxymandelonitrile lyase) catalyze the enantioselective cleavage of cyanohydrins into the corresponding aldehyde or ketone and hydrogen cyanide. The major number of known HNLs has been isolated from diverse plants, where they play a role in the chemical defense system[1,2]. Additional HNLs have been isolated from bacteria[3,4] and animals [5].

HNLs find interest in biocatalysis, due to their ability to synthesize enantiomerically pure α-cyanohydrins, key building blocks for pharmaceuticals and agrochemicals. The introduction of a nitrile group into the molecule provides the possibility of a variety of modifications on both the nitrile and the hydroxyl-group and allows chemical or chemo-enzymatic follow-up reactions[1,2,6,7]. Several other compounds can be easily derived from cyanohydrins (see table in related review[1]). For instance, (*R*)-pantolactone is a key intermediate for the synthesis of (*R*)-pantothenic acid (Vitamin B_5). The chemo enzymatic route consists of the stereoselective formation of (*R*)-hydroxypivaldehyde cyanohydrin performed by an (*R*)-selective HNL (e.g. HNL from *Prunus amygdalus*) followed by an acid-catalyzed hydrolysis[8].

A competitive industrial enzymatic hydrocyanation process requires a robust biocatalyst, which should be easily recombinantly expressed. Avoiding expensive purification steps can reduce costs and therefore high levels of protein production and/or secreted protein expression are preferred. Furthermore, an acidic environment (pH<4.5 preferably) and a reduced water content are necessary for the suppression of the chemical background reaction and product racemization, which both affect the enantio-purity of the cyanohydrin product. Several HNLs have been engineered in order to improve pH stability as well as substrate acceptance of a specific target substrate. The (*R*)-selective HNL from *Prunus amygdalus* (*Pa*HNL) has been modified in order to achieve higher protein secretion levels in *Pichia pastoris*[9]. The same *Pa*HNL variant was further optimized for high activity and stability at low pH - as low as pH 2.5 - by site saturation mutagenesis, obtaining a significant improvement in the synthesis of (*R*)-hydroxypivaldehyde cyanohydrin up to >97% ee[8]. Directed evolution followed by site saturation mutagenesis

enhanced the pH tolerance of *Gt*HNL (*Granulicella tundricola*). The specific activity of the improved mutant is 490-fold higher than that of the wild type enzyme. Furthermore, high conversion and enantiomeric excess of different cyanohydrins of interest were obtained[10].

In addition to mutagenesis approaches, protein immobilization can improve the robustness of a biocatalyst, resulting in a more efficient and competitive process. Cross-linked enzyme aggregates (CLEA) avoid the use of expensive carriers, they are easy to recover and have enhanced shelf life and operational stability[11,12]. HNL-CLEA preparations have already been developed and significant improvement of the biocatalytic processes has been demonstrated. For instance, *Lu*HNL-CLEA was employed on a preparative scale for the synthesis of (*R*)-2-hydroxy-2-methylbutyric acid with 85% yield and 87% ee[13]. Other examples of successful HNL-CLEAs are *Pa*HNL, *Me*HNL and *Hb*HNL[14–16].

HNL from *Davallia tyermannii* (*Dt*HNL1) is a recently discovered (*R*)-selective enzyme belonging to the Bet v 1 fold superfamily. It is a small protein, with high specific activity for (*R*)-mandelonitrile synthesis (373 U/mg). Moreover, *Dt*HNL1 can convert different aldehydes into the respective cyanohydrins with high yields and high enantiomeric excess. Therefore, *Dt*HNL1 is a promising powerful biocatalyst for enantiomerically pure cyanohydrin production. The enzyme is active under acidic conditions; however, it is not stable for a long time and its activity constantly decreases at pH 2.5[i]. Here we describe the high level *Dt*HNL1 production in *Pichia pastoris*. Using the cell lysate, *Dt*HNL1-CLEA was produced in order to improve the pH stability of the enzyme. The CLEA was subsequently tested in the synthesis of (*R*)-mandelonitrile and (*R*)-hydroxypivaldehyde cyanohydrin in a biphasic system.

[i] Chapter 2, Results

Results

*Dt*HNL1 recombinant production in *Pichia pastoris* CBS 7435 Mut^S

The methyltrophic yeast *Pichia pastoris* was chosen as host organism for *Dt*HNL1 production, because of its capability of high level recombinant protein expression, as well as the very well-established molecular biology toolbox. The gene was integrated into *P. pastoris* CBS 7435 Muts^S genome. Expression of *Dt*HNL1 was under control of the strong and tightly regulated AOX1 promoter. Two different shuttle vectors were used: pPpT4_S and pPpB1 in order to obtain strains with optimal gene copy-numbers for efficient heterologous expression. Transformation with the linearized pPpB1 vector likely results in a higher copy number than pPpT4_S, due to a difference in the expression strength of the antibiotic resistance marker[17]. After transformation, zeocin resistant colonies were picked and analyzed in 96 deep well plates for *Dt*HNL1 expression by using the standard HNL assay (**Material and methods**). Selected clones were rescreened in shake flasks and *Dt*HNL1 expression was confirmed by SDS PAGE (**Fig. 1, A**). The intense band at 20 kDa indicates the strain called B1.G4 as the best *Dt*HNL1 producer.

In order to obtain gram quantities of *Dt*HNL1, the production was scaled up to the 5 L scale in a bioreactor. After four days of methanol induction, the specific activity of *Dt*HNL1 reached 300 ± 19 U/mg (degradation of (*RS*)-mandelonitrile). This activity was obtained with the cell free lysate, after Y-PER™ disruption and without any purification or concentration step. Based on this result, we can conclude that *Dt*HNL1 constitutes at least 80% of the total protein content (Standard values for the specific activity of purified *Dt*HNL1 are 320-350 U/mg with a 20% confidence interval for degradation of (*RS*)-mandelonitrile). The high *Dt*HNL1 level is further confirmed by gel electrophoresis (**Figure 1, B**). The most intense band at 20 kDa represents *Dt*HNL1, whereas the bands of other cytoplasmic proteins are weak. Table 1 reports data collected during the fermentation. *Dt*HNL1 levels reached their maximum after 75 hours and remained constant until the end of the biomass production, as confirmed by gel electrophoresis (**Figure 1, B**). 300 g WCW/L were obtained and the biomass was processed by homogenization. Cell disruption was by far not 100% efficient however, high pressure (18 kPSI) and several repeated cycles seems to negatively affect *Dt*HNL1 activity. Therefore, homogenization

was performed at lower pressure (10-13 kPSI) with three repeated cycles and approximately 10 mg of *Dt*HNL1 per 1 g of WCW, with a specific activity of 290 U/mg of total protein was obtained.

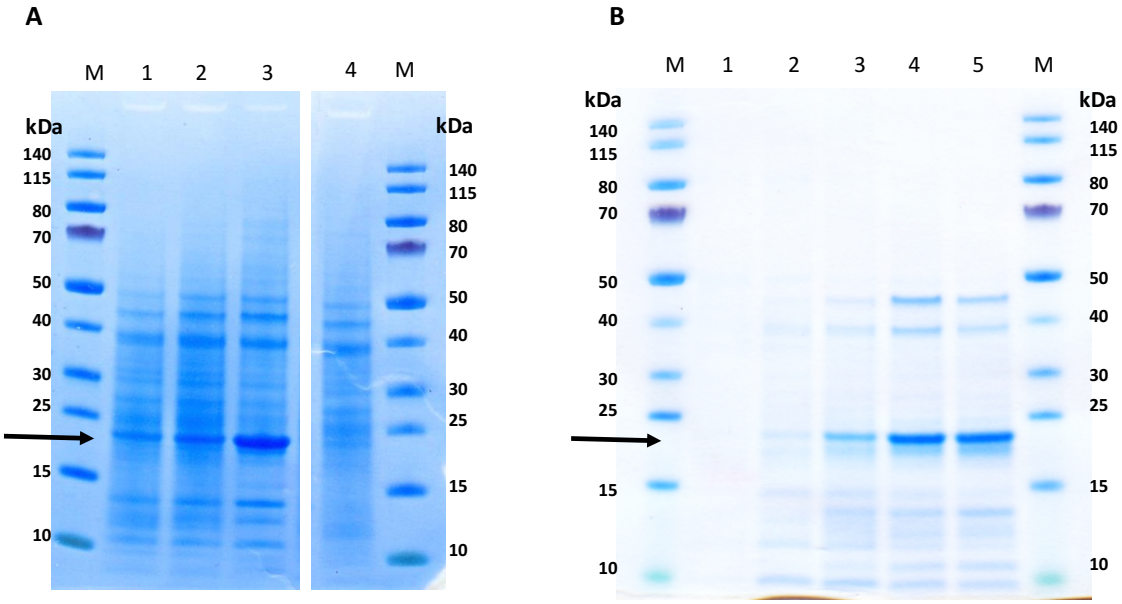


Figure 1. SDS PAGE monitoring recombinant expression of *Dt*HNL1 in *P. pastoris* CBS 7435 Mut^S. A: *Dt*HNL1 expression in shake flask cultivation after 72 hours of methanol induction. PageRuler™ Prestained Protein Ladder (M); 20 µg cell free protein lysate of T4.F1 (1); T4.B7 (2); B1.G4 (3); negative control *P. pastoris* CBS 7435 Mut^S (4). T4 and B1 indicate the original vector constructs pPpT4_s_DtHNL1 or pPpB1_DtHNL1 respectively. **B: Cultivation and induction of *P. pastoris* CBS 7435 Mut^S *Dt*HNL1 B1.G4 bioreactor experiment**. PageRuler™ Prestained Protein Ladder (M); cell free protein lysate of different sampling time: 0 hours (1); 25 hours (2); 49.5 hours (3); 75 hours (4); 96.5 hours (5). The black arrows indicate *Dt*HNL1 (20kDa).

Table 1. Bioreactor cultivation summary of *P. pastoris* CBS 7435 Mut^S *Dt*HNL1 B1.G4. OD: optical density. WCW: wet cell weight. DCW: dry cell weight. Protein mg/mL refer to the protein concentration of the cell free lysate after Y-PER™ disruption. n.d.: not determined (measurement was not performed). Activity in units (U) per mg total protein: 1 U corresponds to 1 µmol of (*RS*)-mandelonitrile, cleaved per minute by the enzyme to benzaldehyde and HCN.

Day	Carbon Source	Cultivation time h	WCW g/L	DCW g/L	Protein mg/mL	Activity U/mg
1	Glycerol	0	178	26 ± 2	0.4	n.d.
2	Methanol	25	412	108 ± 1	0.8	n.d.
3	Methanol	49.5	400	82 ± 2	1.3	n.d.
4	Methanol	75	372	107 ± 2	2.5	n.d.
5	Methanol	96.5	428	89 ± 4	2.3	**300 ± 19**

*Dt*HNL1-CLEA development

Protein precipitation is the first step for CLEA preparation. The screening showed that all *Dt*HNL1 precipitated at 70% saturation with ammonium sulfate (**Fig. 3a**). Starting with 950 U/mL before precipitation, the remaining supernatant after precipitation contained almost no activity (8 U/mL), which was similar than the blank where no enzyme was expressed. Cross-linking is a second important step for an active CLEA preparation, since it "locks" the enzyme in its active state and prevents leaching during the reaction[13]. The activity recovery after cross-linking with different concentrations of glutaraldehyde was investigated (**Fig. 3b**). An optimum range is defined between 52 and 102 mM of glutaraldehyde. Therefore, 52 mM glutaraldehyde was chosen for the production of the *Dt*HNL1-CLEA. A leaching study confirmed that 52 mM was a sufficient concentration for cross-linking without re-dissolution of the enzyme (data not shown).

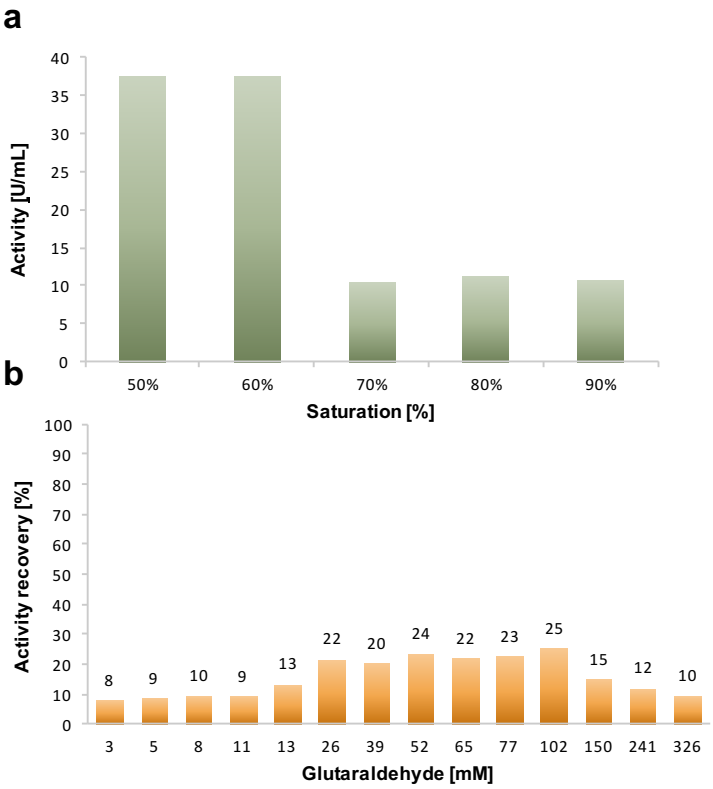


Figure 3. *Dt*HNL-CLEA development. Precipitation screening (A): *Dt*HNL1 activity in the supernatant after ammonium sulfate precipitation. **Cross-Link screening (B):** Recovered activity after cross-linking with glutaraldehyde at different concentrations. HNL activity is based on mandelonitrile cleavage; benzaldehyde formation is followed at 280 nm.

Finally, capping of the CLEA can influence different properties, such as particle size, resuspendability, filterability as well as mechanical stability. Tetra methyl orthosilicate (TMOS) was added to improve the resuspendability of the CLEA and to improve the powder characteristics of the dried CLEA. Capping did not affect the activity recovery and 5% of TMOS resulted in good free flowing material (data not show).

pH stability analysis

The production of enantiomerically pure cyanohydrins is preferably performed under acidic conditions, in order to suppress the chemical background reaction. Therefore, a stable biocatalyst at low pH is a key aspect for successful preparation of enantiomerically pure cyanohydrins. The stability of *Dt*HNL1-CLEA (CLEA prep2) was tested by incubation of the immobilized enzyme at different pHs: 2.4; 3.0; 3.5; 4:0 respectively (**Fig. 3**).

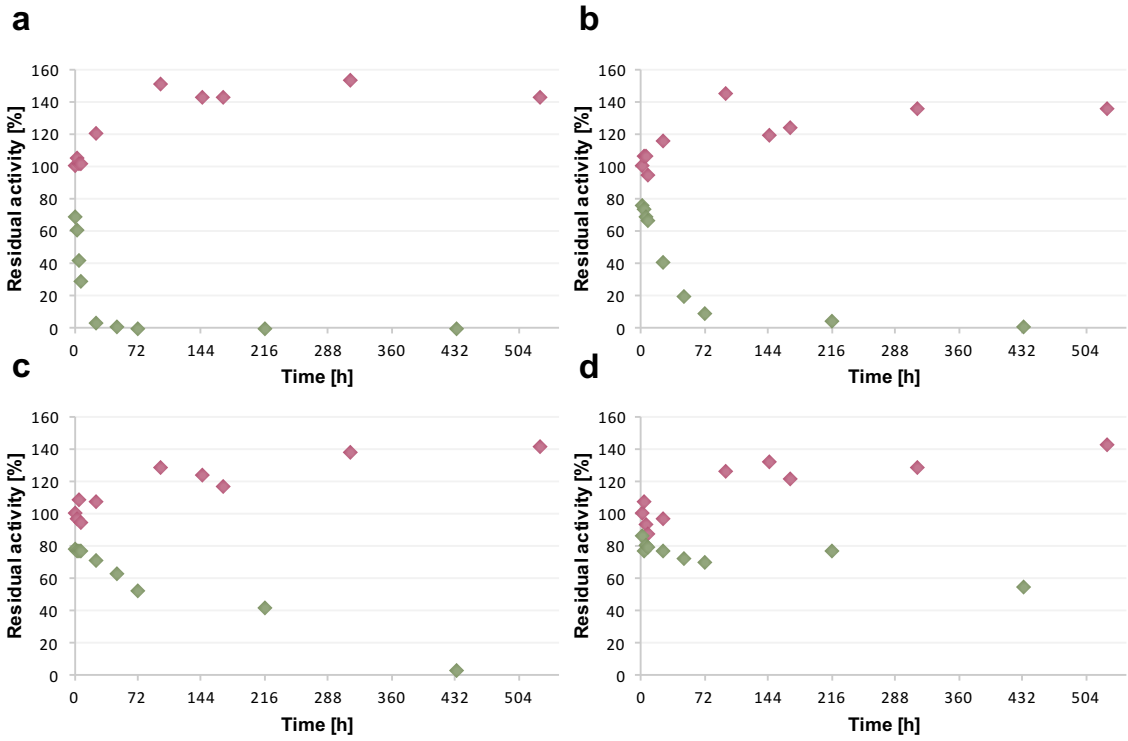


Figure 3. pH Stability of *Dt*HNL1-CLEA (◆) and *Dt*HNL1 free enzyme (◆). pH 2.4 (a); pH 3.0 (b); pH 3.5 (c); pH 4.0 (d). CLEA Prep2 (**Material and Methods**) was used.

*Dt*HNL1-CLEA does not seem to be affected by the pH, and it behaves similar in all four conditions. Figure 3 also shows a comparison between residual activities measured for immobilized *Dt*HNL1 and free enzyme (cell free lysate from *P. pastoris* CBS 7435 Mut^S

*Dt*HNL1 B1.G4). The *Dt*HNL1-CLEA system is substantially more stable than the soluble enzyme. The difference is even more remarkable at pH 2.4 (**Fig 3a**) and 3.0. (**Fig 3b**). *Dt*HNL1 looses activity in c.a. 72 hours, whereas the activity of *Dt*HNL1-CLEA is stable after 504 hours.

Cyanohydrin synthesis

*Dt*HNL1 can find its final application as biocatalyst for the synthesis of enantiomerically pure cyanohydrins. Therefore, three different *Dt*HNL1-CLEAs (**Material and Methods**) were compared with the cell free lysate from *P. pastoris*. We chose the synthesis of (*R*)-mandelonitrile as the benchmark reaction. After 2 hours, substrate conversion and enantiomeric purity of the product were comparable between CLEA preparation 3 (CLEA prep3) and the free enzyme (46.7% and 43.1% conversion of benzaldehyde with CLEA prep3 and free enzyme respectively). Nevertheless, a difference was found after 24 hours, where *Dt*HNL1-CLEA prep3 showed much higher conversion (**Table 2**). Best conversion and enantiomeric excess were achieved with *Dt*HNL1-CLEA prep2 after 24 hours (99.0% conversion and 98.0% enantiomeric excess). This indicates that the CLEA preparation is a more robust biocatalyst than the free enzyme. Organic solvent and the strong mechanical agitation might affect the stability of the free enzyme, which was partially or completely inactivated during the 24 hours. Dry CLEA in organic solvent did not show any product formation, indicating a certain water content to be essential for the catalysis.

Finally, we tested the two successful *Dt*HNL1-CLEAs for (*R*)-hydroxypivaldehyde cyanohydrin synthesis (**Table 3**). Hydroxypivaldehyde exists in equilibrium between its monomer and dimer form. A pH, which is lower than 2.5 is preferred in order to slow down the dimerization process[8] and HNLs which are stable in this highly acidic environment are required. We carried out the reaction in a biphasic system, at pH 2.4. Again, *Dt*HNL1-CLEA showed an improvement, due to its increased stability at low pH. The enzyme catalyzed formation of the product was confirmed by the enantiomeric excess of 18% after 24 hours. The high substrate conversion was due to the strong background reactions: chemical cyanohydrin formation and dimerization of the substrate.

Table 2. Synthesis of (*R*)-mandelonitrile with different *Dt*HNL1 preparations. The synthesis of cyanohydrins was performed with 0.2 mg of biocatalyst dissolved in buffer or TMBE. 0.5 M aldehyde was mixed with 2 M HCN in TBME, 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, 10° C and 1000 rpm. Samples were analyzed by GC after acetylation of the product. Conversion is based on the substrate consumption. Dashes indicate that no measurement was recorded; ND, not determined (no product formation).

Benzaldehyde + HCN → (*R*)-Mandelonitrile

Reaction	Enzyme Preparation	2 hours		24 hours	
		Conversion %	ee %	Conversion %	ee %
Organic solvent*	**CLEA prep1**	15	n.d.	-45.2	n.d.
	Control	-16.5	n.d.	---	---
Biphasic System	**CLEA prep2**	56.2	98.0	99.0	98.0
	CLEA prep3	46.7	96.4	95.2	95.7
	Free enzyme lysate	43.1	97.7	76.9	95.1
	Control	-1.5	0.5	2.4	-2.8

*TBME saturated with acetate buffer pH 4.0.

Table 3. Synthesis of (*R*)-hydroxypivaldehyde with different *Dt*HNL1 preparations. 0.5 M aldehyde was mixed with 2 M HCN in TBME, 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 2.4, 10° C and 1000 rpm. Samples were analyzed by GC after acetylation of the product. Conversion is based on the substrate consumption.

