R-hydroxynitrile lyase from the cyanogenic millipede, Chamberlinius hualienensis—A new entry to the carrier protein family Lipocalines

Fumihiro Motojima[1,2], Atsushi Izumi[1,2], Aem Nuylert[1], Zhenyu Zhai[1,2,*], Mohammad Dadashipour[1,2,†], Sayaka Shichida[1,2], Takuya Yamaguchi[1,2,‡], Shogo Nakano[1,2,¶] and Yasuhisa Asano[1,2]

1 Biotechnology Research Center and Department of Biotechnology, Toyama Prefectural University, Imizu, Toyama, Japan
2 Asano Active Enzyme Molecule Project, ERATO, JST, Imizu, Toyama, 939-0398, Japan

Correspondence
Y. Asano, Biotechnology Research Center and Department of Biotechnology, Toyama Prefectural University, 5180 Kurokawa, Imizu, Toyama 939-0398 JAPAN
Tel. +81 766 56 7500
E-mail: asano@pu-toyama.ac.jp

Present address
*Center for Experimental Medicine, The First Affiliated Hospital of Nanchang University, Nanchang,Jiangxi, 330006, China
†Faculty of Medicine, Health and Life Sciences, School of Biological Sciences, Queen's University Belfast, 19 Chlorine Gardens, Malone Road, Belfast, NI, BT9 5DL, UK
‡Faculty of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8572, Japan
¶Graduate Division of Nutritional and Environmental Sciences, University of Shizuoka, 52-1 Yada, Suruga-ku, Shizuoka, 422-8526, Japan

Hydroxynitrile lyases (HNLs) catalyze the cleavage of cyanohydrin into cyanide and the corresponding aldehyde or ketone. Moreover, they catalyze the synthesis of cyanohydrin in the reverse reaction, utilized in industry for preparation of enantiomeric pure pharmaceutical ingredients and fine chemicals. We discovered a new HNL from the cyanogenic millipede, Chamberlinius hualienensis. The enzyme displays several features including a new primary structure, high stability, and the highest specific activity in (R)-mandelonitrile ((R)-MAN) synthesis (7420 U·mg^{-1}) among the reported HNLs. In this study, we elucidated the crystal structure and reaction mechanism of natural ChuaHNL in ligand-free form and its complexes with acetate, cyanide ion, and inhibitors (thiocyanate or iodoacetate) at 1.6, 1.5, 2.1, 1.55, and 1.55 Å resolutions, respectively. The structure of ChuaHNL revealed that it belongs to the lipocalin superfamily, despite low amino acid sequence identity. The docking model of (R)-MAN with ChuaHNL suggested that the hydroxyl group forms hydrogen bonds with R38 and K117, and the nitrile group forms hydrogen bonds with R38 and Y103. The mutational analysis showed the importance of these residues in the enzymatic reaction. From these results, we propose that K117 acts as a base to abstract a proton from the hydroxyl group of cyanohydrins and R38 acts as an acid to donate a proton to the cyanide ion during the cleavage reaction of cyanohydrins. The reverse mechanism would occur during the cyanohydrin synthesis. (Photo: Dr. Yuko Ishida)

Databases
Structural data are available in PDB database under the accession numbers 6JHC, 6KFA, 6KFB, 6KFC, and 6KFD.

Abbreviations
HNL, hydroxynitrile lyase; MAN, mandelonitrile; NAG, N-acetylglucosamine; PDI, protein disulfide isomerase; SAD, single anomalous dispersion.

Introduction

Hydroxynitrile lyases (HNLs) have been known to occur mostly in plants [1,2], and recently, those from bacteria [3,4] and arthropod (cyanogenic millipedes) have been newly added to the group [5,6]. HNLs of plant origin have been studied well and their new distribution, characterization, and structure determination are reported. We and others discovered HNL activities by screening natural plants [2,7,8]. They include *S*-stereoselective HNL in the leaves of *Baliospermum montanum* and *R*-HNLs from *Chaenomles sinensis*, *Sorbus aucuparia*, *Prunus persica*, *Eriobotrya japonica* and recently from fern *Davallia tyermannii* [9], etc. Metal-dependent HNLs were newly discovered and characterized from bacteria [3,4]. The plant and millipede enzymes catalyze the last step of cyanogenesis in nature in which a cyanohydrin (alpha-hydroxynitrile) is cleaved into HCN and the corresponding aldehyde or ketone. This process is considered to function as one of their defense mechanism against herbivores and microbes, respectively [10]. On the other hand, HNLs are attractive biocatalysts for the synthesis of optically active cyanohydrins as valuable building blocks for the synthesis of pharmaceuticals and agrochemicals, since they catalyze asymmetric syntheses with either of *R*- or *S*-stereoselectivities by an addition of one carbon and the formed chiral cyanohydrin are versatile compound modified into varieties of synthetic intermediates by several chemical procedures [11].