Hydroxypivaldehyde + HCN → (*R*)-Hydroxypivaldehyde cyanohydrin

Reaction	Enzyme Preparation	2 hours		24 hours	
		Conversion %	ee %	Conversion %	ee %
Biphasic system**	**CLEA prep 2** (0.45 mL)	52.0	13.0	77.6	14.4
	CLEA prep2 (0.045 mL)	44.9	-0.9	55.4	1.1
	CLEA prep 3 (0.45 mL)	57.2	20.3	79.9	18.4
	CLEA prep 3 (0.045 mL)	46.8	-0.8	56.8	-0.1
	Free enzyme purified	74.3	2.4	74.3	0.5
	Control	35.3	-1.6	43.5	-0.3

Discussion

*Dt*HNL1 became easy available by recombinant production in *Pichia pastoris* CBS 7435 Mut^S. The overexpressed enzyme was estimated to be ca. 80% of the total protein content. The preparation of CLEAs (*Dt*HNL1-CLEA) resulted in a much more robust biocatalyst under acidic conditions, which is an interesting feature for efficient production of enantiomerically pure cyanohydrin. *Dt*HNL1-CLEA was applied for the synthesis of (*R*)-hydroxypivaldehyde cyanohydrin at pH 2.4 which is one of the most difficult HNL reaction described. An improvement of the enantiomeric excess of the product due to the use of CLEAs was observed. Further engineering of *Dt*HNL1 for a better acceptance of aliphatic substrates, followed by CLEA preparation, could lead to a biocatalyst, stable at low pH and able to synthesize a broad range of different cyanohydrins. Finally, the application of *Dt*HNL1-CLEA for the synthesis of (*R*)-mandelonitrile showed 99% conversion and 98% enantiomeric excess, a significant improvement compared to the free enzyme. In conclusion, a robust biocatalyst was developed, and high conversion and enantiomeric excess were obtained for a standard substrate such as benzaldehyde.

Material and methods

CAUTION: All procedures involving hydrogen cyanide must be performed in a well-ventilated fume-hood equipped with an HCN detector. HCN-containing wastes were neutralized.

General

All chemicals were purchased from Sigma-Aldrich or Carl Roth GmbH, if not stated otherwise. Racemic mandelonitrile was purchased from abcr GmbH & Co. KG. or Sigma Aldrich. Material for molecular biology and protein analysis was obtained from Thermo Fisher Scientific or Promega, if not specifically mentioned. Spectrophotometric measurements were performed with a Synergy Mx plate reader (BioTek) or Shimadzu UV-1800 Spectrophotmeter.

Strains and vectors

E. coli TOP F' (Thermo Fisher Scientific) was used for cloning and propagation of the vectors, *Pichia pastoris* CBS 7435 Mut^S was used for protein expression. pPpT4_S and pPpB1 were used as shuttle vectors for integration of the gene into *Pichia pastoris* CBS 7435 Mut^S [17].

Protein analysis

Protein quantification was performed with Pierce™ BCA Protein Assay Kit. For protein electrophoresis, NuPAGE® Novex® 4-12% Bis-Tris Protein Gels and an XCell SureLock® Mini-Cell equipped with a PowerEase® 500 Programmable Power Supply were used. Gels were stained with SimplyBLue™ Safe Stain (Invitrogen).

*Dt*HNL1 cloning and transformation of *Pichia pastoris*

*Dt*HNL1 gene was codon optimized for expression in *Pichia pastoris* under the control of P_{AOX1} promoter[18] and the self-designed synthetic gene was purchased (GeneArt® Gene synthesis, Thermo Scientific). The gene was cloned into pPpT4_S and pPpB1 shuttle vectors (restriction sites: EcoRI/NotI)[17]. *Pichia pastoris* CBS 7435 Mut^S was transformed with 400 ng (pPpT4_S) or 3 µg (pPpB1) of linearized vectors (SwaI for pPpT4_s or BglII for pPpB1, respectively) as described in detail by Lin-Cereghino *et al.*[17,19]. Selection was done

on zeocin (100 µg/mL and 300 µg/mL). The gene sequence was verified after each cloning and transformation step by Sanger sequencing (LGC Genomics).

Protein expression in *Pichia pastoris*

Cells were cultivated in buffered minimal media in deep well plates or shake flasks at 28°C and 80% humidity. Protein expression was induced by methanol addition as described in detail previously[20]. Cells disruption was performed by Y-PER™ Plus Dialyzable Yeast Protein Extraction Reagent (Thermo Fisher Scientific) according to the provided manual or with glass beads in 50 mM sodium phosphate buffer at pH 6.5. Cell free lysate was stored at -20°C.

*Dt*HNL1 production on 5L scale

Cultivation was performed in 5 L BIOSTAT® CT and BIOSTAT® CTplus fermenter based on the manual provided by Invitrogen™ (Thermo Fisher Scientific) at 28°C, pH 6.0. Fed-batch phase with methanol feeding was performed for four days. Finally, cells were harvested by centrifugation (1 hour 4000 rpm 4°C) and disrupted with a Micro DeBEE High Pressure Homogenizer in 50 mM sodium phosphate buffer at pH 6.5. Cell pellet or cell free lysate was stored at -20°C.

Cells dry weight (CDW)

For cell dry weight determination, 200 µL of culture was harvested in a pre-weighed 1.5 mL tube. The supernatant was discarded and the cell pellet was dried in a 60°C incubator for two days. After drying, the tubes were weighed again and the cell dry weight was calculated.

Hydroxynitrile lyase activity

Screening of P. pastoris transformants

For the determination of the enzymatic activity SynergyMx Plate Reader (Biotek Inc., Winooski, United States) was used. 20 µL of the appropriate dilutions of cell free lysate were mixed with 130 µL 50 mM citrate phosphate buffer pH 5.0. The reaction was initiated by the addition of 50 µL of 60 mM racemic mandelonitrile, freshly prepared in 3 mM citrate phosphate buffer pH 3.5. The activity was determined by detection at 280 nm

of benzaldehyde produced. The molar extinction coefficient of benzaldehyde was used for the calculations: 1.376 L $mmol^{-1}$ cm^{-1}. One unit of HNL activity is defined as the amount of enzyme, which converts 1 µmol of mandelonitrile per minute.

CLEA studies

For the determination of the *Dt*HNL1 activity of the free enzyme and the CLEA, a UV-1800 spectrophotometer was used at a wavelength of 280 nm to measure the absorbance of benzaldehyde at room temperature. A sample of 200 µL of appropriate dilutions of enzyme were mixed with 1300 µL citrate buffer 50 mM pH 5 and 500 µL of 60 mM racemic mandelonitrile, prepared in 3 mM citrate buffer of pH 3.5. The reaction with the CLEA was performed in an Eppendorf tube, where the mixture was centrifuged and the supernatant was measured. The molar extinction coefficient of benzaldehyde was used for the calculations: 1.376 L mmol-1 cm-1. One unit of HnL activity is defined as the amount of enzyme, which converts 1 µmol of mandelonitrile per minute

*Dt*HNL1-CLEA preparation

Precipitation studies

A saturated ammonium sulfate solution of pH 7.5 was prepared by saturating dH2O with ammonium sulfate and subsequently adjusting the pH with 1 M NaOH. A series of precipitations were performed by incubating the enzyme with 50%-90% saturated ammonium sulfate solution. The precipitate was shaken softly for 15 minutes and then centrifuged. The supernatant was measured on activity with mandelonitrile hydrolysis.

Cross linking studies

The HNL1 enzyme was precipitated with 80% saturated ammonium sulfate solution of pH 7.5 and several amounts of 25% glutaraldehyde solution were added, ranging from 3 to 326 mM. The CLEA was cross-linked at room temperature for 3 hours by shaking at 500 rpm. The products were washed 3 times with 50 mM citrate phosphate buffer pH 5 and stored in 50 mM citrate buffer. The activities were measured monitoring hydrolysis of racemic mandelonitrile.

Leaching studies

A CLEA suspension was washed and resuspended in fresh 50 mM citrate phosphate buffer of pH 5. The suspension was stirred overnight at room temperature at 990 rpm. The activity of the CLEA and the supernatant was measured after 18 hours.

Capping studies

To an 80% ammonium sulphate enzyme precipitate 52 mM of a 25% glutaraldehyde solution was added. Various amounts of TMOS were added, ranging between 0 to 5 v% and the CLEA was cross-linked for 18 hours at room temperature by shaking at 500 rpm.

The CLEA was subsequently washed three times with 50 mM citrate phosphate buffer pH 5 or dried.

Drying

The CLEAs were washed one time with 50 mM citrate phosphate buffer pH 5, one time with acetone and two times with diethylether, and were left to dry by airflow for 1 hour.

For determination of the activity, a sample was resuspended in 50 mM citrate phosphate buffer pH 5. The activity was measured by mandelonitrile hydrolysis. Only the best material was measured on activity.

Preparation 1: Dry CLEA 5 v% TMOS capping

Preparation 2: CLEA suspension 5 v% TMOS capping

Preparation 3: CLEA suspension no capping

pH stability

A pH stability test was performed by incubating the CLEA and the free enzyme in 50 mM citrate phosphate buffer at various pH, ranging from 2.4-4. The activities were measured over time by mandelonitrile hydrolysis.

Cyanohydrin synthesis

Synthesis of (*R*)-mandelontrile and (*R*)-hydroypivaldehyde cyanohydrin was carried out in a biphasic system as described in detail by Wiedner *et al*.[10]. For comparative studies the

aqueous phase contained 0.2 mg of *Dt*HNL1 in the different forms. Synthesis of (*R*)-hydroxypivaldehyde was performed with 0.45 mL of CLEA preparation or 0.7 mg of purified *Dt*HNL1. Purified *Dt*HNL1 was prepared as described by Lanfranchi *et. al*. [i].

[i] Chapter 3, Material and methods.

List of Abbreviations

CLEA = Cross Link Enzyme Aggregate

DCW = Dry Cell Weight

ee = enantiomeric excess

HCN = Hydrocyanic acid

HNL = Hydroxynitrile lyase

Prep = Preparation

SDS PAGE = SDS Polyacrylamide gel electrophoresis

WCW = Wet Cell Weight

Acknowledgments

We are grateful to Karin Reicher, Elena Pietrich and Dr. Mandana Gruber for technical support. 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) and the Federal Ministry of Science, Research and Economy (BMWFW), the Federal Ministry of Traffic, Innovation and Technology (bmvit), the Styrian Business Promotion Agency SFG, the Standort agentur 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.

References

1. Dadashipour, M. & Asano, Y. Hydroxynitrile lyases: insights into biochemistry, discovery, and engineering. *ACS Catal.* **1,** 1121–1149 (2011).
2. Winkler, M., Glieder, A. & Steiner, K. in *Comprehensive Chirality* (eds. Carreira, E. M. & Yamamoto, H.) **7,** 350–371 (2012).
3. Hussain, Z. *et al.* Characterization of two bacterial hydroxynitrile lyases with high similarity to cupin superfamily proteins. *Appl. Environ. Microbiol.* **78,** 2053–2055 (2012).
4. 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).
5. 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).
6. Lanfranchi, E. *et al.* Mini-review: recent developments in hydroxynitrile lyases for industrial biotechnology. *Recent Pat. Biotechnol.* **7,** 197–206 (2013).
7. 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).
8. Pscheidt, B. *et al.* Efficient biocatalytic synthesis of (*R*)-pantolactone. *Adv. Synth. Catal.* **350,** 1943–1948 (2008).
9. Glieder, A. *et al.* Comprehensive step-by-step engineering of an (*R*)-hydroxynitrile lyase for large-scale asymmetric synthesis. *Angew. Chem. Int. Ed. Engl.* **42,** 4815–8 (2003).
10. Wiedner, R. *et al.* Improving the properties of bacterial *R*-selective hydroxynitrile lyases for industrial applications. *ChemCatChem* **7,** 325–332 (2015).
11. Sheldon, R. A. Cross-linked enzyme aggregates as industrial biocatalysts. *Org. Process Res. Dev.* **15,** 213–223 (2011).
12. Sheldon, R. A. *Comprehensive Chirality*. *Comprehensive Chirality* **9,** (Elsevier, 2012).
13. Cabirol, F. L. *et al. Linum usitatissimum* hydroxynitrile lyase cross-linked enzyme aggregates: a recyclable enantioselective catalyst. *Adv. Synth. Catal.* **350,** 2329–2338

(2008).

14. Chmura, A. *et al.* Cross-linked aggregates of the hydroxynitrile lyase from *Manihot esculenta*: highly active and robust biocatalysts. *Adv. Synth. Catal.* **348,** 1655–1661 (2006).
15. Cabirol, F. L., Hanefeld, U. & Sheldon, R. A. Immobilized hydroxynitrile lyases for enantioselective synthesis of cyanohydrins: sol-gels and cross-linked enzyme aggregates. *Adv. Synth. Catal.* **348,** 1645–1654 (2006).
16. Van Langen, L. M., Selassa, R. P., Van Rantwijk, F. & Sheldon, R. A. Cross-linked aggregates of (*R*)-oxynitrilase: a stable, recyclable biocatalyst for enantioselective hydrocyanation. *Org. Lett.* **7,** 327–329 (2005).
17. Näätsaari, L. *et al.* Deletion of the *Pichia pastoris* ku70 homologue facilitates platform strain generation for gene expression and synthetic biology. *PLoS One* **7,** (2012).
18. Abad, S. *et al.* Stepwise engineering of a *Pichia pastoris* D-amino acid oxidase whole cell catalyst. *Microb. Cell Fact.* **9,** 24 (2010).
19. Lin-cereghino, J. *et al.* Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast *Pichia pastoris*. *Biotechniques* **38,** 48–52 (2005).
20. Weis, R. *et al.* Reliable high-throughput screening with *Pichia pastoris* by limiting yeast cell death phenomena. *FEMS Yeast Res.* **5,** 179–189 (2004).

Chapter 5
Determination of Hydroxynitrile Lyase Activity after Blue Native PAGE

Elisa Lanfranchi[1]

[1]Austrian Centre Industrial Biotechnology, Petersgasse 14 8010 Graz, Austria

Creation date: 03.02.2014 Restrictions: no Validity: until revoked/cancelled This protocol replaces the version from: -- Developed in project/working group: Kyrobio / Winkler M.	
Notice of modification	--
Written by	Elisa Lanfranchi
Checked: Yes	Eva-Maria Köhler
Approved by	Margit Winkler

Purpose and field of application

This method is used to determine hydroxynitrile lyase activity after a non-denaturing protein gel electrophoresis, in order to distinguish native HNL protein from inactive background proteins. Moreover, the method can reveal the molecular weight of the active enzyme. MS analysis can be done on the detected protein band (e.g. for protein sequence discovery).

Principle

The method comprises of two steps: in the first step, the proteins are separated by electrophoresis. Then, the polyacrylamide gel is directly used for the hydroxynitrile lyase activity assay.

Blue Native PAGE allows the separation of the proteins according to their molecular weight, maintaining their native condition. All proteins acquire negative charge due to Comassie G-250, therefore, is not necessary to know the pI of the protein of interest in order to prepare a suitable buffer (in contrast to the standard native PAGE). Therefore, it can be applied to any protein (unknown proteins too), and it is suitable for separation of crude protein extracts.

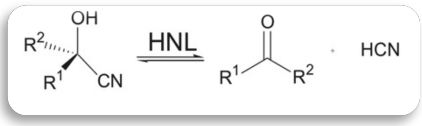


Figure 1. HNL reaction scheme.

Once the electrophoresis run is completed, the hydroxynitrile lyase activity (reaction scheme figure 1) is performed within the polyacrylamide gel. The method is based on the hydroxynitrile lyase colony assay, published by Krammer[1]. The polyacrylamide gel is directly in contact with papers soaked into the substrate (any cyanohydrin is possible – therefore, the assay is not restricted to a particular HNL class. The Feigl–Anger test paper is placed on the top of the "sandwich assembly". When the gaseous cyanide reaches the test paper, blue spots appear. Proteins exhibiting hydroxynitrile lyase activity are identified by comparison of the activity assay and the stained polyacrylamide gel. The corresponding bands can be excised and used for MS analysis.

Key words, definitions & abbreviations

Hydroxynitrile lyase; Enzyme discovery; Activity Assay; Blue Native PAGE; Cyanohydrins

HNL: Hydroxynitrile lyase

*Pa*HNL: *Prunus amygdalus* Hydroxynitrile lyase

*Lu*HNL: *Linum usitatissimum* Hydroxynitrile lyase

BN PAGE: Blue Native polyacrylamide gel electrophoresis

Methodology

Materials

Name	Supplier	Order No.	Comments
NativePAGE™ Novex®4-16% Bis-Tris Gels	Thermo Fisher Scientific	BN1002BOX	Store at 4°C
NativeMark™ Unstained Protein Standard	Thermo Fischer Scientific	LC0725	Store at -20°C
Mandelonitrile lyase from Almonds	Sigma Aldrich	M6782	Store at 4°C
Whatman® filter paper	GE Healthcare		
Mosquito net			Baumax

Apparatus

Name	Supplier
XCell *SureLock*® Mini-Cell	Thermo Fischer Scientific
The PowerEase® 500 Programmable Power Supply	Thermo Fischer Scientific
Digital Camera	

Reagents

NativePAGE™ Novex®

Name	Formula	MW	Purity	Supplier	CAS No.
BisTris	$C_8H_{19}NO_3$	209.24	≥99%	Carl Roth	6976-37-0
Tricine	$C_6H_{13}NO_3$	179.2	≥99%	Carl Roth	5704-04-1
Hydrochloric acid	HCl	36.4	37%	Carl Roth	7647-01-0
Sodium Chloride	NaCl	58.44	≥99.5%	Carl Roth	7647-14-5
Glycerol	$C_3H_8O_3$	92.09	≥99.5%	Carl Roth	56-81-5
Ponceau S					
Coomassie G-250	$C_{47}H_{50}N_3NaO_7S_2$ (Sodium salt)	856.03		Fluka	6104-58-1

HNL activity

Name	Formula	MW	Purity	Supplier	CAS No.
Copper(II) ethylacetoacetate	$C_{12}H_{18}CuO_6$	321.81	99%	Abcr	14284-06-1
4,4 Methylenebis (N,N-dimethylaniline)	$C_{17}H_{22}N_2$	254.37	98%	Sigma Aldrich	101-61-1
Chloroform	$CHCl_3$	119.38	≥99.9%	Carl Roth	67-66-3
Citric Acid – Monohydrate	$C_6H_8O_7$	210.14	≥99.5%	Carl Roth	77-92-9
Trisodium citrate dihydrate	$C_6H_5O_7$ $2H_2O$ 3Na	294.1	≥99%	Carl Roth	6132-04-3
(RS)-Mandelonitrile[i,ii]	C_8H_7NO	133.15	97%	Abcr	532-28-5

Coomassie Staining

Name	Formula	MW	Purity	Supplier	CAS No.
Acetic Acid	$C_2H_4O_2$	60.05	≥99.5%	Carl Roth	64-19-7
Methanol	CH_4O	32.04	≥99.9%	Carl Roth	67-56-1
Coomassie R-250	$C_{45}H_{44}N_3NaO_7S_2$ (Sodium salt)	825.97		Sigma Aldrich	6104-59-2

[i] Toxic

[ii] Or any other cyanohydrin

Solutions – Native-PAGE[2]

Running Buffer (10X) 1L pH 6.8

BisTris 104.6 g

Tricine 89.6 g

Check pH value and correct with HCl. Store at room temperature.

Sample Buffer (4X) 10mL pH 7.2

BisTris	0.418 g
6N HCl	0.107 mL
Glycerol	4 g
NaCl	0.117 g
Ponceau S	0.4 mg

Store at 4°C.

Cathode Additive (20X) 250mL

Coomassie G-250	1g
Deionized Water	250mL

Anode Buffer (1X) 1L (~ 600mL needed for 1 electrophoresis)

Running buffer 10X	100 mL
Deionized Water	900 mL

Light Blue Cathode Buffer 1X 200mL (~ 200mL needed for 1 electrophoresis)

Running buffer 10X	20 mL
Cathode Additive 20X	1 mL
Deionized water	179 mL

Solutions - HNL activity

Reaction buffer: Citrate Buffer 100mM pH 4.5

Citric acid	10.44 g/L
Sodium Citrate	29.41g/L

Autoclave for long storage

In case of background or suspicious unspecific reactions, use pH 4.0 and/or increase molarity of the buffer.

Mandelonitrile stock 15-30 mM

Dissolve the cyanohydrin in 100mM citrate buffer pH 4.5. Prepare freshly

Detection Paper[1]

- Prepare a 1% (w/v) solution of copper (II) ethylacetoacetate in chloroform (solution A).
- Prepare a 1% (w/v) solution of 4,4'-methylenebis (N,N-dimethyl-aniline) in chloroform (solution B).
- Slowly add solution B to solution A while stirring the mixture (ratio 1:1).
- The resulting mixture is dark green and clear.
- Soak Whatman No.1 filter sheets with the prepared solution (~ 2mL) and let them dry under a fume hood. The pale green, almost colourless detection filters can be stored on a dark, cold and dry place for several months.

Procedure

1. *Sample preparation*[2]

- **Prepare cell/tissue lysate** with the most convenient method (e.g.: yeast Y-PER® or glass beads; *E. coli* sonication; plant tissue P-PER® or mechanical disruption).
- Maintain the salt concentration of the sample less than 50 mM (PD10 columns can be employed for desalting or buffer exchange).
- Sample preparation in 1X Sample Buffer is possible.
- Prepare sample for Native PAGE®

Sample	X µL
Sample buffer (4X)	2.5µL
Deionized water	Up to 10µL

- Sample prepared in Sample Buffer 1X can be loaded directly.
- Maximum loading volume: 40 µL.
- Protein amount depends on HNL activity and staining method. Make different dilutions, in order to find the good compromise (e.g. 7 µg of *Pa*HNL from Sigma show a strong signal after few minutes)
- **Do not heat samples!!!**
- For a better electrophoresis run, load the same volume in all wells (load 1X sample buffer in the empty wells).
- Load 6 µL (Coomassie staining) or 3 µL (Silver staining) of the protein ladder.

2. *Electrophoresis run*[2]

- Perform electrophoresis at 4°C (cold room) with pre-chilled buffers.
- Before loading the samples, rinse the wells with cathode buffer 2/3 times. Then fill the wells with cathode buffer. Be sure to displace all air bubbles from the wells, as they will affect sample running.
- Load samples **before** filling the upper buffer chamber!
- Pour:

  ~ 600 mL of anode buffer 1X in the lower (outer) buffer chamber (**Fig. 2**).

  ~ 200 mL of cathode buffer 1X in the upper (inner) buffer chamber (**Fig.2**).

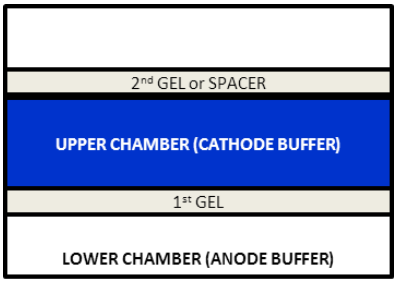


Figure 2. Schematic representation of the assembled electrophoresis chamber.

- Running conditions:

1st step: 150 V constant for 60 minutes (Starting current: 8-15 mA).

2nd step: 250 V constant for 60-90 minutes. The time is variable: keep running until the front reaches the bottom of the gel (Ending current 2-4 mA).

3. Equilibration

- After the electrophoresis, wash the gel 1-2 times with deionized water and equilibrate it 30 minutes in the reaction buffer (citrate buffer 100 mM, pH 4.5) at 4°C.
- NB: If the activity assay is not immediately performed, store the gel at 4°C in an appropriate buffer for the enzyme (e.g. potassium phosphate buffer 50 mM, pH 6.0). Do not store the gel for long time, in order to avoid the decrease or loss of the enzyme activity (ideally, perform the activity assay on the same day).

4. HNL activity

- Assemble the different components in the following order (Fig. 3):

  1.Whatman paper soaked in reaction buffer
  2.NATIVE polyacrylamide gel
  3.Whatman paper soaked in the substrate solution
  4.Mosquito Net
  5.Detection Paper
  6.ransparent plastic (e.g. petri dish's lid) + Weight (e.g. keep press with your fingers, or a small glass bottle)

Notes

- Remove the liquid in excess in each layer (help yourself with a plastic 5mL pipet as rolling pin
- The activity assay is based on HCN release! Therefore, perform the reaction in a well-ventilated fume hood.

- Wait until blue spots appear on the detection paper. Take pictures with a camera or scan the paper. The blue colour disappears over time. Therefore, it is recommended taking the picture within one or few days.

Hints

- When mandelonitrile is used as substrate, *Pa*HNL can be used as positive control. This can be a nice "activity marker", and helpful for interpretation of the results. *Lu*HNL (or any other protein) is an optimal negative control in case of mandelonitrile, because it does not accept mandelonitrile as substrate.
- Other known and purified HNLs can be used as activity marker: *Hb*HNL, *Gt*HNL or *At*HNL are few examples.

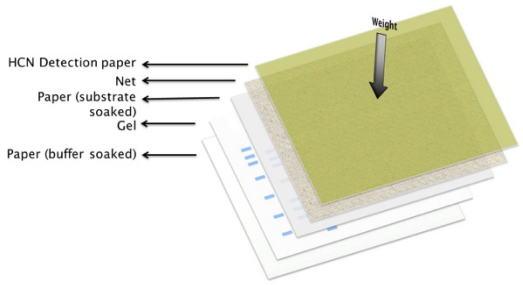


Figure 3. Schematic representation of the assembly for the activity coupled blue native gel assay.

5. Gel Staining

- The polyacrylamide gel can subsequently be either stained with coomassie or silver staining.
- When coomassie staining is chosen, the fixation step is required. The protocol is the following:
- Fix solution (40% methanol, 10% acetic acid) –Place the gel in the fix solution (~150 mL) and microwave on high (950-1100 watts) for 30-45 seconds (the liquid should boil) – shake 15/30 minutes at room temperature (repeat 2 times).
- Staining solution (0.02% coomassie R-250, 30% methanol, 10% acetic acid) - Place the gel in the staining solution (~150 mL) and microwave on high (950-1100 watts) for 30-45 seconds (the liquid should boil) – shake 15/30 minutes at room temperature.
- Destaining solution (8% acetic acid) – Place the gel in the destaining solution (~150 mL) and microwave on high (950-1100 watts) for 30-45 seconds (the liquid should boil) – shake at room temperature until the contrast is acceptable.

- NB: SimplyBlue™ SafeStain (Thermo Fischer Scientific) is **not** compatible with this kind of gels.

Result examples

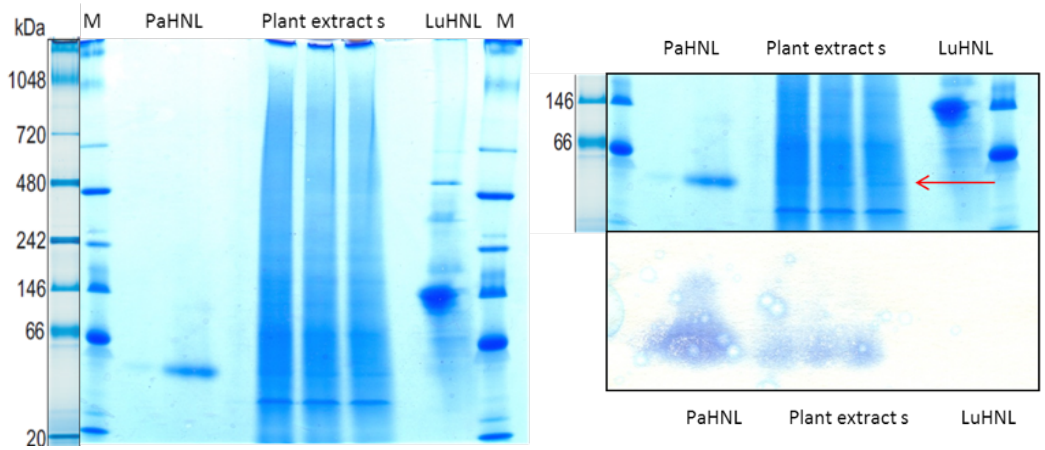


Figure 4. BN-PAGE and HNL in gel-activity assay. M: NativeMark™ Unstained Protein Standard. *Pa*HNL: *Prunus amygdalus* HNL (Sigma); Plant extrats: crude protein extract from cyanogenic plant tissue (Sample in triplicate); LuHNL: *Linum usitatissimum* HNL expressed in *P. pastoris*. Exposition time 15 minutes.

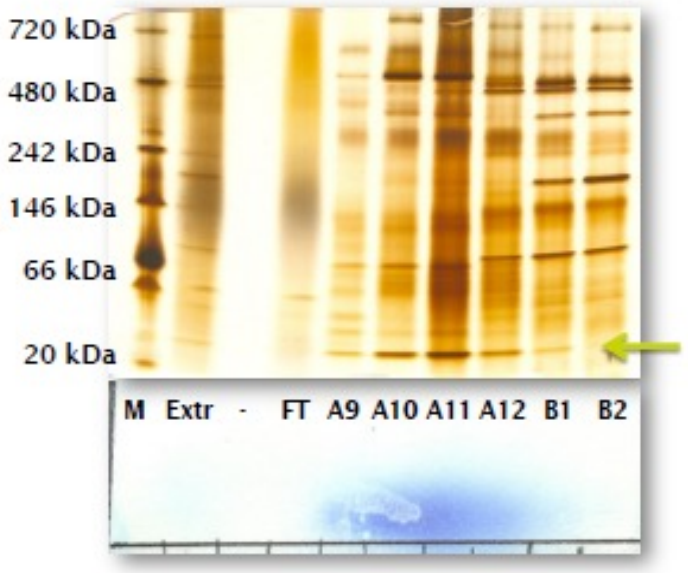


Figure 5. Silver stained BN-PAGE and HNL activity detection. The protein was partially purified by anion exchange. Extr: crude protein extract; FT: flow through; A9-B2: purification fractions.

Safety precautions

CAUTION: All procedures involving hydrogen cyanide must be performed in a well-ventilated fume-hood, ideally equipped with an HCN detector.

References

1. Krammer, B., Rumbold, D., Tschemmernegg, M., Poechlauer, P. & Schwab, H. A novel screening assay for hydroxynitrile lyases suitable for high-throughput screening. J. *Biotechnol.* **129**, 151-161 (2007).
2. NativePAGE™ Novex® Bis-Tris gel system Manual, Thermo Fischer Scientific (2012)

Conclusions and outlook

Biocatalysis strongly demands new enzymes either with novel activities or exhibiting new substrate scope or better biocatalytic features, because of the continuous search for new chemicals for pharmaceutical and agrochemical industries as well as the need of new environmentally sustainable processes. Despite several protein engineering techniques developed in the last decades, the discovery of new enzymes from natural sources gives still a solid and strong contribution to this theme. Hydroxynitrile lyases (HNLs) are a long-known established tools for cyanohydrin synthesis, key building blocks for several chemical and chemoenzymatic follow up reactions. However, the number of HNLs applied in industrial processes is limited and a demand of new HNLs has gained a renewed interest recently[1].

By the development of a new approach for bioprospecting HNLs, the discovering workflow is speeded up and the enzyme of interest can be identified from a protein mixture. The method is versatile for different protein preparations as and independent of the HNL class and even a first substrate scope can be explored. Notably, the integration of this assay with proteomics and transcriptomics allows the determination of the amino acid sequence of target enzymes (Chapter 2).

After twenty years from the isolation of the first fern HNL, the amino acid sequence of the HNL from the fern *Davallia tyermannii* (*Dt*HNL) was discovered (Chapter 3). Four isoenzymes (93.5% identical) were identified and named *Dt*HNL1, 2, 3 and 4. *Dt*HNL is a new 20 kDa HNL, totally different to those known, and it is only 28% identical to the closest sequence within NCBI database[i] and no HNL activity of that sequence was known.

On basis of this sequence information, it was possible to express the four isoenzymes recombinantly in *E.coli*. High-purity protein preparation was obtained by affinity chromatography for an in-depth biochemical characterization. All four isoenzymes showed high catalytic efficiency k_{cat}/K_m for the degradation of the benchmark substrate (*R*)-mandelonitrile of 438.5 ± 58.8 s^{-1} mM^{-1} (data based on the average between *Dt*HNL1, 2, 3 and 4). The reverse reaction showed an apparent K_m of 13 mM for benzaldehyde,

[i] September 2015

and high v_{max} 373 µmol min^{-1} mg^{-1} was observed for the synthesis of (*R*)-mandelonitrile (data based on *Dt*HNL1).

The tertiary structure of *Dt*HNL as well as of complexes of *Dt*HNL with different ligands was solved by high-resolution X-Ray crystallography. *Dt*HNL belongs to the Bet v 1 protein superfamily, which is composed of proteins related to major Birch (*Betula verrucose*) pollen allergen Bet v 1[2]. The fold consists of an anti-parallel β-sheet, which is wrapped around a long C-terminal α-helix. A deep ligand binding cavity is situated between the α-helix and the β-sheet. The architecture of the active site and residues involved in substrate biding were identified, furthermore the catalytic mechanism consisting of the catalytic triad Arg69, a water molecule and Tyr101 was proposed. Furthermore, Tyr117 coordinates the hydroxyl group together with Tyr101, and Asp65 interacts with the cyano group of the substrate together with the guanidinium moiety of Arg69. In conclusion, results revealed a further new strategy evolved for the catalysis of cyanohydrins degradation by fulfilling the required features described in the introduction of this thesis.

In order to acquire a more exhaustive knowledge of ferns and, more generally Bet v 1-like HNLs, we applied the earned information for the investigation of homologs HNLs in the bracken fern *P. aquilinum* and within the Bet v 1 superfamily as well. For this purpose, transcriptome of cyanogenic leaves and fiddleheads of *P. aquilinum* was sequenced. Despite the presence of proteins similar to *Dt*HNL, no HNL with Bet v 1 fold seemed to be present in bracken fern. Data retrieved from proteomics by the method described in chapter 2 confirmed the result. In conclusion, different HNLs evolved in the *pteridophytes*, explainable by the vast numbers of families and genera within the group. Furthermore, an additional new protein fold with HNL activity is expected for *P. aquilinum* HNL (*Pta*HNL). Scouting the Bet v 1 superfamily with different bioinformatic tools (e.g. 3DM) did not deliver a protein with HNL activity. According to the current knowledge, no sequence within the superfamily can satisfy the requirements for HNLs activity in a Bet v 1 – like protein. Nevertheless, the non-result revealed *Dt*HNL is not only a novel HNL, but also a unique protein within the entire superfamily composed of 13,904 sequences[ii].

[ii] According to Bet v 1 3DM database released in October 2014.

Aspects towards the application of *Dt*HNL in biocatalysis were also addressed. The substrate scope was explored using several aldehydes and a ketone. Biotransformations were carried out in a biphasic system for 24 hours. Results exhibit *Dt*HNL as a promising tool for biocatalysis. Without any enzyme optimization, high substrate conversions and enantiomeric excess of the products were achieved. The best results were obtained with the aromatic aldehydes benzaldehyde, 2-chloro-benzaldeyde as well as furan-2-carbaldehyde. Only ca. 50% ee was obtained for the more flexible substrate 3-phenylpropanal, nevertheless the introduction of one double bond on the aliphatic moiety, contributing to a more rigid chain, was sufficient to obtain product with >95% ee (3-phenyl-prop-2-enal).

The successful result led to further optimization of *Dt*HNL production and robustness, to generate a competitive and easily available biocatalyst (Chapter 4). The recombinant expression of *Dt*HNL in *Pichia pastoris* was achieved with high product levels of intracellular protein after both shake flasks and 5 L cultivation. About 80% of the total protein content was *Dt*HNL, according to the calculation of specific activity. This outstanding result was confirmed by SDS-PAGE of the cell free lysate. Due to this efficient overexpression, *Dt*HNL reached a sufficient purity for biocatalytic applications. Therefore, no expensive purification steps are necessary. Finally, *Dt*HNL was used to generate cross linked enzyme aggregate (CLEA). CLEA-*Dt*HNL showed enhanced stability at low pH values as low as pH 2.4 compared to the free enzyme. This is an essential advantage for the production of enantiomerically pure cyanohydrins. Finally, CLEA-*Dt*HNL was applied for the synthesis of (*R*)-mandelonitrile with a significant improvement in substrate conversion and product ee. Notably the enzymatic conversion of the difficult substrate hydroxypivaldeyde to the corresponding cyanohydrin was also observed at pH 2.4.

In conclusion, a truly novel HNL was discovered by combination of enzymatic assay, proteomics and transcriptomics. The enzyme is not only unique, but it reveled to be a promising biocatalyst for cyanohydrin synthesis. *Dt*HNL could find applications also for cyanogenic food plant detoxification or as natural defense system for the agricultural trade owing to its high efficiency for degradation of mandelonitrile. *Dt*HNL demonstrates how starting from scratch can provide novelty, which is certainly limited in sequence

based approaches, and how nature's biodiversity can still provide numerous resources to address different issues.

The outcome of this thesis consists of: One review published in Recent Patent on Biotechnology (Chapter 1), one published research paper in Current Biotechnology (Chapter 2) and two manuscripts (Chapter 3 and 4). Chapter 3 is the core of this thesis and consists of a manuscript intended for submission to Nature Chemical Biology as research paper. A presubmission enquiry resulted in a positive feedback from the journal. One standard operation protocol (Chapter 5). A complete list of posters, talks and IP related to this thesis is reported in the Appendix.

Coworkers Contributions: Anton Glieder from Graz University of Technology, Institute of Molecular Biotechnology, 8010 Graz, Austria and Margit Winkler & Kerstin Steiner from Austrian Centre Industrial Biotechnology, 8010 Graz Austria supported this thesis with day-to-day supervision, scientific discussions, mentoring and writing of the manuscripts. Chapter 3 was performed in close collaboration with Karl Gruber and Tea Pavkov-Keller from University of Graz, Institute of Molecular Biosciences, Nawi Graz, 8010 Graz, Austria and from Austrian Centre Industrial Biotechnology, 8010 Graz, Austria. Specifically, they determined the structure and elucidated the catalytic mechanism of the new enzyme. Finally, they contributed to the writing of the manuscript. Results of the Master Thesis of Eva-Maria Köhler from Graz University of Technology and Matthias Diepold from University of Graz contributed to part of the study in chapter 3: enzyme purification and characterization (E.M.K.) and structure determination (M.D.) respectively. Transcriptomes assembling (Chapter 3) was provided by the group of Gerhard Thallinger, Graz University of Technology, Institute of Knowledge Discovery, Bioinformatics Group, 8010 Graz, Austria. Mass spectrometry and proteomics analysis (Chapter 2 and 3) was performed by the group of Ruth Birner-Gruenberger from the Medical University of Graz, Institute of Pathology, 8010 Graz, Austria. 3DM database (Chapter 3) was built and provided by Henk-Jan Joosten and Tom Van Der Bergh from Bio-prodict BV, 6511 AA Nijmengen, The Netherlands. Chapter 5 was performed in close collaboration with Sander Van Pelt, Zainab Raghoebar and Roger A. Sheldon from CLEA Technologies, 2628 XH Delft, The Netherlands. Specifically, they developed the CLEAs and performed pH

stability experiments. Mandana Gruber-Khadjawi supervised experiments in the HCN laboratory (Chapter 3 and 4). Ivan Hajanal provided translation of Chinese patents (Chapter 1). The summer students Dorota Katarzyna Pomorska and Kladia Jadczak from Lodz university of technology supported the general lab work and completion of activity experiments: chapter 2 and 3 respectively.

My personal contributions.

Chapter 1

Collected patent literature and wrote the manuscript.

Chapter 2

Designed and performed the experiments, collected and analyzed the results and wrote the manuscript.

Chapter 3

Designed and performed the experiments, collected and analyzed the results of: bioinformatics analysis, enzyme discovery; isoenzymes identification and isolation, enzymes expression and purification; biochemical characterization; construction, expression, purification and characterization of the different mutants; analysis of the protein superfamily; investigation of *P. aquilinum* HNL and cyanohydrin synthesis. Wrote the manuscript.

Chapter 4

Designed and performed the experiments, collected and analyzed results of: strain generation, enzyme expression, bioreactor cultivation and cyanohydrin synthesis. Wrote the manuscript.

References

1. 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).
2. Radauer, C., Lackner, P. & Breiteneder, H. The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol.* Biol. **8**, 286 (2008).