Wöhler and Liebig studied 'emulsin', hydrolyzing amygdalin to HCN, suggesting an occurrence of possible β-glucosidase and HNL in *Prunus dulcis* (syn., *P. amygdalus*: *Pa*HNL) in 1832 [12]. HNLs have been purified and characterized from several plants and bacteria, by recent discoveries which have been made since the interesting aspects of biological cyanogenesis [13] and the potential of HNL as industrial catalysts had been recognized by several researchers. From the structural point of view, the class of the enzymes is now known to have several tertiary structures, probably because relatively simple single base mechanisms would allow the ancestral proteins to evolve to HNLs. The crystal structures of HNLs in Protein Data Bank (PDB) exhibit similarity to those

belonging to seven different superfamilies: glucose-methanol-choline (GMC) oxidoreductase [14], α/β-hydrolases [15-17], serine carboxypeptidase [18], cupin [3], Zn^{2+}-containing alcohol dehydrogenase [19], Bet v1 [9], and dimeric α + β barrel folds [20]. The structure of the well-studied HNL from *Prunus dulcis* is classified into GMC oxidoreductase fold [14]. *R-Pa*HNL contains flavin adenine dinucleotide (FAD), which was shown to be not directly involved in the enzyme catalysis [21], but having a stabilizing role for the structure of enzyme. The structure of HNLs from *Manihot esculenta* (*S-Me*HNL) and *Hevea brasiliensis* (*S-Hb*HNL) bears similarity to those belonging to α/β-hydrolase family. Although the α/β-hydrolase type esterases form an oxyanion hole during the catalysis, α/β-hydrolase type HNLs do not have such structures [22]. The HNL from *Sorghum bicolor* (*S-Sb*HNL) belongs to the serine carboxypeptidase fold with a conserved peptidase motif, although it does not exhibit carboxypeptidase activity for potential substrates [17]. The structure of HNL from *Granulicella tundricola* (*Gt*HNL), reported as the first crystallized bacterial HNL, belongs to the cupin fold bearing a small barrel structure as a metal-dependent enzyme [3]. The structure of the newly found HNL from fern, *Dt*HNL, is reported to be a Bet v1-like fold, which is composed of antiparallel β-sheet wrapped by α-helices [9]. We showed that the HNL from *Passiflora edulis* (*Pe*HNL) is the first HNL belonging to dimeric α + β barrel (DABB) superfamily, and it is the shortest HNL ever known [20]. These HNLs are considered to have evolved from different ancestral proteins as expanding example of convergent evolution and using different catalytic residues for the cleavage and synthesis of cyanohydrins [2]. There are still more HNLs reported with unknown tertiary structures, such as the one similar to oxidoreductase containing Zn^{2+} [19], the one with a new primary structure [23], etc. In 2015, our group discovered an HNL from the invasive millipede, *Chamberlinius hualienensis* (ChuaHNL). ChuaHNL showed the highest specific activity during (*R*)-MAN synthesis among the reported HNLs [5]. Since ChuaHNL has no similarity to other HNLs or any sequences in the BLAST

database, its structure and the catalytic mechanism have not been predicted.

In this study, we successfully determined the crystal structure of ChuaHNL purified as a dimer form from the millipedes in ligand-free form, its complexes with acetate, cyanide ion, and inhibitors (thiocyanate or iodoacetate) determined at 1.6, 1.5, 2.1, 1.55 Å resolution, respectively. A structure similarity search indicated that the overall structure of ChuaHNL is similar to lipocalin fold proteins even with low-sequence identities. The docking simulation of (*R*)-MAN and the kinetic parameters of the mutational study using recombinant ChuaHNL expressed in *Pichia pastoris* suggested that K117 abstracts a proton from the hydroxyl group and R38 donates a proton to the released cyanide ion during the cleavage reaction of cyanohydrins. Therefore, ChuaHNL follows the acid/base catalysis mechanism of action during the cleavage and synthesis reactions. This research reports the first crystal structure of an HNL other than those from plants and bacteria, from a cyanogenic millipede. We present various forms of the crystal structures of the enzyme possessing both new primary and tertiary structures. 3D structural determination of this biocatalyst will assist improving its function for any potential application as well as reveals a wider range of protein scaffolds with HNL activity in the nature as a great example of convergent evolution.

Results

Overall structure of ChuaHNL

The crystal structure of the natural ChuaHNL, purified from the millipedes, was determined by SAD method using an iodine-soaked ChuaHNL crystal diffracted to 2.1 Å resolution at an in-house X-ray source (Table 1). The crystal belonged to $P6_122$, and there was one molecule in the asymmetric unit (Table 1 and Fig. 1A). The amino acid residues were numbered from the N-terminal Leu residue of the mature ChuaHNL, whose signal peptide was removed. The electron density maps of Leu1 at the N-terminal end and of the Tyr162 at the C-terminal end of mature ChuaHNL were clearly visible. ChuaHNL contains two α-helices, three 3_{10}-helices, and eight β-strands (Fig. 1 and 2B). The eight β-strands form an antiparallel β-barrel comprising the central active site cavity (Fig. 1B).