Appendix

Appendix Table 1. Posters related to this thesis

TITLE	A NEW HYDROXYNITRILE LYASE FROM FERN: FROM THE PLANT TO THE SEQUENCE
AUTHORS	**E. Lanfranchi**, K. Steiner, B. Darnhofer, R. Birner-Grünberger, A. Glieder, M. Winkler
YEAR	2013
EVENT	**Biotrans**, Manchest Central, UK
	ACIB Science Days, Graz, AT
TITLE	RECOMBINANT HYDROXYNITRILE LYASE: HIGH LEVEL EXPRESSION IN *PICHIA PASTORIS*
AUTHORS	**E. Lanfranchi**, D. K. Pomorska, K. Steiner, H. Schwab, A. Glieder, R. Weis, M. Winkler
YEAR	2013
EVENT	**DocDay**, NAWI Graz Doctoral School Molecular Biology and Biochemistry & Molecular Biosciences and Biotechnology, Graz, AT
TITLE	NOVEL ENZYMES FOR BIOCATALYSIS
AUTHORS	**E. Lanfranchi**, M. Winkler, A. Glieder
YEAR	2014
EVENT	**DocDay**, NAWI Graz Doctoral School Molecular Biology and Biochemistry & Molecular Biosciences and Biotechnology, Graz, AT **Italian Forum for Industrial Biotechnology** (IFIB), Genoa, IT
TITLE	BLUE NATIVE PAGE: BIOPROSPECTING FOR NEW HYDROXYNITRILE LYASES
AUTHORS	**E. Lanfranchi**, E.M. Köhler, K. Steiner, A. Glieder, M. Winkler
YEAR	2014
EVENT	**7th International Conference on Biocatalysis (BIOCAT)**, Hamburg, DE
TITLE	BIOPROSPECTING FOR NEW HYDROXYNITRILE LYASES USING BLUE NATIVE PAGE COUPLED IN-GEL ACTIVITY ASSAY
AUTHORS	E. M. Köhler, **E. Lanfranchi**, K. Steiner, A. Glieder, M. Winkler
YEAR	2014
EVENT	**OEGMBT Annual Meeting**, Vienna, AT
TITLE	STRUCTURAL STUDIES ON THE HYDROXYNITRILE LYASE FROM *DAVALLIA TYERMANNII*
AUTHORS	M. Diepold, **E. Lanfranchi**, K. Steiner, M. Winkler, A. Glieder, T. Pavkov-Keller, K. Gruber
YEAR	2014
EVENT	**Crystal (cl)Year**, Turin, IT
TITLE	THE UNIQUE HYDROXYNITRILE LYASE FROM FERN
AUTHORS	**E. Lanfranchi**, M. Winkler, T. Pavkov-Keller, E.M. Köhler, M. Diepold, B. Darnhofer, K.Steiner, G. G. Thallinger, R. Birner-Gruenbergher, S. Van Pelt, R. A. Sheldon, K. Gruber, A. Glieder
YEAR	2015
EVENT	**Biotrans**, Vienna, AT

Appendix Table 2. Talks related to this thesis

TITLE	NON HOMOLOGOUS ENZYME DISCOVERY: THE HYDROXYNITRILE LYASES CASE
AUTHORS	**E. Lanfranchi**, K.Steiner, B. Darnhofer. R. Birner-Gruenberger, M. Winkler, A. Glieder
PRESENTER	**E. Lanfranchi**
YEAR	2015
EVENT	**Master Class Computational Approaches for Discovery and Engineering of Enzymes for Biocatalysis and Synthetic Biology**, Groningen, NL
TITLE	UNIQUE AND NOVEL ENZYMES CATALYZING CYANIDE RELEASE FROM FERN
AUTHORS	**E. Lanfranchi**
PRESENTER	**E. Lanfranchi**
YEAR	2015
EVENT	**DocDay**, NAWI Graz Doctoral School Molecular Biology and Biochemistry & Molecular Biosciences and Biotechnology, Graz AT
TITLE	HYDROXYNITRILE LYASE FROM FERN: A UNIQUE BIOCATALYST
AUTHORS	**E. Lanfranchi**, M. Winkler., T. Pavkov-Keller, E.M. Köhler, M. Diepold, B. Darnhofer, K. Steiner, G. G. Thallinger, R. Birner-Gruenberger, S. Van Pelt, K. Gruber, A. Glieder
PRESENTER	M. Winkler
YEAR	2015
EVENT	**16. Österreichische Chemietage**, Innsbruck, AT

Appendix Table 3. Intellectual Property related to this thesis

TITLE	NEW HYDROXYNITRILE LYASES
INTERNATIONAL APPLICATION NUMBER	PCT/EP2015/068262
FILING DATE	07 August 2015

Appendix Table 4. Strains of *Escherichia coli* and *Pichia pastoris* deposited in the culture collection of the Institute of Molecular Biotechnology, Graz University of Technology.

CC#	Organism	Strain	Description	Notes
6533	*E. Coli*	K12 TOP10 F'	pPpB1_alpha_S_LuHNL_opt	Optimized gene
6534	*E. Coli*	K12 TOP10 F'	pPpT4_alpha_S_LuHNL_opt	Optimized Gene
6535	*E. Coli*	K12 TOP10 F'	pPpB1_LuHNL_opt	Optimized gene
6536	*E. Coli*	K12 TOP10 F'	pPpT4_S_LuHNL_opt	Optimized Gene
6537	*P. pastoris*	CBS 7435 Mut^S	pPpB1_LuHNL_opt_2mcB12	High expressing strain
6538	*P. pastoris*	CBS 7435 Mut^S	pPpB1_alpha_LuHNL_opt_3scB4	Secretion does not work
6367	*P. pastoris*	CBS 7435 Mut^S	pPpT4_S_LuHNL_opt_4scA4	
6368	*P. pastoris*	CBS 7435 Mut^S	pPpT4_S_LuHNL_opt_4mcD10	
6369	*P. pastoris*	CBS 7435 Mut^S	pPpT4_alpha_S_LuHNL_opt_18mcB4	Secretion does not work
6456	*P. pastoris*	CBS 7435 Mut^S	pPpB1_apha_LuHNL_opt_1mcG3	Secretion does not work
7165	*E. Coli*	K12 TOP10 F'	pPpB1_DtHNL1	Optimized gene
7166	*E. Coli*	K12 TOP10 F'	pPpT4_S_DtHNL1	Optimized Gene
7167	*P. pastoris*	CBS 7435 Mut^S	DtHNL1_B1.G4	High expressing strain
7168	*P. pastoris*	CBS 7435 Mut^S	DtHNL1_T4.B7	
7169	*P. pastoris*	CBS 7435 Mut^S	DtHNL1_T4.F1	
7170	*E. Coli*	BL21 Star (DE3)	pEHisTEV_DtHNL1_opt	Optimized Gene Nterm His Tag TEV cleavage site
7171	*E. Coli*	BL21 Star (DE3)	pEHisTEV_DtHNL2	Nterm His Tag TEV cleavage site
7172	*E. Coli*	BL21 Star (DE3)	pEHisTEV_DtHNL3	Nterm His Tag TEV cleavage site
7173	*E. Coli*	BL21 Star (DE3)	pEHisTeV_DtHNL4	Nterm His Tag TEV cleavage site
7164	*E. Coli*	BL21 Gold (DE3)	pET26b(+)_DtHNL1	Optimized Gene

Appendix Table 5. List of deposited sequences to GenBank® database

Entry	GenBank Accession Number	Description	Ref. Protein (Supplementary Table 12)
1	KT804569	*Davallia tyermannii* Hydroxnyitrile lyase Isoform 1	DtHNL1
	Sequence		
	ATGGCGGGAACGGGAGGGGGCGCAGAACAGTTCCAGCTCCGGGGAGTGCTGTGGGGGAAAGCCTACTCTTGGAAGATA ACCGGAACGACAATCGACAAGGTGTGGTCGATTGTGGGCGATTATGTGCGCGTCGACAACTGGGTCTCTTCCGTCGTG AAGAGCTCGCACGTCGTGTCTGGCGAGGCCAACCAGACGGGGTGCGTGAGGAGGTTCGTCTGCTACCCAGCCTCCGAG GGAGAGTCGGAGACTGTGGACTACTCGGAGCTCATCCACATGAACGCTGCCGCGCACCAGTACATGTACATGATCGTG GGAGGTAACATCACTGGCTTCTCTCTCATGAAGAACTATGTGAGCAATATCTCGCTGTCTTCTCTTCCTGAGGAGGAC GGTGGTGGTGTAATCTTTTACTGGAGCTTCACAGCCGAGCCTGCCTCTAACCTCACGGAACAAAAATGCATAGAAATT GTGTTCCCTCTCTATACCACTGCCCTGAAGGATTTATGCACTCACCTTTCCATACCCGAAAGCTCTGTTACACTTCTC GATGATTAA		
	GenBank Accession Number	**Description**	**Ref. Protein**
2	KT805919	*Davallia tyermannii* Hydroxnyitrile lyase Isoform 2	DtHNL2
	Sequence		
	ATGGCGGGAACGAGAGGAGGCGCTGAAGAGTTCCAGCTCCGGGGAGTGCTGTGGGGGAAAGCCTACTCTTGGAAGATA ACGGGAACGACAATCGACAAGGTGTGGTCGATTGTGGGTGATTATGTGCGCGTCGACAACTGGGTCTCTTCCGTCGTG AAGAGCTCGCACGTCGTGTCCGGCGATGCCAACCAGACGGGGTGCGTGAGGAGGTTCGTCTGCTACCCAGCCTCCGAT GGAGAGTCGGAGACTGTGGACTACTCGGAGCTCATCCACATGAACGCCGCCGCTCACCAATACATGTACATGATTGTG GGAGGTAACATCACTGGCTTCTCTCTCATGAAGAACTATGTGAGCAATATCTCGCTGTCTTCTCTTCCTGAGGAGGAC GGTGGTGGTGTCATCTTTTACTGGAGCTTCACAGCCGAGCCTGCCTCTAACCTCACGGAACAAAAATGCATAGAAATT GTGTTTCCTCTCTACACCACTGCCCTGAAGGATTTATGCACTCACCTTTCCATACCCGAAAGCTCTGTTACACTTCTC GATGATTAA		
	GenBank Accession Number	**Description**	**Ref. Protein**
3	KT805921	*Davallia tyermannii* Hydroxnyitrile lyase Isoform 3	DtHNL3
	Sequence		
	ATGGCAGGAACGGGAGGGGGCGCAGAAGAGTTCCAGCTGCGGGGAGTGCTGTGGGGGAAAGCCTACTCGTGGAAGATA TCGGGAACGACAATTGACAAGGTGTGGGCGATTGTGGGCGACTATGTGCGCGTCGACAACTGGGTCTCTTCTGTAGTG AAGAGCTCGCACGTCGTGTCTGGCGACGCTAACAAGACGGGGTGCGTGAGGAGGTTCGTCTGCTACCCAGCCTCCGAG GGAGAGTCGGAGACTGTGGACTACTCGGAGCTCATCCACATGAACGCGGCCGCGCACCAGTACATGTACATGATTGTG GGAGGTAACATCACTGGCTTCTCTCTCATGAAGAACTATGTGAGCAATATATCGCTCAATTCTCTTCCTGAGGCGGAC GGTGGTGGTGTCATCCTCCACTGGAGCTTCACAGCCGAGCCTGCTTCTAACCTCACCGAACAAAAATGTATAGAAATT GTGTTCCCTCTCTATACCACTGCCTTGAAGGATTTATGCACTCACCTTTCTATTCCGGAAAGCTCTGTTACACTCCTC GGTGATTAA		
	GenBank Accession Number	**Description**	**Ref. Protein**
4		*Davallia tyermannii* Hydroxnyitrile lyase Isoform 4	DtHNL4
	Sequence		
	ATGGCAGGAACGGGAGGGGGCGCAGAAGAGTTCCAGCTGCGGGGAGTGCTGTGGGGGAAAGCCTACTCTTGGAAGATA ACGGGAACGACAATCGACAAGGTGTGGTCGATTGTGGGCGACTATGTCCGCGTCGATAACTGGGTCTCTTCCGTAGTG AAGAGCTCGCACGTCGTGTCTGGCGATGCGAACAAGACGGGGTGCGTGAGGAGGTTCGTCTGCTACCCAGCCTCCGAG GGAGAGTCGGAGACTGTGGACTACTCGGAGCTCATCCACATGAATGCGGCCGCGCACCAGTACATGTACATGATCGTG GGAGGTAACATCACTGGCTTCTCTCTCATGAAGAACTATGTGAGCAATATATCGCTCAATTCTCTTCCTGAGGCGGAC GGAGGTGGTGTCATCTTCCACTGGAGCTTCACAGCCGAGCCTGCTTCTAACCTCACCGAACAAAAATGCATCGAAATT GTGTTCCCTCTCTATACCACTGCCTTGAAGGATTTATGCACTCACCTTTCTATCCCGGAAAGCTCTGTTACACTCCTC GGTGATTAA		

	GenBanK Accession Number	Description	Ref. Protein
5	KT818577	*Davallia tyermannii* Unknown Protein	DtIsotig07200
	Sequence		
	ATGAGTTATTGGAAGAGCAAGGTTGTGCCCAAAATCAAGAAGTTTTTTGACAAGGGGAAGAAGAAAGGAGCTGCTGAG TTCTCCAAAAACTTTGATTCGTCCAAGGAGTCTTTGGACAAGGAGATAGGAGAGAAGAGTTCAGATCTGAGCCCCAAG GTTGTGGAGATATACAGATCCTCTTCCACCTTCATTGCCAAGAAGTTGCTGAAGGAACCCAATGAGGCAACAGTGAAG GAAAATTCCGATGCAACGCAAGGCGTGCTTCAGGAGCTGGCAACAGCAGGCTTTCCTGGAGCGCAGGGCATCGCTGAT GCAGGCAAAAAGTATGGACCGGCGCTTTTACCGGGGCCGGTCGTGTACTTGTTCGAGAAAGCATCCGTGTTTTTGGCG GAGGAGCCTTTGCCAGAGGAGCCCAAGGCAGAGACTAGAGAGGTGAGTGCGGAGGACGTCAAGCCAGCAGAAGCGCCA GCGACGACTTCGGAAACTCCGCCGCCGCCACCTGTAGCAGACGTGCCTCCCCCAGCTGTCGTCGAGGAAGAAAAGAAG GAAGCAGAGCCCATTGTTGCTGCCCCGCCGCCCGAAGCTGCCCCCCCTGCAGCTGTTGAGATCCCCACCTCCGTTGAC CCTACGCCCCCGCCTCCTGCTCCGCCCGCCGACAAACCTGAGTAG		
	GenBanK Accession Number	**Description**	**Ref. Protein**
6	KT818578	*Davallia tyermannii* Unknown Protein	DtIsotig06604
	Sequence		
	ATGGCAGACTCTGCTCGCACAACCGTACTCGTAACTGGTGCTGGTGGAAGAACGGGGCACATTGTATATGAGAAGCTG AAACATAAGGCAGACAAGTTTCATGTGAGAGGTCTTGTGAGGTCAGAGCCAAGCAAGGCAAAGATTGGAGGGGGTGAG GATGTGTACATAGGTGACATAACAAAGGCAGAGAGCCTGGGCCCAGCATTTGCAGGGGTGGATGTGCTCATCATCCTC ACAAGTGCTGCTCCTCAAATGAAGCCAGGGTTTGATCCAAGCAAAGGAGGGCGCCCCGAGTTCTACTACGAAGAAGGC GCATATCCTGAGCAGGTTGATTGGATCGGGCAAAAGAATCAAATCGATGCAGCCAAAGAAGCGGGGGTGAAGCATATC ATTTTGGTTGGCTCGATGGGTGGCACAAATCCGAATCATCCTCTGAACTCTTTCGGAAATGGGAAGATTTTGATTTGG AAAAGGAAGTCGGAGCAGTATTTGGCAGACTCTGATACAACCTACACCTTAATCAGGCCGGGTGGGCTGCTAGACAAA GAAGGTGGGTTGCGAGAGCTCTTGATTGGAAACAACGACGAACTTCTCGCCACAGATACCAAAACGGTTCCACGAGCA GATGTTGCGGAAGTCTGTGTACAGGCAATTGTTCATGATGCAGTAAAGAATAAGGCTTTTGACTTGACCTCGAAACCG GAGGGTGAAGGTACGCCGACGACAGATTTCAAGATTCTCTTCAGTCAAGTAACTGCAAGCTTTTAG		
	GenBanK Accession Number	**Description**	**Ref. Protein**
7	KT818579	*Davallia tyermannii* Unknown Protein	DtIsotig04379
	Sequence		
	ATGGGCACCTCAACGTGGGTGGTATGGAGCATTCTGTTGCTAGCAGTGGCGCAAGTAGCGGGGAGCATCCCCATCCCA AGACGCTACGATGGCTTCGTCTTCAATGCTTCTTCATCCTCGTCGCCTGTGTTGCTGGAGGCCTTCTTCGATCCCCTC TGCCCGGATAGCGCAGACGCTTGGCCTGTTGTCAAGAAAATCGCCCAATACTTCCAGGACGATCTGCTCCTCATTGTC CACCCCTTCCCTCTCCCGTACCATCACAATGCATATTTTGCAAGTAGAGCATTGCACATCATCAATAACCTGAACAGT TCTCTCACTTATCCATTGCTTGAGTTGTTTTTTGAAAACCAGGATAGCTTTTCAACGAGTGAAACGCTAGCGGAGGCA CCATCCTCCGTCGTAGACAGAATCGTTCAACTGGCAGCAGATAGCTTGAATGAACTCGTGTCTTCCGATTTTGAGAGC CAGTTCAAAGCAGGATTTTCTGATACAGGAACGGATCTGATCACTCGTGTTTCGTTCAAGTTTGGGTGCTCGCGCGTT GTGGTCGGTACGCCGTACTTTTTTGTCAATGGCATACCTCTTTATGATCGGATTCGGCATGGACCTTCTCTGAATGGG CAGAGATTATTGAGCCATTAACAGGGGCGCAAAAAATCGCGCTTGCATAA		
	GenBanK Accession Number	**Description**	**Ref. Protein**
8	KT818580	*Davallia tyermannii* Unknown Protein	DtIsotig04065
	Sequence		
	ATGGCCAAGGTGCACATTATGCTCCTATGTACATTATGCGCTCTCTCCCTCTCCTCCCTCTCTGCCCCACAGCCAGCT TTGGCTGCCTCGGGTCCTGACCATCTTGACTTCTACATGTACATTGCTGTTCAGAATAACAGCAATCTCGACAACCCC AATGTCACCTTCACAGCCGTGCAGTCCGCGCAGCCGCTTTCGACGCAACCCAACTCATTCGGCATCATCCACACCTTC GACAACCCACTCACCAGTGCCGCAGATCTCAACTCTACGCAGCTCGGGCACGTGCAGGGCTGGTATGGTGATGTGGGG CAGAATCTTTTGACGTTGTTCCTGGCGCAGACCTTTACCTGGAACGACGGCACCTACAATGGCACATTCAGCCTGTTG GGGGTGGACGTCGCGTCAGATGCGGTAAAGTTTGCACCCATTGTTGGCGGCACGGGCGACTTTGCGTATGTGCGCGGA GTGGCTCAACAGTCCCTTGTCTCAACCGCTACTGTAAACATGGAAACGGTCTCATGGTTCTTTTATGCCATTGACTTT GTATACTAG		

	GenBank Accession Number	Description	Ref. Protein
9	KT818581	*Davallia tyermannii* Unknown Protein	DtContig00505
	Sequence		
	ATGCCCTTTGCTCAATCCTTGATAGTGCTTTTCCTTGCGGCTTCAGCACTCAGCTATGGAGGAGTGTTGGCAACAACC ATAACGGCTGTGAACAACTGTGGAACAAGCGGCCCACTTGAGTTCACAGGCACTAGTGCCAACGGCATGAACTTGGCA CCTGCACAATCGTCTGGCCCCATCGGTGTACCTGACGGATGGTCGGGCCGAGTTTCGCTGGACCCTTCGCCGTCCACT TTAGCAGAGTTCAGCATCGTCCAAAACAACAAGAATACCATGGATATTAGTCTGGTGGATGGCTTCAACGTTGCTCTG GGAATCTCATACACCGGTGGTAATTGCATAAGGAATGGTGAAGCTGCAGCTAGCAACGTGGCATGCCACATTTCTATC GACCAGTGTCCTGCAAGCTACAGACAAGGCGACCGATGCGTCAACCCTAACAAAGACGCCCAGACTGACTACTCCGCT ACTGTGAAGGGGATATGTCCGGACGCCTATAGCTGGTCCAAGGATGATGCAACTAGCACGTTCACGTGCGATGTTGGT GGTGACTTTACCGTCACATTCTGCCCTCCATGA		
	GenBank Accession Number	Description	Ref. Protein
10	KT818582	*Pteridium aquilinum* Unknown Protein	PtaIsotig02775
	Sequence		
	ATGGAGACGATTCAAACAGCGACGGAGTCGATGACAGCGGCTAGCAGGAGCTATGGAGAGGAGGAGGTATTATGGGGG AAGGCGTTCAAGTGGGAGATAAAGGGTGTAGGGGAGGACGAGGTGTGGGAGGTAACCGGAGACTTTCTGGGAGTGGCC AGGTGGGCAACCTCGCTGGTGGAGAGCTGTGAGCTTATAGAAGGAGAGGCCCATAAGCCAGGCTGCGTGAGAAGGGTC CTTGTTTATCCCCAGGCTCCTGGGGAGGCCTCCACTTTTGCCCTTGAAAAGCTCTTAGAAATGGACGCGCTACACCAC CGTTACTCTTACACTATCCTTGGCGGAAGCACCTTGCCTGGCTTCTCTCTCATGCAGGACTATGTCAGTACCTTCAAG CTCTCTTCCCTACGCCTGGTGTACCCCTCTGCAGAAATTGACCAAGAAAATGGTACCCTCCTCCATTGGAGCTTTGTT TGTCGCCCAGTCTCTACCTTGTCTGAGGAGGAAACCCACAACATTGCCTTCTCTCTCTACCAGGCTGCAGTCAACGAT CTCAAAGCTCGCCTCTCCTTGTCTGACGACCGCATTACTCTCATCCCGTAA		

Genes List

Hydroxynitrile lyase *Linum usitatissimum*

Optimized gene: *P. pastoris*

>LuHNL_opt
ATGGCTTCTCTTCCTGTTAGCTTTGCTAAACCCGACAAGAATGGAGTCATCACTTGCAAGGCTATTATGCTGAAGGAGGCTAAGTTGC
CAGGAATGTCATACGCTGATACCGTCCAGATAATAGACATCCAAGTGGACCCGCCGCAAAATGTCGAGCTTAGGGTTAAGATGTTGTG
TGCAAGTGTGTGCAGAACCGATATTTTAACCATTGAAGGCTTCATGGCCCCGACTCAATTCCCTAAAATCAATGGGCATGAAGGTGTT
GGGATAATCGAGAGCATGGGCCCGGACACGAAGAACTTCAAAGTGGGTGACGTCATCGTGGCTCCAACGTTAGGAGAGTGCCAGACCT
GCAGCAGCTGCAGGTCCGGCCGAACCAACTTCTGCCAGAACTACGGGGCCAATGAATCGGCTCTCGAACCGGACGGTACCTCTAGATT
CTCGTACATCGATTCGGACGGTAAGAAGAAGCTCCTTTACTACAAGCTGGGATGCTCCACGTGGACGCAGTACATGGTGGTCGACTCC
AACTACGCCACCAAGCTCAACGAGATCGCACCGGAGCTCCCTCCGCCACATGGCAGCATCCTCTCATGTGCTTTTGCCACTGGATACG
GTGCTGTGTGGCTCGACGCCGCCGTCCAGGAAGGCGACTCCGTCGCCATTTTTGGAGTCGGCTCCGTCGGCATATCTGCTGTGATTGC
TGCCAAAGAGCTGAAAGCTAAGCAGATAATAGTGGTGGACAGGAACGAGTACAAGCTGAAAATGGCCATGGAACTGGGGGCCACCCAC
TGCATCAACTCCGAGAAACTACCCGAGGGGGTCACCCCTTCGCAAGCTGTTAGGAAACTCACCCCCAAGGAAGTCGGAGTCGATGCTA
GCATCGAATCCTCGGGCTACGACGTCTTCATGAACGAAGCCATGAAAGCCGCCATCCACGGGAAAGCCAAGACCGTGATTACCGGAGA
AGGAATTTACGAAAACGACCGAATCTTCTTCGATTTCAAGGACTTCCTGTTCGGCGGGAACGTAGTCGGAAACGTCACGGGTAGAGTT
AGAATCCATAGCGATTTCCCAGGGTTGCTGAGAAAGGCTCAAGAACCGGTAATCAGAGCTGGAATGGATAAAATCTTGGGGTACGATG
CTGCAACTATGAAGTGCAAGTACGAGGTCGACATTCGTGAGGGTACTCCTGCATTACTGAAAGCATTGGAAGAGGTGGAGAATGTGGA
TTGCGTCAAACTCGTGATCAAGTTGAACGATTATTGA

Hydroxynitrile lyase isoform 1 *Davallia tyermannii*

>DtIsotig02643
CTCTGGACCTTGGCTGTCACTCAAAAGAGAAGGTAAGGAGAGGAAAGCCATTAGGGGTTTGAGGAGAGTGAGGCGAGGTAGGGATTAT
TAAGGGGGCAGCTCCCTAGCAAGTCATGGCGGGAACGGGAGGGGGCGCAGAACAGTTCCAGCTCCGGGGAGTGCTGTGGGGGAAAGCC
TACTCTTGGAAGATAACCGGAACGACAATCGACAAGGTGTGGTCGATTGTGGGCGATTATGTGCGCGTCGACAACTGGGTCTCTTCCG
TCGTGAAGAGCTCGCACGTCGTGTCTGGCGAGGCCAACCAGACGGGGTGCGTGAGGAGGTTCGTCTGCTACCCAGCCTCCGAGGGAGA
GTCGGAGACTGTGGACTACTCGGAGCTCATCCACATGAACGCTGCCGCGCACCAGTACATGTACATGATCGTGGGAGGTAACATCACT
GGCTTCTCTCTCATGAAGAACTATGTGAGCAATATCTCGCTGTCTTCTCTTCCTGAGGAGGACGGTGGTGGTGTAATCTTTTACTGGA
GCTTCACAGCCGAGCCTGCCTCTAACCTCACGGAACAAAAATGCATAGAAATTGTGTTTCCTCTCTACACCACTGCCCTGAAGGATTT
ATGCACTCACCTTTCCATACCCGAAAGCTCTGTTACACTTCTCGATGATTAAGCTTTTCATCCTCCCCCCCCCCAAAAAAAAACAAAA
CAAAACAAAACAAACTTTCTGCGTTCCCTAATACTTGTGCTGCGTTTCCTATCTTTTCCTTAATGCTCTACATCTTGCGCTTCCTTAC
CCTAATAATTGCCGCTTTACACTTTCTAATAGATTTTCTGCTTTCTTGACTTGCCAAGGACTAAATATAT

>DtHNL1
ATGGCGGGAACGGGAGGGGGCGCAGAACAGTTCCAGCTCCGGGGAGTGCTGTGGGGGAAAGCCTACTCTTGGAAGATAACCGGAACGA
CAATCGACAAGGTGTGGTCGATTGTGGGCGATTATGTGCGCGTCGACAACTGGGTCTCTTCCGTCGTGAAGAGCTCGCACGTCGTGTC
TGGCGAGGCCAACCAGACGGGGTGCGTGAGGAGGTTCGTCTGCTACCCAGCCTCCGAGGGAGAGTCGGAGACTGTGGACTACTCGGAG
CTCATCCACATGAACGCTGCCGCGCACCAGTACATGTACATGATCGTGGGAGGTAACATCACTGGCTTCTCTCTCATGAAGAACTATG
TGAGCAATATCTCGCTGTCTTCTCTTCCTGAGGAGGACGGTGGTGGTGTAATCTTTTACTGGAGCTTCACAGCCGAGCCTGCCTCTAA
CCTCACGGAACAAAAATGCATAGAAATTGTGTTCCCTCTCTATACCACTGCCCTGAAGGATTTATGCACTCACCTTTCCATACCCGAA
AGCTCTGTTACACTTCTCGATGATTAA

Optimized gene: *E. coli*

>DtHNL1
ATGGCAGGCACCGGTGGTGGTGCAGAACAGTTTCAGCTGCGTGGTGTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCA
CCATTGATAAAGTTTGGAGCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCATGTTGTTAG
CGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCAAGCGAAGGTGAAAGCGAAACCGTTGATTATAGCGAA
CTGATTCACATGAATGCAGCAGCACATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTACG
TGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTATTGGAGCTTTACCGCAGAACCGGCAAGCAA
TCTGACCGAACAGAAATGTATTGAAATTGTGTTTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAA
AGCAGCGTTACCCTGCTGGATGATTAA

Optimized gene: *P. pastoris*

>DtHNL1
ATGGCGGGCACTGGAGGCGGAGCAGAGCAATTTCAACTCAGGGGCGTCTTGTGGGGAAAAGCATACTCTTGGAAGATAACGGGCACTA
CTATAGACAAGGTTTGGTCGATTGTCGGCGATTACGTTCGAGTGGACAACTGGGTTAGCAGCGTCGTGAAGTCATCTCATGTCGTATC
TGGCGAGGCTAATCAGACCGGATGCGTTAGACGTTTCGTATGCTACCCAGCCTCTGAAGGCGAGTCGGAAACTGTCGACTATAGCGAG
TTAATTCACATGAATGCAGCTGCCCACCAATACATGTACATGATTGTCGGGGGAAACATAACCGGATTTAGCCTTATGAAGAACTATG
TGTCCAACATATCTCTTAGCTCTCTTCCCGAGGAAGATGGCGGCGGGGTGATATTCTATTGGAGCTTCACAGCTGAGCCAGCTTCTAA
TTTGACAGAACAAAAATGTATTGAGATAGTTTTCCCCCTTTACACAACCGCTCTGAAGGATCTGTGCACCCACTTGAGTATCCCAGAG
TCCTCTGTAACGTTGCTCGATGACTAA

Hydroxynitrile lyase isoform 2 *Davallia tyermannii*

>DtIsotig02641
GTCTCTGGACCTTGGCTGTCACTCAAAAAAGAGAGGCGGAGGGAAGCCATTAGGAGTTGTGGGTTGAGAGAGTGAGGCGAGGTAGGGA
TTATTAAGGGGGCAAGCCATGGCGGGAACGAGAGGAGGCGCTGAAGAGTTCCAGCTCCGGGGAGTGCTGTGGGGGAAAGCCTACTCTT
GGAAGATAACGGGAACGACAATCGACAAGGTGTGGTCGATTGTGGGTGATTATGTGCGCGTCGACAACTGGGTCTCTTCCGTCGTGAA
GAGCTCGCACGTCGTGTCCGGCGATGCCAACCAGACGGGGTGCGTGAGGAGGTTCGTCTGCTACCCAGCCTCCGATGGAGAGTCGGAG
ACTGTGGACTACTCGGAGCTCATCCACATGAACGCCGCCGCTCACCAATACATGTACATGATTGTGGGAGGTAACATCACTGGCTTCT
CTCTCATGAAGAACTATGTGAGCAATATCTCGCTGTCTTCTCTTCCTGAGGAGGACGGTGGTGTGTCAATCTTTTACTGGAGCTTCAC
AGCCGAGCCTGCCTCTAACCTCACGGAACAAAAATGCATAGAAATTGTGTTTCCTCTCTACACCACTGCCCTGAAGGATTTATGCACT
CACCTTTCCATACCCGAAAGCTCTGTTACACTTCTCGATGATTAAGCTTTTCATCCTCCCCCCCCCCAAAAAAAAACAAAACAAAACA
AAACAAACTTTCTGCGTTCCCTAATACTTGTGCTGCGTTTCCTATCTTTTCCTTAATGCTCTACATCTTGCGCTTCCTTACCCTAATA
ATTGCCGCTTTACACTTTCTAATAGATTTTCTGCTTTCTTGACTTGCCAAGGACTAAATATAT
>DtHNL2
ATGGCGGGAACGAGAGGAGGCGCTGAAGAGTTCCAGCTCCGGGGAGTGCTGTGGGGGAAAGCCTACTCTTGGAAGATAACGGGAACGA
CAATCGACAAGGTGTGGTCGATTGTGGGTGATTATGTGCGCGTCGACAACTGGGTCTCTTCCGTCGTGAAGAGCTCGCACGTCGTGTC
CGGCGATGCCAACCAGACGGGGTGCGTGAGGAGGTTCGTCTGCTACCCAGCCTCCGATGGAGAGTCGGAGACTGTGGACTACTCGGAG
CTCATCCACATGAACGCCGCCGCTCACCAATACATGTACATGATTGTGGGAGGTAACATCACTGGCTTCTCTCTCATGAAGAACTATG
TGAGCAATATCTCGCTGTCTTCTCTTCCTGAGGAGGACGGTGGTGGTGTCATCTTTTACTGGAGCTTCACAGCCGAGCCTGCCTCTAA
CCTCACGGAACAAAAATGCATAGAAATTGTGTTTCCTCTCTACACCACTGCCCTGAAGGATTTATGCACTCACCTTTCCATACCCGAA
AGCTCTGTTACACTTCTCGATGATTAA

Hydroxynitrile lyase isoform 3 *Davallia tyermannii*

>DtIsotig07602
AGAATTAATGGTAGTAGGAACGAGTTGGTCCTGCCAACTATGTGCATATATTTAATAAAACCTTGGCAAACAAAACATTACTAACGGT
ATAGCAGCAATTATTAGGTGAGGTAGCTGCCGAGTTGGCTAAGGTAGAAAATGTAATTAGGGGGGTGAGATAAAGCTTAATCACCGAG
GAGTGTAACAGAGCTTTCCGGAATAGAAAGGTGAGTGCATAAATCCTTCAAGGCAGTGGTATAGAGAGGGAACACAATTTCTATACAT
TTTTGTTCGGTGAGGTTAGAAGCAGGCTCGGCTGTGAAGCTCCAGTGGAGGATGACACCACCACCGTCCGCCTCAGGAAGAGAATTGA
GCGATATATTGCTCACATAGTTCTTCATGAGAGAGAAGCCAGTGATGTTACCTCCCACAATCATGTACATGTACTGGTGCGCGGCCGC
GTTCATGTGGATGAGCTCCGAGTAGTCCACAGTCTCCGACTCTCCCTCGGAGGCTGGGTAGCAGACGAACCTCCTCACGCACCCCGTC
TTGTTAGCGTCGCCAGACACGACGTGCGAGCTCTTCACTACAGAAGAGACCCAGTTGTCGACGCGCACATAGTCGCCCACAATCGCCC
ACACCTTGTCAATTGTCGTTCCCGATATCTTCCACGAGTAGGCTTTCCCCCACAGCACTCCCCGCAGCTGGAACTCTTCTGCGCCCCC
TCCCGTTCCTGCCATGGCTTGCAAGCTTGCCTACTGCTCCTTACTCTCTCTCACTCTCTACCCTCAGCTCCCTCCTCCGACTCTACTT
TCCTTCTCTACTTCCCCAGGCTGATTGCAGCTGCTTTATAAAGTTT

>DtHNL3
ATGGCAGGAACGGGAGGGGGCGCAGAAGAGTTCCAGCTGCGGGGAGTGCTGTGGGGGAAAGCCTACTCGTGGAAGATATCGGGAACGA
CAATTGACAAGGTGTGGGCGATTGTGGGCGACTATGTGCGCGTCGACAACTGGGTCTCTTCTGTAGTGAAGAGCTCGCACGTCGTGTC
TGGCGACGCTAACAAGACGGGGTGCGTGAGGAGGTTCGTCTGCTACCCAGCCTCCGAGGGAGAGTCGGAGACTGTGGACTACTCGGAG
CTCATCCACATGAACGCGGCCGCGCACCAGTACATGTACATGATTGTGGGAGGTAACATCACTGGCTTCTCTCTCATGAAGAACTATG
TGAGCAATATATCGCTCAATTCTCTTCCTGAGGCGGACGGTGGTGGTGTCATCCTCCACTGGAGCTTCACAGCCGAGCCTGCTTCTAA
CCTCACCGAACAAAAATGTATAGAAATTGTGTTCCCTCTCTATACCACTGCCTTGAAGGATTTATGCACTCACCTTTCTATTCCGGAA
AGCTCTGTTACACTCCTCGGTGATTAA

Hydroxynitrile lyase isoform 4 *Davallia tyermannii*

>DtContig00751
ATTATTTAGCGGTGTAGCAGCAATTATTAGCTGAGGACATTAAGGGGTAGCTGCCGAGTTGGCTAAGGTAGAAAATATAATTAGGGAG
GGGTGAGATAAAGCTTAATCACCGAGGAGTGTAACAGAGCTTTCCGGGATAGAAAGGTGAGTGCATAAATCCTTCAAGGCAGTGGTAT
AGAGAGGGAACACAATTTCGATGCATTTTTGTTCGGTGAGGTTAGAAGCAGGCTCGGCTGTGAAGCTCCAGTGGAAGATGACACCACC
TCCGTCCGCCTCAGGAAGAGAATTGAGCGATATATTGCTCACATAGTTCTTCATGAGAGAGAAGCCAGTGATGTTACCTCCCACGATC
ATGTACATGTACTGGTGCGCGGCCGCATTCATGTGGATGAGCTCCGAGTAGTCCACAGTCTCCGACTCTCCCTCGGAGGCTGGGTAGC
AGACGAACCTCCTCACGCACCCCGTCTTGTTCGCATCGCCAGACACGACGTGCGAGCTCTTCACTACGGAAGAGACCCAGTTATCGAC
GCGGACATAGTCGCCCACAATCGACCACACCTTGTCGATTGTCGTTCCCGTTATCTTCCAAGAGTAGGCTTTCCCCCACAGCACTCCC
CGCAGCTGGAACTCTTCTGCGCCCCCTCCCGTTCCTGCCATGGCTTGGAAGCTTGCCTACTGCCCCTTACTCTCTCTCACTCTCTCAA
TGCGCAGATCCCTCTTCCGGCTCTACTTACCTTATCTTTTGAGTGACAGCCAAGGTCCAGAGACGAGT

>DtHNL4
ATGGCAGGAACGGGAGGGGGCGCAGAAGAGTTCCAGCTGCGGGGAGTGCTGTGGGGGAAAGCCTACTCTTGGAAGATAACGGGAACGA
CAATCGACAAGGTGTGGTCGATTGTGGGCGACTATGTCCGCGTCGATAACTGGGTCTCTTCCGTAGTGAAGAGCTCGCACGTCGTGTC
TGGCGATGCGAACAAGACGGGGTGCGTGAGGAGGTTCGTCTGCTACCCAGCCTCCGAGGGAGAGTCGGAGACTGTGGACTACTCGGAG
CTCATCCACATGAATGCGGCCGCGCACCAGTACATGTACATGATCGTGGGAGGTAACATCACTGGCTTCTCTCTCATGAAGAACTATG
TGAGCAATATATCGCTCAATTCTCTTCCTGAGGCGGACGGAGGTGGTGTCATCTTCCACTGGAGCTTCACAGCCGAGCCTGCTTCTAA
CCTCACCGAACAAAAATGCATCGAAATTGTGTTCCCTCTCTATACCACTGCCTTGAAGGATTTATGCACTCACCTTTCTATCCCGGAA
AGCTCTGTTACACTCCTCGGTGATTAA

Mutants - Hydroxynitrile lyase isoform 1 *Davallia tyermannii*

Optimized genes: *E. coli*

>HisTEV_DtHNL1_opt_R69A
ATGGCAGGCACCGGTGGTGGTGCAGAACAGTTTCAGCTGCGTGGTGTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCA
CCATTGATAAAGTTTGGAGCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCATGTTGTTAG
CGGTGAAGCAAATCAGACCGGTTGTGTTGCACGTTTTGTTTGTTATCCGGCAAGCGAAGGTGAAAGCGAAACCGTTGATTATAGCGAA
CTGATTCACATGAATGCAGCAGCACATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTACG
TGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTATTGGAGCTTTACCGCAGAACCGGCAAGCAA
TCTGACCGAACAGAAATGTATTGAAATTGTGTTTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAA
AGCAGCGTTACCCTGCTGGATGATTAA

>DtHNL1_opt_R69K
ATGGCAGGCACCGGTGGTGGTGCAGAACAGTTTCAGCTGCGTGGTGTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCA
CCATTGATAAAGTTTGGAGCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCATGTTGTTAG
CGGTGAAGCAAATCAGACCGGTTGTGTTAAACGTTTTGTTTGTTATCCGGCAAGCGAAGGTGAAAGCGAAACCGTTGATTATAGCGAA
CTGATTCACATGAATGCAGCAGCACATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTACG
TGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTATTGGAGCTTTACCGCAGAACCGGCAAGCAA
TCTGACCGAACAGAAATGTATTGAAATTGTGTTTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAA
AGCAGCGTTACCCTGCTGGATGATTAA

>DtHNL1_opt_D85A
ATGGCAGGCACCGGTGGTGGTGCAGAACAGTTTCAGCTGCGTGGTGTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCA
CCATTGATAAAGTTTGGAGCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCATGTTGTTAG
CGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCAAGCGAAGGTGAAAGCGAAACCGTTGCATATAGCGAA
CTGATTCACATGAATGCAGCAGCACATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTACG
TGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTATTGGAGCTTTACCGCAGAACCGGCAAGCAA
TCTGACCGAACAGAAATGTATTGAAATTGTGTTTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAA
AGCAGCGTTACCCTGCTGGATGATTAA

>DtHNL1_opt_S87A
ATGGCAGGCACCGGTGGTGGTGCAGAACAGTTTCAGCTGCGTGGTGTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCA
CCATTGATAAAGTTTGGAGCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCATGTTGTTAG
CGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCAAGCGAAGGTGAAAGCGAAACCGTTGATTATGCAGAA
CTGATTCACATGAATGCAGCAGCACATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTACG
TGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTATTGGAGCTTTACCGCAGAACCGGCAAGCAA
TCTGACCGAACAGAAATGTATTGAAATTGTGTTTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAA
AGCAGCGTTACCCTGCTGGATGATTAA

>DtHNL1_opt_Y101A
ATGGCAGGCACCGGTGGTGGTGCAGAACAGTTTCAGCTGCGTGGTGTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCA
CCATTGATAAAGTTTGGAGCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCATGTTGTTAG
CGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCAAGCGAAGGTGAAAGCGAAACCGTTGATTATAGCGAA
CTGATTCACATGAATGCAGCAGCACATCAGTATATGGCAATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTACG
TGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTATTGGAGCTTTACCGCAGAACCGGCAAGCAA
TCTGACCGAACAGAAATGTATTGAAATTGTGTTTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAA
AGCAGCGTTACCCTGCTGGATGATTAA

>DtHNL1_opt_Y101F
ATGGCAGGCACCGGTGGTGGTGCAGAACAGTTTCAGCTGCGTGGTGTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCA
CCATTGATAAAGTTTGGAGCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCATGTTGTTAG
CGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCAAGCGAAGGTGAAAGCGAAACCGTTGATTATAGCGAA
CTGATTCACATGAATGCAGCAGCACATCAGTATATGTTTATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTACG
TGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTATTGGAGCTTTACCGCAGAACCGGCAAGCAA
TCTGACCGAACAGAAATGTATTGAAATTGTGTTTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAA
AGCAGCGTTACCCTGCTGGATGATTAA

>DtHNL1_opt_Y117A
ATGGCAGGCACCGGTGGTGGTGCAGAACAGTTTCAGCTGCGTGGTGTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCA
CCATTGATAAAGTTTGGAGCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCATGTTGTTAG
CGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCAAGCGAAGGTGAAAGCGAAACCGTTGATTATAGCGAA
CTGATTCACATGAATGCAGCAGCACATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACGCAG
TGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTATTGGAGCTTTACCGCAGAACCGGCAAGCAA
TCTGACCGAACAGAAATGTATTGAAATTGTGTTTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAA
AGCAGCGTTACCCTGCTGGATGATTAA