The homodimer structure was generated by the crystallographic symmetry (Fig 1B). The buried dimer interface area was 1,717 $Å^2$ corresponding to 20 % of the surface area in the monomer structure. In the dimer interfaces, hydrogen bonds formed mainly among loops connecting β1-β2, β2-β3, β3-β4, β4-β5, and β5-β6, and hydrophobic

interactions formed among residues in β2-β4 (Fig 1C). There were two salt bridges between the carboxyl group of the C-terminal end (Y162) in one molecule and H81 in another molecule (Fig. 1C). There were three intramolecular disulfide bonds between C3-C108, C35-C152, and C104-C118 and two intermolecular disulfide bonds between C80 of one molecule and C158 of another (Fig 1B and 1C). The root-mean-square deviation (RMSD) among whole structure was 0.13 Å at 162 Cα atoms. Therefore, there was no significant change in the structure of ChuaHNL upon ligand binding.

Although no homologous sequence with ChuaHNL has been found in BLAST search [5], the structure comparison search using DALI server [24] revealed that the structure of ChuaHNL resembles those of lipocalin family proteins. Many typical lipocalins gave *Z*-scores higher than 10, indicating significant structural similarity. The *Z*-scores is a composite of several evaluations including RMSD of three-dimensional alignment and numbers of residues matched. A *Z*-score of more than 10 is typically given to a match of very strong similarities [25]. The secondary structure-based multiple sequence alignment between ChuaHNL and typical lipocalins is shown in Fig. 2A. Overall amino acid sequence identity of ChuaHNL to that of lipocalins was less than 8%. Lipocalins generally contain three structurally and sequentially conserved motifs known as structurally conserved regions (SCRs) 1 to 3 [26]. Even though the amino acid similarity among the SCRs was very low (Fig. 2A), the secondary structure of ChuaHNL overlapped well with that of human retinol binding protein 4, RBP4 (Fig. 2B).

Previous research has indicated that the natural ChuaHNL is glycosylated [5]. In the thiocyanate-bound form, *N*-acetylglucosamine (NAG) moiety with an (α1, 6)-fucose was observed at N123 (Fig. S1A). In other structures of ChuaHNL, one well-ordered NAG moiety linking N109 and N123 was observed (Fig. S1B and S1C). The recombinant ChuaHNL has been also confirmed that it was glycosylated at the same glycosylation sites (at position N109 and N123) in ChuaHNLs produced in *Pichia pastoris* expression system [27]. Although several Ser and Thr residues in the ChuaHNL structure are exposed to the solvent and there is only one predicted *O*-glycosylation site [5], the electron density of *O*-linked glycosylation site was not observed.

Ligands bound in the active site of ChuaHNL

Acetate-bound form

The acetate-bound form of ChuaHNL structure was determined at 1.5 Å resolution. The acetate contained

Table 1. Statistics for data collection and refinement

Data set	Acetate	Iodide-SAD	Ligand free	Cyanide ion	Thiocyanate	Iodoacetate
PDB code	***		***	***	***	***
Data Collection						
Beam line	PF BL1A	R-AXIS VII	PF BL5A	PF BL5A	PF BL1A	PF BL1A
Space group	$P6_122$	$P6_122$	$P6_122$	$P6_122$	$P6_122$	$P6_122$
Cell dimensions						
a = b, c (Å)	58.25, 224.98	58.42, 226.41	58.38, 226.09	58.59, 226.23	58.10, 225.96	58.09, 225.46
Wavelength (Å)	1.1	1.54	1.0	1.0	1.1	1.1
Resolution (Å)	50.44-1.50 (1.58-1.50)*	46.19-2.10 (2.21-2.10)*	50.56-1.60 (1.69-1.60)*	50.74-2.10 (2.21-2.10)*	50.31-1.55 (1.61-1.55)*	50.31-1.55 (1.63-1.55)*
R_{merge}	0.046 (0.501)*	0.117 (0.657)*	0.039 (0.229)*	0.116 (0.598)*	0.049 (0.525)*	0.066 (0.624)*
I/ σ(I)	26.5 (4.1)*	27.3 (5.0)*	36.1 (9.5)*	18.8 (5.3)*	34.7 (6.1)*	25.1 (4.2)*
Completeness (%)	100.0 (100.0)*	100.0 (100.0)*	99.9 (100.0)*	100.0 (100.0)*	99.9 (99.6)*	100 (100)*
Redundancy	10.4 (10.8)*	19.1 (15.8)*	11.3 (11.6)*	19.5 (20.5)*	18.7 (16.9)*	17.9 (16.6)*
Phasing						
Anomalous completeness (%)		100 (100)*				
Anomalous multiplicity		10.8 (8.5)*				
FOM after SHELX		0.63				
Refinement						
No. reflections	35572		29697	13571	32225	32191
R_{work}/R_{free}	0.174/0.187		0.159/0.185	0.191/0.233	0.171/0.191	0.165/0.179
Average B-factor/Number of atoms						
Protein	16.8/1303		18.5/1314	31.5/1287	18.1/1305	17.6/1310
Ligands	35.6/32		41.4/28	84.6/38	39.0/30	43.0/33
Waters	33.2/305		34.7/297	39.6/141	33.3/227	31.8/231
R.m.s. deviations						
Bond lengths (Å)	0.007		0.007	0.014	0.007	0.007
Bond angles (°)	1.25		1.30	1.67	1.30	1.28

*Values in parentheses are for the highest-resolution shell. R.m.s. deviations: root-mean-square deviations.

in the reservoir solution was observed in the active site. The carboxylic group's oxygen forms salt bridges with R38 and Y40 (Fig. 3A).