>DtHNL1_opt_Y117F
ATGGCAGGCACCGGTGGTGGTGCAGAACAGTTTCAGCTGCGTGGTGTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCA
CCATTGATAAAGTTTGGAGCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCATGTTGTTAG
CGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCAAGCGAAGGTGAAAGCGAAACCGTTGATTATAGCGAA
CTGATTCACATGAATGCAGCAGCACATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTTTG
TGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTATTGGAGCTTTACCGCAGAACCGGCAAGCAA
TCTGACCGAACAGAAATGTATTGAAATTGTGTTTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAA
AGCAGCGTTACCCTGCTGGATGATTAA

>DtHNL1_opt_Y161A
ATGGCAGGCACCGGTGGTGGTGCAGAACAGTTTCAGCTGCGTGGTGTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCA
CCATTGATAAAGTTTGGAGCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCATGTTGTTAG
CGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCAAGCGAAGGTGAAAGCGAAACCGTTGATTATAGCGAA
CTGATTCACATGAATGCAGCAGCACATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTACG
TGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTATTGGAGCTTTACCGCAGAACCGGCAAGCAA
TCTGACCGAACAGAAATGTATTGAAATTGTGTTTCCGCTGGCAACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAA
AGCAGCGTTACCCTGCTGGATGATTAA

>DtHNL1_opt_Y161F
ATGGCAGGCACCGGTGGTGGTGCAGAACAGTTTCAGCTGCGTGGTGTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCA
CCATTGATAAAGTTTGGAGCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCATGTTGTTAG
CGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCAAGCGAAGGTGAAAGCGAAACCGTTGATTATAGCGAA
CTGATTCACATGAATGCAGCAGCACATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTACG
TGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTATTGGAGCTTTACCGCAGAACCGGCAAGCAA
TCTGACCGAACAGAAATGTATTGAAATTGTGTTTCCGCTGTTTACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAA
AGCAGCGTTACCCTGCTGGATGATTAA

>DtHNL1_opt_D85S_S87D
ATGGCAGGCACCGGTGGTGGTGCAGAACAGTTTCAGCTGCGTGGTGTTCTGTGGGGTAAAGCATATAGCTGGAAAATTACCGGCACCA
CCATTGATAAAGTTTGGAGCATTGTTGGTGATTATGTGCGTGTTGATAATTGGGTTAGCAGCGTTGTTAAAAGCAGCCATGTTGTTAG
CGGTGAAGCAAATCAGACCGGTTGTGTTCGTCGTTTTGTTTGTTATCCGGCAAGCGAAGGTGAAAGCGAAACCGTTAGCTATGATGAA
CTGATTCACATGAATGCAGCAGCACATCAGTATATGTATATGATTGTGGGTGGCAACATTACCGGTTTTAGCCTGATGAAAAACTACG
TGAGCAATATTAGCCTGAGCAGCCTGCCGGAAGAGGATGGTGGTGGCGTTATCTTTTATTGGAGCTTTACCGCAGAACCGGCAAGCAA
TCTGACCGAACAGAAATGTATTGAAATTGTGTTTCCGCTGTATACCACCGCACTGAAAGACCTGTGTACCCATCTGAGCATTCCGGAA
AGCAGCGTTACCCTGCTGGATGATTAA

Inactive Sequences

Proteins from *Davallia tyermannii* with unknown function

>DtIso06604
ATGGCAGACTCTGCTCGCACAACCGTACTCGTAACTGGTGCTGGTGGAAGAACGGGGCACATTGTATATGAGAAGCTGAAACATAAGG
CAGACAAGTTTCATGTGAGAGGTCTTGTGAGGTCAGAGCCAAGCAAGGCAAAGATTGGAGGGGGTGAGGATGTGTACATAGGTGACAT
AACAAAGGCAGAGAGCCTGGGCCCAGCATTTGCAGGGGTGGATGTGCTCATCATCCTCACAAGTGCTGCTCCTCAAATGAAGCCAGGG
TTTGATCCAAGCAAAGGAGGGCGCCCCGAGTTCTACTACGAAGAAGGCGCATATCCTGAGCAGGTTGATTGGATCGGGCAAAAGAATC
AAATCGATGCAGCCAAAGAAGCGGGGGTGAAGCATATCATTTTGGTTGGCTCGATGGGTGGCACAAATCCGAATCATCCTCTGAACTC
TTTCGGAAATGGGAAGATTTTGATTTGGAAAAGGAAGTCGGAGCAGTATTTGGCAGACTCTGATACAACCTACACCTTAATCAGGCCG
GGTGGGCTGCTAGACAAAGAAGGTGGGTTGCGAGAGCTCTTGATTGGAAACAACGACGAACTTCTCGCCACAGATACCAAAACGGTTC
CACGAGCAGATGTTGCGGAAGTCTGTGTACAGGCAATTGTTCATGATGCAGTAAAGAATAAGGCTTTTGACTTGACCTCGAAACCGGA
GGGTGAAGGTACGCCGACGACAGATTTCAAGATTCTCTTCAGTCAAGTAACTGCAAGCTTTTAG

Gene optimized: *E. coli*

>DtIso06604_opt
ATGGCAGATAGCGCACGTACCACCGTTCTGGTTACCGGTGCCGGTGGTCGTACCGGTCATATTGTTTATGAAAAACTGAAACATAAAG
CCGACAAATTTCATGTTCGTGGTCTGGTTCGTAGCGAACCGAGCAAAGCAAAAATTGGTGGTGGTGAAGATGTGTATATTGGCGATAT
TACCAAAGCAGAAAGCCTGGGTCCGGCATTTGCCGGTGTTGATGTTCTGATTATTCTGACCAGCGCAGCACCGCAGATGAAACCGGGT
TTTGATCCGAGTAAAGGTGGTCGTCCTGAATTTTACTATGAAGAAGGTGCATATCCGGAACAGGTTGATTGGATTGGCCAGAAAAATC
AGATTGATGCAGCAAAAGAAGCAGGCGTTAAACATATTATTCTGGTTGGTAGCATGGGTGGCACCAATCCGAATCATCCGCTGAATAG
TTTTGGCAATGGCAAAATTCTGATCTGGAAACGTAAAAGCGAACAGTATCTGGCCGATAGCGATACCACCTATACCCTGATTCGTCCG
GGTGGTCTGCTGGATAAAGAAGGTGGCCTGCGCGAACTGCTGATTGGTAATAATGATGAACTGCTGGCAACCGATACCAAAACCGTTC
CGCGTGCAGATGTTGCCGAAGTTTGTGTTCAGGCAATTGTTCATGATGCCGTTAAAAACAAAGCCTTTGATCTGACCAGTAAACCGGA
AGGTGAAGGCACCCCGACCACCGATTTCAAAATCCTGTTTAGCCAGGTTACCGCCAGCTTTTAA

>DtIso07200
ATGAGTTATTGGAAGAGCAAGGTTGTGCCCAAAATCAAGAAGTTTTTTGACAAGGGGAAGAAGAAAGGAGCTGCTGAGTTCTCCAAAA
ACTTTGATTCGTCCAAGGAGTCTTTGGACAAGGAGATAGGAGAGAAGAGTTCAGATCTGAGCCCCAAGGTTGTGGAGATATACAGATC
CTCTTCCACCTTCATTGCCAAGAAGTTGCTGAAGGAACCCAATGAGGCAACAGTGAAGGAAAATTCCGATGCAACGCAAGGCGTGCTT
CAGGAGCTGGCAACAGCAGGCTTTCCTGGAGCGCAGGGCATCGCTGATGCAGGCAAAAAGTATGGACCGGCGCTTTTACCGGGGCCGG
TCGTGTACTTGTTCGAGAAAGCATCCGTGTTTTTGGCGGAGGAGCCTTTGCCAGAGGAGCCCAAGGCAGAGACTAGAGAGGTGAGTGC
GGAGGACGTCAAGCCAGCAGAAGCGCCAGCGACGACTTCGGAAACTCCGCCGCCGCCACCTGTAGCAGACGTGCCTCCCCCAGCTGTC
GTCGAGGAAGAAAAGAAGGAAGCAGAGCCCATTGTTGCTGCCCCGCCGCCCGAAGCTGCCCCCCCTGCAGCTGTTGAGATCCCCACCT
CCGTTGACCCTACGCCCCCGCCTCCTGCTCCGCCCGCCGACAAACCTGAGTAG

Gene optimized: *E. coli*

>DtIso07200
CATATGAGCTACTGGAAAAGCAAAGTGGTGCCGAAAATCAAAAAATTCTTCGATAAAGGCAAAAAAAAAGGCGCAGCCGAGTTTAGCA
AAAACTTTGATAGCAGCAAAGAGAGCCTGGATAAAGAAATTGGTGAAAAAAGCAGCGATCTGAGCCCGAAAGTTGTTGAAATTTATCG
TAGCAGCAGCACCTTTATCGCAAAAAAACTGCTGAAAGAACCGAATGAAGCCACCGTGAAAGAAAATAGTGATGCAACCCAGGGTGTT
CTGCAAGAACTGGCAACCGCAGGTTTTCCGGGTGCACAGGGTATTGCAGATGCTGGCAAAAAATACGGTCCGGCACTGCTGCCTGGTC
CGGTTGTTTACCTGTTTGAAAAAGCAAGCGTTTTTCTGGCAGAAGAACCGCTGCCGGAAGAACCTAAAGCAGAAACCCGTGAAGTTAG
CGCAGAAGATGTTAAACCGGCAGAAGCACCGGCAACCACCAGTGAAACACCGCCTCCGCCTCCGGTTGCAGATGTTCCGCCACCGGCA
GTGGTTGAAGAAGAAAAAAAAGAAGCCGAACCGATTGTTGCAGCACCTCCGCCTGAAGCAGCACCGCCTGCAGCAGTTGAAATTCCGA
CCAGCGTTGATCCGACCCCTCCTCCTCCGGCACCGCCAGCAGATAAACCGGAATAAAAGCTT

>DtContig00505
ATGCCCTTTGCTCAATCCTTGATAGTGCTTTTCCTTGCGGCTTCAGCACTCAGCTATGGAGGAGTGTTGGCAACAACCATAACGGCTG
TGAACAACTGTGGAACAAGCGGCCCACTTGAGTTCACAGGCACTAGTGCCAACGGCATGAACTTGGCACCTGCACAATCGTCTGGCCC
CATCGGTGTACCTGACGGATGGTCGGGCCGAGTTTCGCTGGACCCTTCGCCGTCCACTTTAGCAGAGTTCAGCATCGTCCAAAACAAC
AAGAATACCATGGATATTAGTCTGGTGGATGGCTTCAACGTTGCTCTGGGAATCTCATACACCGGTGGTAATTGCATAAGGAATGGTG
AAGCTGCAGCTAGCAACGTGGCATGCCACATTTCTATCGACCAGTGTCCTGCAAGCTACAGACAAGGCGACCGATGCGTCAACCCTAA
CAAAGACGCCCAGACTGACTACTCCGCTACTGTGAAGGGGATATGTCCGGACGCCTATAGCTGGTCCAAGGATGATGCAACTAGCACG
TTCACGTGCGATGTTGGTGGTGACTTTACCGTCACATTCTGCCCTCCATGA

Gene optimized: *E. coli*

>DtContig00505
ATGCCGTTTGCACAGAGCCTGATTGTTCTGTTTCTGGCAGCAAGCGCACTGAGCTATGGTGGTGTTCTGGCAACCACCATTACCGCAG
TGAATAATTGTGGCACCAGTGGTCCGCTGGAATTTACCGGCACCAGCGCAAATGGTATGAATCTGGCACCGGCACAGAGCAGCGGTCC
GATTGGTGTTCCGGATGGTTGGAGCGGTCGTGTTAGCCTGGATCCGAGCCCGAGCACCCTGGCAGAATTTAGCATTGTTCAGAATAAC
AAAAACACGATGGATATCAGCCTGGTGGATGGTTTTAATGTTGCACTGGGTATTAGCTATACCGGTGGTAATTGTATTCGTAATGGTG
AAGCAGCAGCAAGCAATGTTGCATGTCATATTAGTATTGATCAGTGCCCTGCAAGCTATCGTCAGGGTGATCGTTGTGTTAATCCGAA
TAAAGATGCACAGACCGATTATAGCGCAACCGTTAAAGGTATTTGTCCGGATGCATATAGCTGGTCAAAAGATGATGCAACCAGCACC
TTTACCTGTGATGTTGGTGGTGATTTCACCGTTACCTTTTGTCCGCCTTAA

>DtIso04065
ATGGCCAAGGTGCACATTATGCTCCTATGTACATTATGCGCTCTCTCCCTCTCCTCCCTCTCTGCCCCACAGCCAGCTTTGGCTGCCT
CGGGTCCTGACCATCTTGACTTCTACATGTACATTGCTGTTCAGAATAACAGCAATCTCGACAACCCCAATGTCACCTTCACAGCCGT
GCAGTCCGCGCAGCCGCTTTCGACGCAACCCAACTCATTCGGCATCATCCACACCTTCGACAACCCACTCACCAGTGCCGCAGATCTC
AACTCTACGCAGCTCGGGCACGTGCAGGGCTGGTATGGTGATGTGGGGCAGAATCTTTTGACGTTGTTCCTGGCGCAGACCTTTACCT
GGAACGACGGCACCTACAATGGCACATTCAGCCTGTTGGGGGTGGACGTCGCGTCAGATGCGGTAAAGTTTGCACCCATTGTTGGCGG
CACGGGCGACTTTGCGTATGTGCGCGGAGTGGCTCAACAGTCCCTTGTCTCAACCGCTACTGTAAACATGGAAACGGTCTCATGGTTC
TTTTATGCCATTGACTTTGTATACTAG
Gene optimized: *E. coli*

>DtIso04065
ATGGCCAAAGTGCATATTATGCTGCTGTGTACCCTGTGTGCACTGAGCCTGAGCAGCCTGAGTGCACCGCAGCCTGCACTGGCAGCAA
GCGGTCCGGATCATCTGGATTTCTATATGTATATTGCCGTGCAGAATAACAGCAATCTGGATAATCCGAATGTTACCTTTACCGCAGT
TCAGAGCGCACAGCCGCTGAGCACCCAGCCGAATAGCTTTGGTATTATTCATACCTTTGATAACCCGCTGACCAGCGCAGCCGATCTG
AATAGCACACAGCTGGGTCATGTTCAGGGTTGGTATGGTGATGTTGGTCAGAATCTGCTGACCCTGTTTCTGGCACAGACCTTTACCT
GGAATGATGGCACCTATAACGGCACCTTTAGCCTGCTGGGTGTTGATGTTGCAAGTGATGCAGTTAAATTTGCACCGATTGTTGGTGG
CACCGGTGATTTTGCCTATGTTCGTGGTGTTGCACAGCAGAGCCTGGTTAGCACCGCAACCGTTAATATGGAAACCGTTAGCTGGTTT
TTCTATGCCATCGATTTCGTGTACTAA

>DtIso04379
ATGGGCACCTCAACGTGGGTGGTATGGAGCATTCTGTTGCTAGCAGTGGCGCAAGTAGCGGGGAGCATCCCCATCCCAAGACGCTACG
ATGGCTTCGTCTTCAATGCTTCTTCATCCTCGTCGCCTGTGTTGCTGGAGGCCTTCTTCGATCCCCTCTGCCCGGATAGCGCAGACGC
TTGGCCTGTTGTCAAGAAAATCGCCCAATACTTCCAGGACGATCTGCTCCTCATTGTCCACCCCTTCCCTCTCCCGTACCATCACAAT
GCATATTTTGCAAGTAGAGCATTGCACATCATCAATAACCTGAACAGTTCTCTCACTTATCCATTGCTTGAGTTGTTTTTTGAAAACC
AGGATAGCTTTTCAACGAGTGAAACGCTAGCGGAGGCACCATCCTCCGTCGTAGACAGAATCGTTCAACTGGCAGCAGATAGCTTGAA
TGAACTCGTGTCTTCCGATTTTGAGAGCCAGTTCAAAGCAGGATTTTCTGATACAGGAACGGATCTGATCACTCGTGTTTCGTTCAAG
TTTGGGTGCTCGCGCGTTGTGGTCGGTACGCCGTACTTTTTTGTCAATGGCATACCTCTTTATGATGCGGATTCGGCATGGACCTTCT
CTGAATGGGCAGAGATTATTGAGCCATTAACAGGGGCGCAAAAAATCGCGCTTGCATAA

Gene optimized: *E. coli*

>DtIso04379
CATATGGGCACCAGCACCTGGGTTGTTTGGAGCATTCTGCTGCTGGCAGTTGCACAGGTTGCAGGTAGCATTCCGATTCCGCGTCGTT
ATGATGGTTTTGTTTTTAATGCAAGCAGCAGCAGCTCACCGGTTCTGCTGGAAGCATTTTTTGATCCGCTGTGTCCGGATAGCGCAGA
TGCATGGCCTGTTGTGAAAAAAATCGCACAGTATTTTCAGGATGACCTGCTGCTGATTGTTCATCCGTTTCCGCTGCCGTATCATCAT
AATGCATATTTTGCAAGCCGTGCCCTGCATATTATCAATAATCTGAATAGCAGCCTGACCTATCCGCTGCTGGAACTGTTTTTTGAAA
ATCAGGATAGCTTTAGCACCAGCGAAACCCTGGCAGAAGCACCGAGCAGCGTTGTTGATCGTATTGTTCAGCTGGCAGCAGATAGCCT
GAATGAACTGGTTAGCAGCGATTTTGAAAGCCAGTTTAAAGCAGGTTTTAGCGATACCGGCACCGATCTGATTACCCGTGTTAGCTTT
AAATTCGGTTGTAGCCGTGTTGTTGTTGGCACCCCGTATTTTTTCGTTAATGGTATTCCGCTGTATGATGCAGATAGCGCATGGACCT
TTAGCGAATGGGCAGAAATTATTGAACCGCTGACCGGTGCACAGAAAATTGCACTGGCATAAAAGCTT

Protein from *Pteridium aquilinum* with unknown function

>PtaIso02775
ATGGAGACGATTCAAACAGCGACGGAGTCGATGACAGCGGCTAGCAGGAGCTATGGAGAGGAGGAGGTATTATGGGGGAAGGCGTTCA
AGTGGGAGATAAAGGGTGTAGGGGAGGACGAGGTGTGGGAGGTAACCGGAGACTTTCTGGGAGTGGCCAGGTGGGCAACCTCGCTGGT
GGAGAGCTGTGAGCTTATAGAAGGAGAGGCCCATAAGCCAGGCTGCGTGAGAAGGGTCCTTGTTTATCCCCAGGCTCCTGGGGAGGCC
TCCACTTTTGCCCTTGAAAAGCTCTTAGAAATGGACGCGCTACACCACCGTTACTCTTACACTATCCTTGGCGGAAGCACCTTGCCTG
GCTTCTCTCTCATGCAGGACTATGTCAGTACCTTCAAGCTCTCTTCCCTACGCCTGGTGTACCCCTCTGCAGAAATTGACCAAGAAAA
TGGTACCCTCCTCCATTGGAGCTTTGTTTGTCGCCCAGTCTCTACCTTGTCTGAGGAGGAAACCCACAACATTGCCTTCTCTCTCTAC
CAGGCTGCAGTCAACGATCTCAAAGCTCGCCTCTCCTTGTCTGACGACCGCATTACTCTCATCCCGTAA

Genes optimized: *E. coli*

>PtaIso02775
ATGGAAACCATTCAGACCGCAACCGAAAGCATGACCGCAGCAAGCCGTAGCTATGGTGAAGAAGAGGTTCTTTGGGGTAAAGCCTTTA
AATGGGAAATTAAAGGTGTGGGCGAAGATGAAGTTTGGGAAGTTACCGGTGATTTTCTGGGTGTTGCACGTTGGGCAACCAGCCTGGT
GGAAAGCTGTGAACTGATTGAAGGTGAAGCACATAAACCGGGTTGTGTTCGTCGTGTTCTGGTTTATCCGCAGGCACCGGGTGAAGCA
AGCACCTTTGCACTGGAAAAACTGCTGGAAATGGATGCACTGCATCATCGTTATAGTTATACCATTCTGGGTGGTAGCACCCTGCCTG
GTTTTAGCCTGATGCAGGATTATGTTAGCACCTTTAAACTGAGCAGCCTGCGTCTGGTGTATCCGAGCGCAGAAATTGATCAAGAAAA
TGGCACCCTGCTGCATTGGAGCTTTGTTTGTCGTCCGGTGAGCACCCTGAGCGAAGAAGAAACCCATAATATTGCATTTAGCCTGTAT
CAGGCAGCCGTGAATGATCTGAAAGCACGTCTGAGCCTGAGTGATGATCGTATTACCCTGATTCCGTAA

>Ptaiso02775_A92S
ATGGAAACCATTCAGACCGCAACCGAAAGCATGACCGCAGCAAGCCGTAGCTATGGTGAAGAAGAGGTTCTTTGGGGTAAAGCCTTTA
AATGGGAAATTAAAGGTGTGGGCGAAGATGAAGTTTGGGAAGTTACCGGTGATTTTCTGGGTGTTGCACGTTGGGCAACCAGCCTGGT
GGAAAGCTGTGAACTGATTGAAGGTGAAGCACATAAACCGGGTTGTGTTCGTCGTGTTCTGGTTTATCCGCAGGCACCGGGTGAAGCA
AGCACCTTTAGCCTGGAAAAACTGCTGGAAATGGATGCACTGCATCATCGTTATAGTTATACCATTCTGGGTGGTAGCACCCTGCCTG
GTTTTAGCCTGATGCAGGATTATGTTAGCACCTTTAAACTGAGCAGCCTGCGTCTGGTGTATCCGAGCGCAGAAATTGATCAAGAAAA
TGGCACCCTGCTGCATTGGAGCTTTGTTTGTCGTCCGGTGAGCACCCTGAGCGAAGAAGAAACCCATAATATTGCATTTAGCCTGTAT
CAGGCAGCCGTGAATGATCTGAAAGCACGTCTGAGCCTGAGTGATGATCGTATTACCCTGATTCCGTAA

>Ptaiso02775_A92D_E94S
ATGGAAACCATTCAGACCGCAACCGAAAGCATGACCGCAGCAAGCCGTAGCTATGGTGAAGAAGAGGTTCTTTGGGGTAAAGCCTTTA
AATGGGAAATTAAAGGTGTGGGCGAAGATGAAGTTTGGGAAGTTACCGGTGATTTTCTGGGTGTTGCACGTTGGGCAACCAGCCTGGT
GGAAAGCTGTGAACTGATTGAAGGTGAAGCACATAAACCGGGTTGTGTTCGTCGTGTTCTGGTTTATCCGCAGGCACCGGGTGAAGCA
AGCACCTTTGATCTGAGCAAACTGCTGGAAATGGATGCACTGCATCATCGTTATAGTTATACCATTCTGGGTGGTAGCACCCTGCCTG
GTTTTAGCCTGATGCAGGATTATGTTAGCACCTTTAAACTGAGCAGCCTGCGTCTGGTGTATCCGAGCGCAGAAATTGATCAAGAAAA
TGGCACCCTGCTGCATTGGAGCTTTGTTTGTCGTCCGGTGAGCACCCTGAGCGAAGAAGAAACCCATAATATTGCATTTAGCCTGTAT
CAGGCAGCCGTGAATGATCTGAAAGCACGTCTGAGCCTGAGTGATGATCGTATTACCCTGATTCCGTAA

Appendix Table 6. Primers List

Entry	Primer Name	Sequence	Purpose
1	LuHNL_fw	AATGCGAATTCGCCACCATGGCTTCTCTTCCTGTTAGCTTTG	Cloning pPpB1-pPpT4S
2	LuHNL_rev	AATGCGCGGCCGCTCAATAATCGTTCAACTTGATC	Cloning pPpB1-pPpT4S
3	LuHNL_alpha_fw	AATTCTCGAGAAGAGAGAGGCCGAAGCTGCTTCTCTTCCTGTTAGCTTTG	Cloning pPpB1-pPpT4S
4	LuHNL_intra_fw	ATTCGGACGGTAAGAAGAAG	Sequencing
5	DtIsotig02643_fw	AGCTCCCTAGCAAGTCATG	amplification from gDNA
6	DtIsotig02641_fw	AGAGAGTGAGGCGAGGTAG	amplification from gDNA
7	DtIsotig02641/3_rev	GGAGGATGAAAAGCTTAATC	amplification from gDNA
8	DtIsotig07602_fw	TAGAAAATGTAATTAGGGGGGTGAGATAAAG	amplification from gDNA
9	DtIsotig07602_rev	GAGAGTAAGGAGCAGTAGGCAAGC	amplification from gDNA
10	DtContig00751_fw	TATAATTAGGGAGGGGTGAGATAAAGC	amplification from gDNA
11	DtContig00751_rev	GTAAGGGGCAGTAGGCAAGC	amplification from gDNA
12	DtHNL1_Ec_NcoI_fw	AATGCCCATGGCAGGCACCGGTGGTG	Cloning pEHisTEV
13	DtHNL1_Ec_HindIII_rev	AATGCAAGCTTTTAATCATCCAGCAGGGTAACGC	Cloning pEHisTEV
14	DtHNL2_Dt_NcoI_fw	AATGCCCATGGCGGGAACGAGAGGAGGCG	Cloning pEHisTEV
15	DtHNL2_Dt_HindIII_rev	AATGCAAGCTTTTAATCATCGAGAAGTGTAACAGAGC	Cloning pEHisTEV
16	DtHNL3/4_Dt_NcoI_fw	AATGCCCATGGCAGGAACGGGAGGGGGC	Cloning pEHisTEV
17	DtHNL3/4_Dt_HindIII_rev	AATGCAAGCTTTTAATCACCGAGGAGTGTAACAG	Cloning pEHisTEV
18	PtaIsotig02775_Ec_NcoI_fw	AATGCCCATGGAAACCATTCAGACC	Cloning pEHisTEV
19	PtaIsotig02775_Ec_HindIII_rev	AATGCAAGCTTTTACGGAATCAGGGTAATAC	Cloning pEHisTEV
20	pEHisTEV_DtHNL1_gibson_fw	TACCCATCTGAGCATTCCGGAAAG	Plasmid amplification: pEHistTEV_DtHNL1_Ec
21	pEHisTEV_DtHNL1_gibson_rev	CAGCTATATGCTTTACCCCACAGAAC	Plasmid amplification: pEHistTEV_DtHNL1_Ec
22	pEHisTEV_PtaIsotig02775_gibson_fw	ATATTGCATTTAGCCTGTATCAGGCAG	Plasmid amplification: pEHistTEV_PtaIso02775_Ec
23	pEHisTEV_PtaIsotig02775_gibson_rev	AGGCTTTACCCCAAAGAACCTCTTC	Plasmid amplification: pEHistTEV_PtaIso02775_Ec
24	pEHisTEVseq1	CTTTAATAGTGGACTCTTGTTC	Sequencing
25	pEHisTEVseq2	GTTTATGCATTTCTTTCCAGAC	Sequencing
26	pEHisTEVseq3	GCAAGACGTTTCCCGTTGAATATG	Sequencing
27	pEHisTEVseq4	AGATACCTACAGCGTGAGCTATG	Sequencing
28	pEHisTEVseq5	GTGACTGGGTCATGGCTGCG	Sequencing
29	pEHisTEVseq6	TTCCACAGGGTAGCCAGCAGCATC	Sequencing
30	pEHisTEVseq7	TTGAAGGCTCTCAAGGGCATCG	Sequencing
31	pEHisTEVseq8	CGGCTGAATTTGATTGCGAGTG	Sequencing
32	pEHisTEVseq9	CACTTTTTCCCGCGTTTTCGCAG	Sequencing
33	pEHisTEVseq10	GGAATTGTGAGCGGATAACAATTC	Sequencing

Appendix Dataset 1. Proteomics *D. tyermannii*

Entry	Sequence name	Signal Peptide	Target	Molecular weight (Da)	Theoretical pI	Molecular weight w/o signal peptide	Putative conserved domains	Blastp accesion number	Description	Identity	Query Cover	E value
1	isotig02643 gene=isogroup00558 length=862 numContigs=2			20142,7	4,77		SRPBCC superfamily	XP_009405224.1	PREDICTED: lachrimatory-factor synthase-like [*Musa acuminata* subsp. *Malaccensis*]	28	72	8,00E-12
2	isotig06604 gene=isogroup02899 length=1098 numContigs=1			27397	6,09		SDR superfamily - NAD_binding_10 multi domain	NP_001051733.1	Os03g0822200 [*Oryza sativa Japonica* Group]	77	100	6,00E-141
3	isotig07200 gene=isogroup03495 length=921 numContigs=1			23517,6	4,86		DREPP multi domain	XP_006859057.2	PREDICTED: plasma membrane-associated cation-binding protein 1 [*Amborella trichopoda*]	43	100	1,00E-35
4	contig00505 gene=isogroup00001 length=912	24 residues	Secret. Pathway	19795	4,25	17373	GH64-TLP-SF superfamily	XP_010555234.1	PREDICTED: thaumatin-like protein 1b [*Tarenaya hassleriana*] [*Cleome hassleriana*]	35	96	1,00E-20
5	isotig04065 gene=isogroup01241 length=1066 numContigs=2	28 residues	Secret. Pathway	19759	4,24	16893,6	Dirigent superfamily	XP_002313728.1	Disease resistance-responsive family protein [*Populus trichocarpa*]	48	55	6,00E-19
6	isotig04379 gene=isogroup01398 length=846 numContigs=2	21 residues	Secret. Pathway	24700	4,45	22516,3	Thioredoxin_like superfamily - DsbG muti domain	XP_002971933.1	Hypotetical protein SELMODRAFT_270941 [*Selaginella moelledorffii*]	49	93	2,00E-68
7	isotig06433 gene=isogroup02728 length=1175 numContigs=1			32467,2	5,4		No putative conserved domains have been detected	XP_002969319.1	Hypotetical protein SELMODRAFT_41091 [*Selaginella moelledorffii*]	26	90	4,00E-07
8	isotig02919 gene=isogroup00663 length=1492 numContigs=3			41594,6	5,31		NDB_sugar-kinase_HSP70_actin superfamily - Actin multi domain	AAC64127.1	Actin [*Anemia phyllitidis*]	99	100	0,00E+00
9	isotig06737 gene=isogroup03032 length=1052 numContigs=1			35356,2	4,77		Inihibitor_I29 superfamily (N-term) & Peptidase_C1 superfamily	KCW76106.1	Hypotetical protein EUGRSUZ_D00484 [*Eucalyptus grandis*]	53	96	6,00E-100
10	isotig06929 gene=isogroup03224 length=996 numContigs=1			33769,5	6,11		Aldo_ket_red superfamily	ABR18140.1	Unknown [*Picea sitchensis*]	57	98	3,00E-115

11	isotig02628 gene=isogroup00553 length=1468 numContigs=2	22 residues	Secret. Pathway	39951,4	4,71	37388,3	Inihibitor_I29 superfamily (N-term) & Peptidase_C1 superfamily	XP_001775992.1	Predicted protein [*Physcomitrella patens*]	64	88	4,00E-150
12	isotig00559 gene=isogroup00044 length=1431 numContigs=6	20 residues	Secret. Pathway	42107,8	4,88	39986,2	AmyAc_family superfamily	XP_009605413.1	PREDICTED: alpha-galactosidase [Nicotiana tomentosiformis]	71	96	0,00E+00
13	contig00187 gene=isogroup00001 length=820	39 residues	Chloroplast	23052,6	9,76	19274,1	PsbQ superfamily	KFK32219.1	Hypohetical protein AALP_AA6G2133000 [*Arabis alpina*]	63	84	6,00E-72
14	isotig04319 gene=isogroup01368 length=877 numContigs=2	56 residues	Chloroplast	≥27311.4	n.d.	³21776	SugarP_isomerase superfamily	XP_002983271.1	Hypotetical protein SELMODRAFT_117990 [*Selaginella moellendorffii*]	91	76	1,00E-126
15	isotig09166 gene=isogroup05461 length=652 numContigs=1	16 residues	Chloroplast	≥19538.3	n.d.	³18104.6	SDR superfamily	ABK23403.1	Unknown [*Picea sitchensis*]	72	55	1,00E-32
16	isotig07085 gene=isogroup03380 length=950 numContigs=1	63 residues	Chloroplast	≥24234.7	n.d.	³17751.3	Periplasmic_Binding_Protein_Type_2 superfamily	XP_011626606.1	PREDICTED: porphobilinogen deaminase, chloroplastic [*Amborella trichopoda*]	77	86	1,00E-98
17	isotig01910 gene=isogroup00325 length=1005 numContigs=3	50 residues	Chloroplast	20663,1	6,03	15735,4	Cu-Zn_Superoxide_Dismutase superfamily	AEO27875.1	SOD2 [*Scutellaria baicalensis*]	80	89	1,00E-100
18	isotig00600 gene=isogroup00049 length=1246 numContigs=6	43 residues	Chloroplast	34986,6	5,17	30693,6	PLN00037 superfamily - MSP mutli domain	XP_006338257.1	PREDICTED: choloroplast manganese stabilizing protein-II [*Solanum tuberosum*]	70	100	4,00E-164
19	isotig06364 gene=isogroup02659 length=1223 numContigs=1	70 residues	Chloroplast	35200	5,13	27361,9	Abhydrolase_5 multi domain	ABK22367.1	Unknown [*Picea sitchensis*]	60	78	4,00E-101
20	contig00608 gene=isogroup00001 length=658	50 residues	Chloroplast	19409,2	9,13	14468,6	RuBisCO_small_like superfamily - PLNO2289 multi domain	CAA67061.1	Ribulose-bisphosphate carboxylase [*Pteris vittata*]	83	98	2,00E-103
21	isotig06698 gene=isogroup02993 length=1065 numContigs=1			26771,8	5,5		TIM_phosphate_binding superfamily	ABC59698.1	Triosephsphate isomerase [*Pteris vittata*]	88	99	7,00E-159
22	isotig00170 gene=isogroup00009 length=927 numContigs=5	73 residues	Chloroplast	18034,4	4,94	11094,3	Cupredoxin superfamily	Q7SIB8.1	RecName: Full=Plastocyanin [*Dryopteris classirhizoma*]	93	57	6,00E-62
23	isotig07124 gene=isogroup03419 length=940 numContigs=1	39 residues	Secret. Pathway	24445	5,85	20593,4	ABD superfamily	BAD86497.1	Germin-like protein [*Physcomitrella patens*]	48	94	7,00E-62

24	isotig07154 gene=isogroup03449 length=932 numContigs=1	20 residues	Secret. Pathway	23630,9	4,78	21238,3	Chitin_bind_3 superfamily	AFR32946.1	Chitin binding protein [*Tectaria macrodonta*]	79	100	8,00E-121
25	contig00359 gene=isogroup00001 length=1130			27922,6	5,26		C2 superfamily	XP_008238904.1	PREDICTED: elicitor-responsive protein 3-like [*Prunus mume*]	42	99	6,00E-28
26	isotig02337 gene=isogroup00450 length=967 numContigs=2	26 residues	Secret. Pathway	28079,6	4,74	25220,1	HAD_like superfamily	EYU30686.1	Hypotetical protein MINGU_mgv1a011950 [*Erythranthe guttata*] [*Mimulus guttatus*]	51	94	4,00E-76
27	isotig06518 gene=isogroup02813 length=1136 numContigs=1	26 residues	Secret. Pathway	29451,3	4,56	4,41	ChtBD1 superfamily & lysozyme_like superfamily	AAP03088.1	Class la chitinase [*Galega orientalis*]	47	100	4,00E-95
28	isotig01107 gene=isogroup00129 length=784 numContigs=4			16569	5,41		Cyanase_C superfamily - PRK02866 multi domain	XP_001780115.1	Predicted protein [*Physcomitrella patens*]	63	92	1,00E-59
29	isotig00064 gene=isogroup00003 length=865 numContigs=6	6 residues	Mitochondrion	16912	5,21	16222,1	Biotinyl_lipoyl_domains superfamily	XP_001769387.1	Predicted protein [*Physcomitrella patens*]	49	98	5,00E-44
30	isotig07553 gene=isogroup03848 length=850 numContigs=1			17105,1	5,63		No putative conserved domains have been detected	XP_012065866.1	PREDICTED: uncharacterized protein LOC105628969 [*Jatropha curas*]	30	83	0,43
31	isotig07191 gene=isogroup03486 length=952 numContigs=1			21589,7	6,42		FMM_red superfamily	XP_008452342.1	PREDICTED: minor allergen Alt a 7 [*Cucumis melo*]	88	100	1,00E-125
32	isotig04005 gene=isogroup01211 length=1029 numContigs=2	73 residues	Chloroplast	21672,1	8,58	14447,6	YjgF_YER057c_UK114_family superfamily	XP_010923900.1	PREDICTED: reactive Intermediate Deaminase A, chloroplastic [*Elaeis guineensis*]	63	93	2,00E-72
33	isotig01336 gene=isogroup00175 length=1197 numContigs=5	24 resdues	Secret. Pathway	14958,9	9,07	12310,8	No putative conserved domains have been detected	WP_002614061.1	TenA family transcriptional regulator [*Stigmatella aurantiaca*]	28	68	0,16
34	isotig07515 gene=isogroup03810 length=848 numContigs=1	26 residues	Secret. Pathway	≥26553.8	n.d.	[3]23633.2	PAE superfamily	XP_007040900.1	Pectinacetylesterase family protein [*Theobroma cacao*]	54	84	4,00E-69
35	isotig01087 gene=isogroup00125 length=1302 numContigs=5	n.d.	n.d.	≥17731.3	n.d.		No putative conserved domains have been detected	XP_006850823.1	PREDICTED: UPF0603 protein At1g54780, chloroplastic [*Amborella trichopoda*]	76	69	1,00E-53
36	isotig08555 gene=isogroup04850 length=703 numContigs=1	37 residues	Chloroplast	≥20923.5	n.d.	[3]17060.2	Thioredoxin_like superfamily & GST_C_family superfamily - Gst multi domain	XP_003558194.1	REDICTED: protein IN2-1 homolog B-like isoform X1 [*Brachypodium distachyon*]	57	100	8,00E-70

Appendix Dataset 2. Proteomics *P. aquilinum*

Entry	Accession	Description	Σ# Proteins	Σ# Unique Peptides	Σ# Peptides	Σ# PSMs
1	isotig27020_2	gene=isogroup01931 length=1185 numContigs=4	2	13	13	58
2	contig07161_5	gene=isogroup00001 length=1004	5	5	5	31
3	isotig12606_1	gene=isogroup00580 length=905 numContigs=5	21	3	3	18
4	isotig00909_5	gene=isogroup00110 length=447 numContigs=3	45	3	3	10
5	isotig10580_2	gene=isogroup00499 length=1298 numContigs=11	8	3	3	10
6	isotig19508_4	gene=isogroup00976 length=1128 numContigs=5	4	3	3	6
7	isotig25581_1	gene=isogroup01657 length=894 numContigs=4	4	2	2	39
8	isotig35325_5	gene=isogroup05475 length=653 numContigs=1	1	2	2	2
9	isotig39620_4	gene=isogroup09770 length=380 numContigs=1	1	6	15	449
10	isotig41972_4	gene=isogroup12122 length=279 numContigs=1	1	5	14	258
11	isotig25270_1	gene=isogroup01604 length=530 numContigs=5	6	5	12	80
12	isotig39620_6	gene=isogroup09770 length=380 numContigs=1	1	6	11	76
13	isotig40347_1	gene=isogroup10497 length=362 numContigs=1	1	4	9	103
14	isotig45655_1	gene=isogroup15805 length=116 numContigs=1	1	2	9	64
15	isotig40644_3	gene=isogroup10794 length=349 numContigs=1	1	5	8	60
16	isotig32680_2	gene=isogroup03823 length=679 numContigs=2	2	6	7	30
17	isotig35485_1	gene=isogroup05635 length=617 numContigs=1	1	3	4	8
18	isotig40976_6	gene=isogroup11126 length=339 numContigs=1	1	4	4	6
19	isotig31074_3	gene=isogroup03034 length=479 numContigs=3	1	3	3	10
20	isotig36859_3	gene=isogroup07009 length=483 numContigs=1	1	3	6	9
21	isotig00921_4	gene=isogroup00111 length=700 numContigs=4	7	4	5	39
22	isotig26079_5	gene=isogroup01760 length=646 numContigs=5	1	3	4	17
23	isotig30152_4	gene=isogroup02722 length=825 numContigs=3	2	4	4	8
24	isotig00131_2	gene=isogroup00041 length=793 numContigs=3	1	3	4	7
25	isotig13501_5	gene=isogroup00624 length=436 numContigs=4	7	4	4	4
26	isotig15087_6	gene=isogroup00701 length=1724 numContigs=8	8	3	3	10
27	isotig42441_6	gene=isogroup12591 length=268 numContigs=1	1	2	2	7
28	isotig29719_3	gene=isogroup02580 length=174 numContigs=3	96	2	2	6
29	isotig22755_3	gene=isogroup01286 length=1170 numContigs=5	10	1	2	5
30	isotig38889_5	gene=isogroup09039 length=408 numContigs=1	1	2	2	2
31	isotig26707_5	gene=isogroup01870 length=820 numContigs=4	1	14	14	149
32	isotig23641_2	gene=isogroup01394 length=1294 numContigs=5	3	11	11	145
33	isotig14166_2	gene=isogroup00659 length=1227 numContigs=4	11	11	11	75
34	isotig25401_1	gene=isogroup01624 length=1031 numContigs=5	3	3	9	86
35	isotig28614_5	gene=isogroup02282 length=1328 numContigs=2	3	3	9	84
36	contig07365_1	gene=isogroup00001 length=676	1	9	9	50
37	isotig21454_2	gene=isogroup01160 length=1173 numContigs=7	1	5	8	48
38	isotig00033_4	gene=isogroup00021 length=551 numContigs=7	9	8	8	42