Ligand-free form

The ligand-free form of ChuaHNL structure was determined at 1.6 Å resolution. To remove the acetate bound in the active site, the crystal was soaked in the solution containing bis-tris propane-citric acid buffer as described in the Materials and Methods. Three water molecules were observed in the active site (Fig. 3B). They formed hydrogen bonds with R38, Y103, and K117.

Cyanide ion-bound form

The crystal structure of ChuaHNL complexed with a cyanide ion was determined at 2.1 Å resolution (Fig. 3C). The orientation of the cyanide ion was determined by the temperature factor and a possible electrostatic interaction between the negatively charged carbon and R38. The nitrogen atom of the cyanide ion formed a hydrogen bond with Y40-Oη at the distance of 3.3 Å. The negatively charged carbon of cyanide ion electrostatically interacts with R38-Nη1 at the distance of 4.2 Å.

Inhibitor-bound form

Our group had previously reported that iodoacetate and thiocyanate inhibited the (R)-MAN synthesis catalyzed by ChuaHNL [5]. We determined the structures of the complexes formed with iodoacetate and thiocyanate at 1.55 Å resolution each. In the complex structure with iodoacetate, the oxygen of the carboxylic acid moiety formed salt bridges with R38-N1η and -N2η with bond lengths of 3.1 and 3.0 Å, respectively, and with K117-Nε with bond length of 3.1 Å (Fig. 3D). The iodine atom of iodoacetate may form a weak π-interaction with F67 with bond length of 4.3 Å. A strong anomalous difference
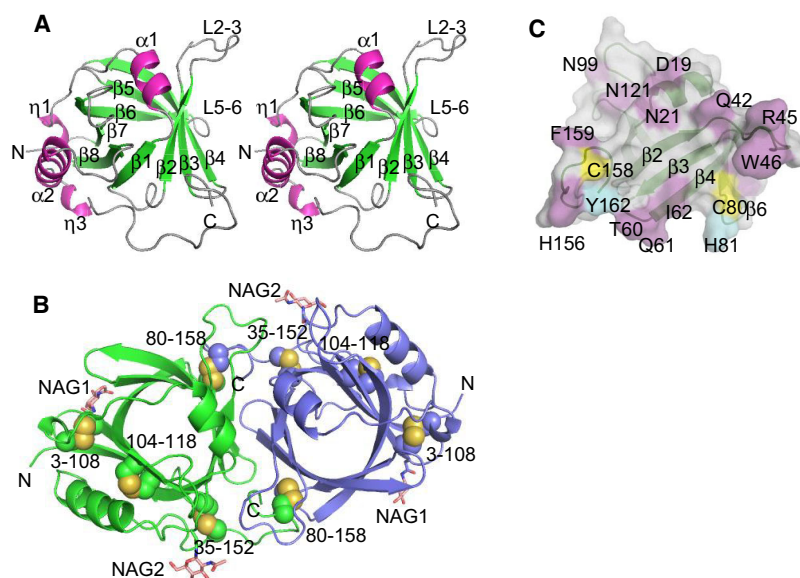
Fig. 1. Overall structure of ChuaHNL. (A) The stereo view of the monomeric structure of ChuaHNL. The helices (α1-2 for α-helices and η1-3 for 3_{10}-helices), the β-strands (β1-8), and the loop structures are colored in magenta, green, and gray, respectively. 'N' and 'C' symbols indicate the N- and C-terminal ends of ChuaHNL, respectively. (B) The dimer structure of ChuaHNL. One chain is colored in green and the other is colored in blue. Disulfide bonds are labeled with the residue number and are shown in a sphere representation. *N*-acetylglucosamine moieties connecting to N109 and N123 are labeled as NAG1 and NAG2, respectively. N109, N123, and sugar moieties are shown via a stick representation. (C) The dimer interface of ChuaHNL. The disulfide bonds, hydrogen bonds, and salt bridges in the interface are colored in yellow, magenta, and cyan, respectively. The interface residues and β-strands are labeled. The protein structures were designed using the program PyMOL [61]

map between F67 and R38 implies the presence of an iodine atom.

In the complex structure with thiocyanate, two alternate binding patterns were observed. The orientation of the two thiocyanate molecules in the crystal structure was determined by the lower *B*-factor values after the structure refinement. In one binding model (Fig. 3E, SCN1), the sulfur atom of thiocyanate formed a hydrogen bond each with Y103-Oη and R38-Nη1 with bond lengths of 3.0 and 3.2 Å, respectively. In the other binding model (Fig. 3E, SCN2), the sulfur atom of thiocyanate formed salt bridges with R38-Nη2 and K117-Nε both with bond length of 3.1 Å (Fig. 3E).