39	isotig05262_1	gene=isogroup00297 length=896 numContigs=12	22	8	8	30
40	isotig23342_6	gene=isogroup01348 length=1135 numContigs=6	4	8	8	27
41	isotig31435_4	gene=isogroup03199 length=1119 numContigs=2	2	7	7	37
42	isotig22528_5	gene=isogroup01268 length=1014 numContigs=6	4	7	7	29
43	isotig11587_2	gene=isogroup00537 length=982 numContigs=6	22	7	7	18
44	contig23522_4	gene=isogroup00133 length=557	1	6	6	56
45	isotig27663_6	gene=isogroup02043 length=1160 numContigs=4	2	6	6	15
46	isotig42980_5	gene=isogroup13130 length=243 numContigs=1	1	5	5	110
47	isotig02928_2	gene=isogroup00232 length=668 numContigs=6	20	3	5	72
48	isotig23335_3	gene=isogroup01347 length=1421 numContigs=4	4	5	5	25
49	isotig28912_3	gene=isogroup02378 length=1431 numContigs=4	2	5	5	25
50	isotig24917_5	gene=isogroup01559 length=680 numContigs=4	4	4	5	17
51	isotig19166_3	gene=isogroup00944 length=792 numContigs=6	6	4	4	15
52	isotig14105_5	gene=isogroup00652 length=882 numContigs=4	12	4	4	13
53	isotig26718_2	gene=isogroup01873 length=1047 numContigs=4	3	4	4	12
54	isotig32396_5	gene=isogroup03680 length=560 numContigs=2	2	4	4	10
55	isotig22508_4	gene=isogroup01264 length=660 numContigs=3	7	4	4	9
56	isotig05180_1	gene=isogroup00294 length=760 numContigs=4	6	4	4	9
57	isotig24397_4	gene=isogroup01493 length=1310 numContigs=4	5	3	3	8
58	contig11100_4	gene=isogroup00001 length=605	4	3	3	8
59	isotig09712_6	gene=isogroup00464 length=825 numContigs=8	24	3	3	6
60	isotig06562_1	gene=isogroup00341 length=730 numContigs=11	46	3	3	5
61	isotig02940_2	gene=isogroup00232 length=399 numContigs=4	10	1	3	3
62	isotig27612_2	gene=isogroup02030 length=104 numContigs=4	4	2	2	6
63	contig09843_2	gene=isogroup00001 length=745	1	3	3	8
64	isotig08035_5	gene=isogroup00404 length=609 numContigs=4	2	3	3	6
65	isotig37791_4	gene=isogroup07941 length=452 numContigs=1	1	3	3	6
66	isotig05390_3	gene=isogroup00303 length=680 numContigs=9	8	2	3	5
67	contig14211_4	gene=isogroup00001 length=576	1	3	3	4
68	isotig34852_4	gene=isogroup05002 length=933 numContigs=1	1	3	3	4
69	isotig07330_2	gene=isogroup00374 length=991 numContigs=5	4	3	3	3
70	isotig32794_6	gene=isogroup03880 length=755 numContigs=2	1	2	2	8
71	isotig05389_3	gene=isogroup00303 length=680 numContigs=9	8	1	2	7
72	isotig26059_1	gene=isogroup01756 length=797 numContigs=5	1	2	2	7
73	isotig20999_3	gene=isogroup01098 length=1205 numContigs=7	2	2	2	6
74	isotig34884_5	gene=isogroup05034 length=897 numContigs=1	1	2	2	6
75	isotig11954_5	gene=isogroup00561 length=678 numContigs=4	4	2	2	5
76	contig20123_2	gene=isogroup00061 length=560	1	2	2	5
77	isotig03436_4	gene=isogroup00246 length=425 numContigs=3	19	2	2	4
78	isotig13936_3	gene=isogroup00641 length=654 numContigs=6	5	2	2	4
79	isotig24915_5	gene=isogroup01559 length=677 numContigs=4	4	1	2	4

80	isotig28529_4	gene=isogroup02265 length=546 numContigs=4	3	2	2	4
81	isotig14043_2	gene=isogroup00647 length=1609 numContigs=10	2	2	2	4
82	isotig36821_1	gene=isogroup06971 length=485 numContigs=1	1	2	2	4
83	isotig17074_4	gene=isogroup00802 length=896 numContigs=3	16	2	2	3
84	isotig25423_6	gene=isogroup01627 length=582 numContigs=5	3	2	2	3
85	isotig32479_1	gene=isogroup03722 length=573 numContigs=2	2	2	2	3
86	isotig30794_1	gene=isogroup02938 length=547 numContigs=3	2	2	2	3
87	contig05122_4	gene=isogroup00001 length=577	1	2	2	3
88	isotig40701_4	gene=isogroup10851 length=347 numContigs=1	19	2	2	2
89	isotig09881_5	gene=isogroup00469 length=627 numContigs=9	12	2	2	2
90	isotig11882_4	gene=isogroup00554 length=514 numContigs=6	12	2	2	2
91	isotig12914_6	gene=isogroup00600 length=485 numContigs=6	12	2	2	2
92	isotig08972_6	gene=isogroup00437 length=420 numContigs=3	10	2	2	2
93	isotig26813_5	gene=isogroup01893 length=470 numContigs=4	10	2	2	2
94	isotig07507_3	gene=isogroup00391 length=1366 numContigs=8	10	2	2	2
95	isotig21255_4	gene=isogroup01129 length=786 numContigs=4	4	2	2	2
96	isotig43542_3	gene=isogroup13692 length=211 numContigs=1	3	2	2	2
97	isotig26534_5	gene=isogroup01829 length=1133 numContigs=4	2	2	2	2
98	isotig36514_2	gene=isogroup06664 length=495 numContigs=1	1	2	2	2
99	isotig35809_1	gene=isogroup05959 length=557 numContigs=1	1	2	2	2
100	isotig21419_5	gene=isogroup01155 length=907 numContigs=6	1	2	2	2
101	isotig35638_3	gene=isogroup05788 length=586 numContigs=1	1	2	2	2
102	contig02431_1	gene=isogroup00001 length=562	1	2	2	2
103	isotig35244_4	gene=isogroup05394 length=670 numContigs=1	1	2	2	2
104	isotig34991_6	gene=isogroup05141 length=795 numContigs=1	1	2	2	2
105	contig07034_4	gene=isogroup00001 length=707	1	2	2	2
106	isotig34757_5	gene=isogroup04907 length=1132 numContigs=1	1	2	2	2

Appendix Dataset 3. Pf 10604: List of proteins which existence has been proved at protein level

Entry	Accession	Protein names	Organism	Entry name	3D	Cross-reference (PDB)
1	Q8VZS8	Abscisic acid receptor PYL1 (ABI1-binding protein 6) (PYR1-like protein 1) (Regulatory components of ABA receptor 9)	*Arabidopsis thaliana* (Mouse-ear cress)	PYL1_ARATH	X-ray crystallography (7)	3JRQ;3JRS;3KAY;3KDJ;3NEF;3NEG;3NMN;
2	Q9FJ49	Abscisic acid receptor PYL12 (PYR1-like protein 12) (Regulatory components of ABA receptor 6)	*Arabidopsis thaliana* (Mouse-ear cress)	PYL12_ARATH		
3	O80992	Abscisic acid receptor PYL2 (PYR1-like protein 2) (Regulatory components of ABA receptor 14)	*Arabidopsis thaliana* (Mouse-ear cress)	PYL2_ARATH	X-ray crystallography (19)	3KAZ;3KB0;3KB3;3KDH;3KDI;3KL1;3NJ0;3NJ1 3NMH;3NMP;3NMT;3NMV;3NR4;3NS2;3UJL; 4LA7;4LG5;4LGA;4LGB;
4	Q9SSM7	Abscisic acid receptor PYL3 (PYR1-like protein 3) (Regulatory components of ABA receptor 13)	*Arabidopsis thaliana* (Mouse-ear cress)	PYL3_ARATH	X-ray crystallography (6)	3KLX;3OJI;4DS8;4DSB;4DSC;4JDA;
5	Q8H1R0	Abscisic acid receptor PYL10 (ABI1-binding protein 8) (PYR1-like protein 10) (Regulatory components of ABA receptor 4)	*Arabidopsis thaliana* (Mouse-ear cress)	PYL10_ARATH	X-ray crystallography (4)	3R6P;3RT0;3RT2;3UQH;
6	Q8S8E3	Abscisic acid receptor PYL6 (ABI1-binding protein 5) (PYR1-like protein 6) (Regulatory components of ABA receptor 9)	*Arabidopsis thaliana* (Mouse-ear cress)	PYL6_ARATH		
7	Q9FLB1	Abscisic acid receptor PYL5 (ABI1-binding protein 3) (PYR1-like protein 5) (Regulatory components of ABA receptor 8)	*Arabidopsis thaliana* (Mouse-ear cress)	PYL5_ARATH	X-ray crystallography (1)	4JDL;
8	Q9FGM1	Abscisic acid receptor PYL8 (ABI1-binding protein 1) (PYR1-like protein 8) (Regulatory components of ABA receptor 3)	*Arabidopsis thaliana* (Mouse-ear cress)	PYL8_ARATH		
9	O80920	Abscisic acid receptor PYL4 (ABI1-binding protein 2) (PYR1-like protein 4) (Regulatory components of ABA receptor 10)	*Arabidopsis thaliana* (Mouse-ear cress)	PYL4_ARATH		
10	Q1ECF1	Abscisic acid receptor PYL7 (ABI1-binding protein 7) (PYR1-like protein 7) (Regulatory components of ABA receptor 2)	*Arabidopsis thaliana* (Mouse-ear cress)	PYL7_ARATH		
11	Q84MC7	Abscisic acid receptor PYL9 (ABI1-binding protein 4) (PYR1-like protein 9) (Regulatory components of ABA receptor 1)	*Arabidopsis thaliana* (Mouse-ear cress)	PYL9_ARATH	X-ray crystallography (2)	3OQU;3W9R;
12	P9WM73	Uncharacterized protein Rv0088	*Mycobacterium tuberculosis* (strain ATCC 25618 / H37Rv)	Y088_MYCTU		

13	P9WLU7	Uncharacterized protein Rv1546	*Mycobacterium tuberculosis* (strain ATCC 25618 / H37Rv)	Y1546_MYCTU		
14	P9WL87	Uncharacterized protein Rv2574	*Mycobacterium tuberculosis* (strain ATCC 25618 / H37Rv)	Y2574_MYCTU		
15	P9WJ05	Toxin Rv0910	*Mycobacterium tuberculosis* (strain ATCC 25618 / H37Rv)	Y910_MYCTU		
16	P59082	Lachrymatory-factor synthase	*Allium cepa* (Onion)	LFS_ALLCE		
17	Q5Z8S0	Bet v I allergen-like (Os06g0562200 protein)	*Oryza sativa subsp. japonica* (Rice)	Q5Z8S0_ORYSJ	X-ray crystallography (1)	4OIC;
18	I6XI16	Mycobacterium tuberculosis H37Rv complete genome (Polyketide cyclase)	*Mycobacterium tuberculosis* (strain ATCC 25618 / H37Rv)	I6XI16_MYCTU		
19	O53961	Mycobacterium tuberculosis H37Rv complete genome (Polyketide cyclase)	*Mycobacterium tuberculosis* (strain ATCC 25618 / H37Rv)	O53961_MYCTU		
20	I6Y1P2	Mycobacterium tuberculosis H37Rv complete genome (Polyketide cyclase)	*Mycobacterium tuberculosis* (strain ATCC 25618 / H37Rv)	I6Y1P2_MYCTU		
21	O53408	Mycobacterium tuberculosis H37Rv complete genome (Uncharacterized protein)	*Mycobacterium tuberculosis* (strain ATCC 25618 / H37Rv)	O53408_MYCTU		
22	I6X666	Cyclase (Mycobacterium tuberculosis H37Rv complete genome)	*Mycobacterium tuberculosis* (strain ATCC 25618 / H37Rv)	I6X666_MYCTU		
23	O06816	Mycobacterium tuberculosis H37Rv complete genome	*Mycobacterium tuberculosis* (strain ATCC 25618 / H37Rv)	O06816_MYCTU		
24	Q7CZ16	Uncharacterized protein	*Agrobacterium tumefaciens* (strain C58 / ATCC 33970)	Q7CZ16_AGRT5	X-ray crystallography (1)	2QPV;
25	Q6HG14	Uncharacterized protein	*Bacillus thuringiensis subsp. konkukian* (strain 97-27)	Q6HG14_BACHK	X-ray crystallography (1)	3F08;
26	Q47KK8	Uncharacterized protein	*Thermobifida fusca* (strain YX)	Q47KK8_THEFY	NMR spectroscopy (1)	2LE1;
27	Q81AY6	XoxI	*Bacillus cereus* (strain ATCC 14579 / DSM 31 / JCM 2152 / NBRC 15305 / NCIMB 9373 / NRRL B-3711)	Q81AY6_BACCR	X-ray crystallography (1)	3CNW;
28	Q8NN40	Putative uncharacterized protein Cgl2373	*Corynebacterium glutamicum* (strain ATCC 13032 / DSM 20300 / JCM 1318 / LMG 3730 / NCIMB 10025)	Q8NN40_CORGL	NMR spectroscopy (1)	2M47;
29	Q2Y8N9	Uncharacterized protein	*Nitrosospira multiformis* (strain ATCC 25196 / NCIMB 11849)	Q2Y8N9_NITMU	X-ray crystallography (1)	3P51;
30	A1U5H9	Uncharacterized protein	*Marinobacter hydrocarbonoclasticus* (strain ATCC 700491 / DSM 11845 / VT8)	A1U5H9_MARHV	X-ray crystallography (1)	3P9V;
31	Q9F6D3	Cyclase	*Streptomyces* sp. R1128	Q9F6D3_9ACTO	X-ray crystallography (1)	3TFZ;

BENTHAM SCIENCE PUBLISHERS LICENSE TERMS AND CONDITIONS

Oct 13, 2015

This is a License Agreement between Elisa Lanfranchi ("You") and Bentham Science Publishers ("Bentham Science Publishers") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Bentham Science Publishers, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3726480744490
License date	Oct 12, 2015
Licensed content publisher	Bentham Science Publishers
Licensed content publication	Recent Patents on Biotechnology
Licensed content title	Mini-Review: Recent Developments in Hydroxynitrile Lyases for Industrial Biotechnology
Licensed copyright line	Bentham Science Publisher
Licensed content author	Elisa Lanfranchi, Kerstin Steiner, Anton Glieder et al
Licensed content date	December 2013
Type of Use	Thesis/Dissertation
Requestor type	Author of requested content
Format	Print, Electronic
Portion	chapter/article
Rights for	Main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	no
In the following language(s)	Original language of publication

With incidental promotional use	no
The lifetime unit quantity of new product	0 to 499
The requesting person/organization is:	Elisa Lanfranchi
Title of your thesis / dissertation	New Hydroxynitrile lyase from White Rabbit's foot fern: The route from discovery to perspectives for applications
Expected completion date	Oct 2015
Expected size (number of pages)	184
Billing Type	Invoice
Billing Address	Elisa Lanfranchi c/c Acib GmbH Petersgasse 14 Graz, Austria 8010 Attn: Elisa Lanfranchi
Total	0.00 EUR
Terms and Conditions	

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL

Introduction

The publisher for this copyrighted material is Bentham Science Publishers Ltd. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your CCC account and that are available at any time at http://myaccount.copyright.com).

Limited License

Publisher hereby grants to you a non-exclusive license to use this material. Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process; any form of republication must be completed within 180 days from the date hereof (although copies prepared before then may be distributed thereafter); and any electronic posting is limited to a period of

180 days.

Geographic Rights: Scope

Licenses may be exercised anywhere in the world.

Altering/Modifying Material: Not Permitted

You may not alter or modify the material in any manner, nor may you translate the material into another language without permission from the Publisher.

Reservation of Rights

Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

License Contingent on Payment

While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

Copyright Notice: Disclaimer

It is a condition of this license that on all copies you republish, display or distribute you (i) provide full bibliographical information for the licensed material, including but not limited to author name, title and publication date of the work, volume and number if applicable, publisher name (Eureka Science Ltd.), page reference, (ii) include the statement, "Reprinted by permission of Eureka Science Ltd."; and (iii) include the copyright notice that appears in the work as published by Eureka Science Ltd..

Warranties: None
Publisher makes no representations or warranties with respect to the licensed material.
Indemnity
You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
No Transfer of License
This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
No Amendment Except in Writing
This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
Objection to Contrary Terms
Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
Jurisdiction: Not Required*
This license transaction shall be governed by and construed in accordance with the laws of the United Arab Emirates. You hereby agree to submit to the jurisdiction of the federal and state courts located in Dubai for purposes of resolving any disputes that may arise in connection with this licensing transaction.
Other Terms and Conditions:
v1.0

Curriculum Vitae

Elisa Lanfranchi
Hans Brandstettergasse 9/9, 8010 Graz Austria
e.lanfranchi@hotmail.t - AT: +43-(0)6605120029 – IT: +39-3491272156

Birth: Lecco (Italy), 25th December 1985
Citizenship: Italian

Education/academic qualification

2012 – 2015
Doctoral School of Molecular Biosciences and Biotechnology
Institute of Molecular Biotechnology
Graz University of Technology, Nawi Graz, Graz Austria
Supervisor: Prof. Anton (Toni) Glieder

2008 – 2011
Master Degree with distinction in Industrial Biotechnology
University Milano – Bicocca, Milan Italy
Supervisor: Prof. Giuseppina Bestetti
Thesis: "Purification and characterization of *trans-o*-hydroxybenzylidenpyruvate hydratase-aldolase enzyme of *Pseudomonas fluorescens* N3 and its application in aldolic condensation reactions."

2004 – 2008
Bachelor Degree in Biotechnology
University Milano – Bicocca, Milan Italy

Work Experience

2012 – 2015
Junior Researcher (PhD Student)
Austrian Centre Industrial Biotechnology (ACIB GmbH), Graz Austria
Main Research Field: Identification of new hydroxynitrile lyases and transaminases from non-conventional plants transcriptomes. Enzymes discovery, proteins expression in *P. pastoris* and *E. coli*, purification and characterization.

2011 – 2012
Project Student
Institute of Molecular Biotechnology, Graz University of Technology, Graz Austria
Research Field: Protein expression in *Pichia Pastoris.*

2010 – 2011
Master thesis
Organic chemistry department, University of Milan, Milan Italy
Research field: enzymatic characterization and biocatalysis.

Projects experience

Kyrobio: Developing biocatalysts for industrial chiral chemicals. European Union's Seventh Framework Programme for research. www.kyrobio.eu

Publications

Unique and Novel Enzymes Catalyze Cyanide Release from Fern.
Elisa Lanfranchi, Tea Pavkov-Keller, Margit Winkler, Eva-Maria Köhler, Matthias Diepold, Kerstin Steiner, Barbara Darnhofer, Juergen Hartler, Tom Van Der Berg, Henk-Jan Joosten, Mandana Gruber-Khadjawi, Gerhard G. Thallinger, Ruth Birner-Gruenberger, Karl Gruber, Anton Glieder. Manuscript in preparation

Production and immobilization of hydroxynitrile lyase from white bunny's foot fern (*Davallia tyermannii*) for a robust biocatalyst generation.
Elisa Lanfranchi, Sander Van Pelt, Roger A. Sheldon, Margit Winkler, Anton Glieder. Manuscript in preparation.

Bioprospecting for Hydroxynitrile Lyases by Blue Native PAGE Coupled HCN Detection.
Elisa Lanfranchi, Eva M. Köhler, Barbara Darnhofer, Kerstin Steiner, Ruth Birner-Gruenberger, Anton Glieder, Margit Winkler. Current Biotechnology. 4 (2015) 2, 111 - 117.

Mini-review: recent developments in hydroxynitrile lyases for industrial biotechnology.
Elisa Lanfranchi, Kerstin Steiner, Anton Glieder, Ivan Hajnal, Roger A. Sheldon, Sander Van Pelt, Margit Winkler. Recent Patents on Biotechnology. 7 (2013) 3, 197 - 206.

Patents

New Hydroxynitrile lyases. Elisa Lanfranchi, Anton Glieder, Margit Winkler, Kerstin Steiner, Tea Pavkov-Keller, Matthias Diepold. CT/EP2015/068262

Posters and presentations

2015

Biotrans 2015, Vienna, AT. Poster

DocDay, NAWI Graz Doctoral School Molecular Biology and Biochemistry & Molecular Biosciences and Biotechnology, Graz, AT. Talk

ACIB science Days, AT. Talk

Master Class Computational Approaches for Discovery and Engineering of Enzymes for Biocatalysis and Synthetic Biology, Groningen, NL. Talk

2014

Italian Forum of Industrial Biotechnology (IFIB), Genoa, IT. Poster

7th International Conference on Biocatalysis (BIOCAT), Hamburg, DE. Poster

DocDay, NAWI Graz Doctoral School Molecular Biology and Biochemistry & Molecular Biosciences and Biotechnology, Graz, AT. Poster

2013

The next generation of biocatalysis for industrial chemical synthesis & industrial biotechnology for Europe, Brussels, BE. Poster

Biotrans 2013, Manchester Central, UK. Poster

DocDay, NAWI Graz Doctoral School Molecular Biology and Biochemistry & Molecular Biosciences and Biotechnology, Graz, AT. Poster

Grants and Awards

September 2015

Forschung Stipendium 2015/2016 from Bundesministerium für Wissenschaft, Forschung und Wirtschaft (BMWFW), Austria. (Declined).

July 2015

Best speaker award: 14th Doc Day organized by Nawi Graz Doctoral School of Molecular Biosciences and Biotechnology

References

Anton Glieder

University of technology, Graz Austria.

Phone: +43 (316) 873 - 4074; 9300 a.glieder@tugraz.at

Margit Winkler

Austrian Centre Industrial Biotechnology, Graz Austria.

Phone: +43 316 873 9333 margit.winkler@acib.at

Languages

Italian, English, German

Personal Interests

Hiking and spend time in the wild nature, travel, learn culinary arts from different cultures (although Italian cuisine remains my first love).