The docking simulation of (R)-MAN

To model how ChuaHNL catalyzes the synthesis and cleavage of cyanohydrins, we performed a docking simulation of (*R*)-MAN into the active site of ChuaHNL by MOE (Molecular Operating Environment version 2016.8, Montreal, Canada). The model giving the most prominent affinity score (*S* = −5.44) is shown in Fig. 4. The binding of (*R*)-MAN in the cavity of active site is shown in Fig. 4A. The hydroxyl

and nitrile group of (*R*)-MAN interact with hydrophilic residues of ChuaHNL at the bottom of the active site. The benzene group of (*R*)-MAN is surrounded by many hydrophobic and aromatic residues (such as I11, F17, F25, A54, F67, A75, L77, W88, F90, A105, and A119). The deep hydrophobic cavity might facilitate the acceptance of various bulky cyanohydrins as observed previously [5].

The interactions between hydrophilic residues and (*R*)-MAN are shown in Fig. 4B. The hydroxyl group of (*R*)-MAN forms hydrogen bonds with K117-Nε and R38-Nη1 both with bond length of 3.3 Å. The nitrile group forms hydrogen bonds with R38-Nη1, R38-Nη2, and Y103-Oη with bond lengths of 3.4, 3.3, and 3.3 Å, respectively. D56 forms salt bridges with R38-Nη2, R38-Nε, and K117-Nε. Y40-Oη forms hydrogen bonds with R38-Nη1 and Y103-Oη. The bond lengths and orientations of the residues involved in the ligand binding were almost identical in all five ChuaHNL structures.

Site-directed mutational analysis of ligand-binding residues and kinetic studies

To prove the catalytic roles of the hydrophilic residues that interact with (*R*)-MAN as indicated by the results
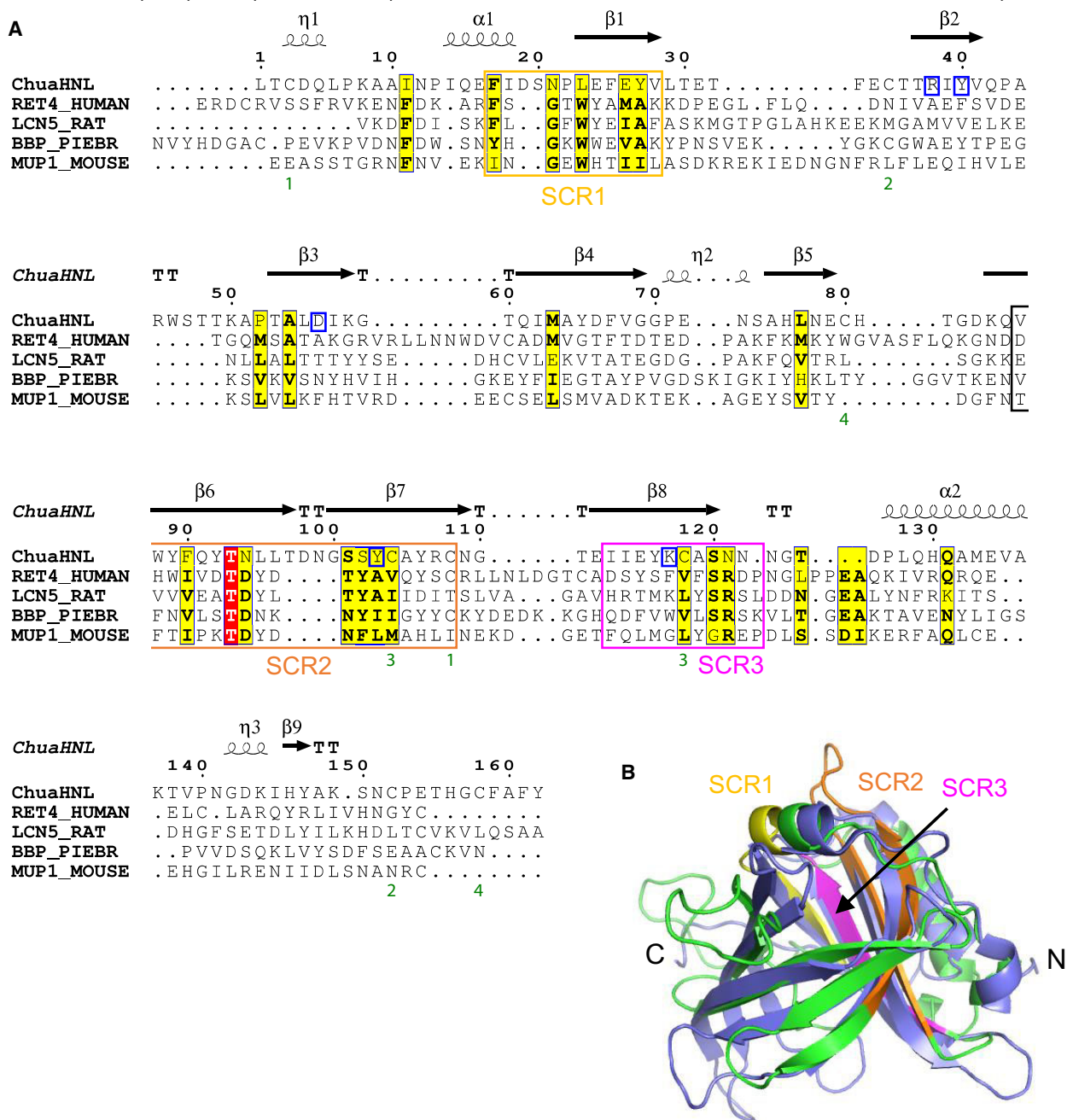
A

B

Fig. 2. Structural and sequential comparison of ChuaHNL with lipocalins. (A) The secondary structure-based multiple sequence alignment among ChuaHNL and typical lipocalins, such as human retinol binding protein 4 (RET4_HUMAN, PDB ID: 1JYD, DALI Z-score: 13.2) [62,63], rat epididymal-specific lipocalin 5 (LCN5_RAT, PDB ID: 1EPA, DALI Z-score: 12.5) [52], bilin binding protein (BBP_PIER, PDB ID: 1BBP, DALI Z-score: 12.3) [64], and mouse major urinary protein 1 (MUP1_MOUSE, PDB ID: 2DM5, DALI Z-score: 10.2) [65]. Numbers refer to the ChuaHNL sequence. The three major structurally conserved regions (SCRs 1-3) are indicated by rectangles. The secondary structure elements are shown as follows: α-helices, medium squiggles with α symbols; 3₁₀-helices, squiggles with η symbols; β-strands, arrows with β symbols; strict β-turns, TT letters. Identical residues are highlighted with red boxes and white texts. Similar groups are highlighted with yellow boxes. Similar amino acids in each group are shown as bold characters. The pairs of cysteine residues forming disulfide bonds are shown as green digits at the bottom of the alignment. The proposed catalytic residues are surrounded by blue boxes. The alignment was done using PROMALS3D [66] and illustrated using ESPript 3.0 (http://espript.ibcp.fr) [67]. (B) The superposed structures of monomer ChuaHNL (green) and human retinol binding protein 4 (PDB ID: 1JYD) are shown. The SCRs 1-3 of ChuaHNL are colored in yellow, orange, and magenta, respectively. The N- and C-terminal ends of ChuaHNL polypeptide chain are labeled as N and C, respectively. The alignment was done using PROMALS3D [67] and illustrated using ESPript 3.0 (http://espript.ibcp.fr) [67]. Superposition with secondary structure matching was evaluated using superpose program of CCP4 suites [54]
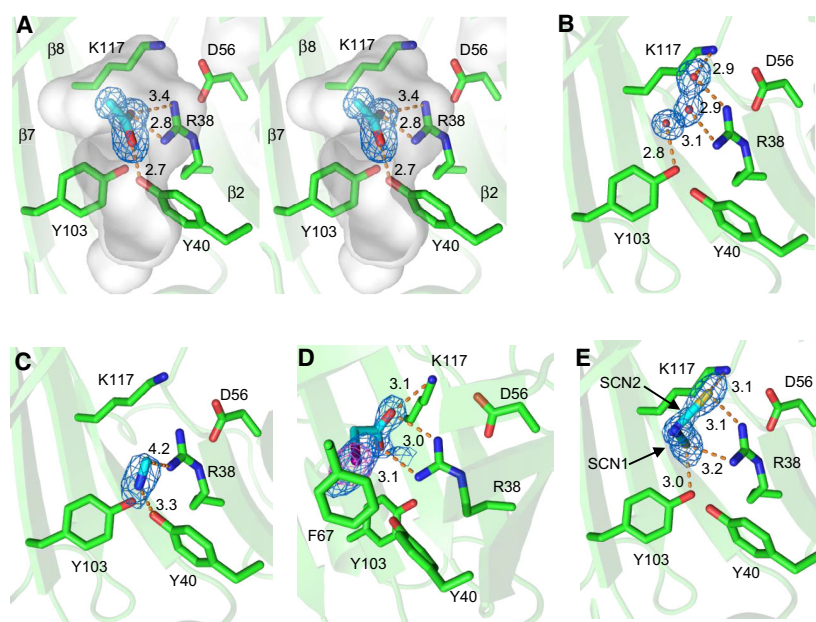
Fig. 3. The active sites of ChuaHNL complexed with various ligands. Amino acid residues in the active site and bound ligands are shown as a stick representation in CPK color scheme (carbon atoms of protein and ligands are shown in green and cyan, respectively). Hydrogen bonds and electrostatic interactions between amino acid residues and ligands are shown as dashed lines (orange) with bond length in Å. σ_A-weighted omit maps for ligands and water molecules are shown as a blue mesh representation and contoured at 3.0 σ level except D (5.0 σ). (A) The wall-eye stereo view of the active site bound to acetate. The surface of the active site cavity is shown as transparent gray surface. The entrance to the active site is located at the bottom of the images. (B) The ligand-free form. Water molecules bound to the active site are shown as a sphere representation. (C) The complex with cyanide ion. (D) The complex with iodoacetate. The anomalous difference map is shown as a violet mesh representation and is contoured at 4.0 σ level. (E) The complex with thiocyanate. Two alternate forms of the complex with thiocyanate are shown and labeled (SCN1 and SCN2). The protein structures were designed using the program PyMOL [61]

of the docking simulation, site-directed mutagenesis was performed using the recombinant His-ChuaHNL generated in *P. pastoris* by coexpression with protein disulfide bond isomerase (PpPDI) as described in our previous report [27]. The hydrophilic residues surrounding (*R*)-MAN in the docking model (R38, Y103, Y40, D56, and K117) were altered (Fig. 4A and 4B). The Ala variant of R38 (R38A) was successfully expressed and purified, but none of the other Ala variants were expressed in *P. pastoris* as confirmed by western blotting analysis (data not shown). Alternatively, analogous variants of His-ChuaHNL (Y40F, D56E, Y103F, and K117R) were similarly prepared and their kinetic parameters shown in the synthesis and cleavage toward benzaldehyde and (*R*)-MAN, respectively, were determined and compared with those of the wild-type enzyme.

The kinetic parameters of the cleavage and synthesis of (*R*)-MAN by His-ChuaHNL and its variants are shown in Table 2. The turnover number of His-ChuaHNL for (*R*)-MAN synthesis ($k_{cat} = 2110 \pm 180$ s^{-1}) was higher than that of the (*R*)-

MAN cleavage ($k_{cat} = 142 \pm 0.85$ s^{-1}) activity. All of the parameters obtained with mutations on the hydrophilic residues surrounding (*R*)-MAN in the docking model were decreased or impaired both in the synthesis and cleavage activities. The cleavage activity of the R38A variant was completely lost, indicating the critical importance of R38 which interacts with both the hydroxyl and the nitrile groups of (*R*)-MAN in the binding model. The V_{max}, k_{cat}, and k_{cat}/K_m of two variants, Y103F and K117R, were significantly decreased as compared with the wild-type, supporting the above model that Y103 and K117 directly interact with the nitrile and the hydroxyl groups, respectively, of (*R*)-MAN. Some unexpected results of the K_m values of the proposed mutations substitute with a bulky aromatic amino acid side chain, Y40F and Y103F, showed improved (*R*)-MAN binding affinities probably because of hydrophobic π-π interactions between F40 or F103, generating and the (*R*)-MAN (Fig. 4B). Another analogous variants, D56E and K117R, also resulted in preferential recognition of (*R*)-MAN, exhibiting the lower K_m value than that of the wild-
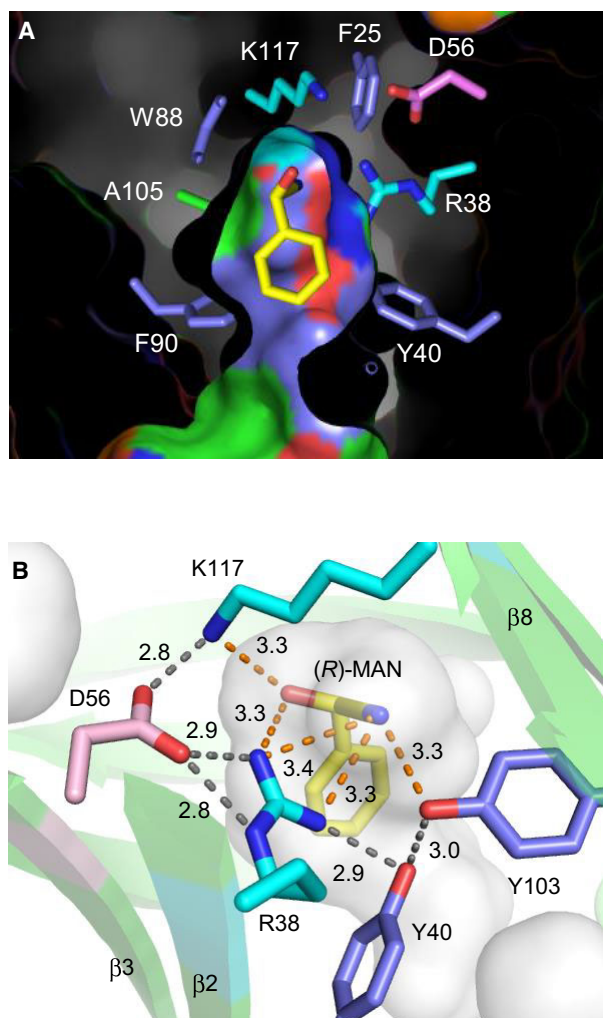
Fig. 4. The docking simulation and mutational analysis of ChuaHNL. (A) The surface representation of docked (*R*)-MAN in the active site of ChuaHNL. (*R*)-MAN and the residues in the active site are shown as a stick representation in CPK color scheme. The carbon atoms of (*R*)-MAN are colored in yellow. The carbon atoms of hydrophilic, negatively charged, positively charged, aliphatic, and aromatic residues of ChuaHNL are colored in orange, pink, cyan, green, and purple, respectively. (B) The interactions between (*R*)-MAN and the residues in the active site. Color scheme is the same as in A. The hydrogen bonds interacting with (*R*)-MAN are shown as orange dotted lines along with their lengths. The hydrogen bonds and salt bridges between the residues are shown as gray dotted lines. The docking simulation was done and the figures drawn by MOE (Molecular Operating Environment version 2016.8, Montreal, Canada)

type, possibly due to an improvement of the interaction in the bond length and/or the favorable orientation of (*R*)-MAN.

In the synthesis reaction, the R38A variant also showed no activity, indicating that R38 would be a

key residue that is involved in an acid/base catalysis in the ChuaHNL reaction mechanism. The lowest V_{max}, k_{cat}, and k_{cat}/K_m toward benzaldehyde was calculated to be 40 ± 0.81 μmol·min^{-1}·mg^{-1}, 14 ± 0.28 s^{-1} and 3.0 ± 0.63 mm^{-1}·s^{-1}, respectively, with Y103F; these values were 125-fold, 150-fold, and 143-fold lower than those of wild-type His-ChuaHNL, respectively. The K_m of Y103F (4.90 ± 1.1 mm) was almost identical to that of the wild-type (4.91 ± 0.66 mm), suggesting that the Y103 interacts with the cyanide ion, not with benzaldehyde. The V_{max} and k_{cat} of mimic function of K117R variant was 403 ± 45 μmol·min^{-1}·mg^{-1} and 141 ± 15 s^{-1} which exhibited the highest activities retaining synthesis activity among the variants, whereas the K_m value for benzaldehyde was eight times higher than that the wild-type. Consequently, the enzyme efficiency of this variant ($k_{cat}/K_m = 3.6 \pm 0.24$ mm^{-1}·s^{-1}) was dramatically decreased, similar to that of Y103F variant. It is obvious in this variant that K117 could have an effect on the accessibility and affinity toward benzaldehyde.

Another analogous alteration at Y40F and D56E, which do not look to interact with (*R*)-MAN directly, the turnover numbers for (*R*)-MAN cleavage were 18 ± 0.35 and 6 ± 0.3 s^{-1}, which retain 12.6% and 4.2% activity of the wild-type, respectively. These results suggest that the mutations do not disrupt the interaction with (*R*)-MAN and that D56 has a role in the cleavage activity, whereas Y40 has an effect on the synthesis activity, coordinating with benzaldehyde.

Discussion

We have proved in *C. hualienensis* that (*R*)-Man is biosynthesized from L-Phe, catalyzed by several P450 enzymes characterized by expression in *E. coli* of their cDNA cloned from mRNA [28]. We established the molecular mechanisms of the synthesis of defense allomones and shown that adult millipedes produce a mixture of benzaldehyde, benzyl alcohol, benzoylcyanide, mandelonitrile, and benzoic acid [29,30]. They employ various mechanisms to defend themselves using enzyme-mediated secretions such as HCN from (*R*)-MAN by HNL [5] and benzoyl cyanide and hydrogen peroxide from (*R*)-MAN using a new enzyme mandelonitrile oxidase [31]. Benzoyl cyanide further accepts one more molecule of (*R*)-MAN to form an ester (*R*)-MAN benzoate by a nonenzymatic Schotten–Bauman reaction releasing one molecule of cyanide [32].

Since these arthropods comprise highly new and potential enzymes, we are interested in precisely studying them from isolation and molecular characterization of these valuable biocatalysts to the structural

Table 2. Kinetic parameters of wild-type and mutants His-ChuaHNL

Enzyme[*,a]	Cleavage reaction[b] (*R*-MAN as substrate)				Synthesis reaction[c] (Benzaldehyde as substrate)			
	V_{max} (μmol min^{-1} mg^{-1})	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (mM^{-1} s^{-1})	V_{max} (μmol min^{-1} mg^{-1})	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (mM^{-1} s^{-1})
Wild-type	406 ± 2.11	142 ± 0.85	8.7 ± 0.40	16.4 ± 0.88	5030 ± 517	2110 ± 180	4.9 ± 0.6	430 ± 23
R38A	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Y40F	54 ± 0.93	18 ± 0.35	0.9 ± 0.04	20.4 ± 0.56	190 ± 16	66 ± 5.6	1.6 ± 0.26	40 ± 2.6
D56E	18 ± 0.88	6 ± 0.30	1.3 ± 0.06	4.8 ± 0.01	292 ± 2.03	102 ± 0.71	7.1 ± 1.1	14.6 ± 2.3
Y103F	4 ± 0.42	1.4 ± 0.14	1.0 ± 0.22	1.4 ± 0.15	40 ± 0.81	14 ± 0.28	4.9 ± 1.1	3.0 ± 0.63
K117R	1.3 ± 0.81	0.44 ± 0.1	2.93 ± 0.5	0.15 ± 0.03	403 ± 45	141 ± 15	40 ± 6.6	3.6 ± 0.24

[a]His-ChuaHNL was prepared and purified by using a recombinant ChuaHNL in *P. pastoris.*; [b]The kinetic parameters were obtained by monitoring cleavage of (*R*)-MAN by a UV-vis spectrometer at 280 nm.; [c]The kinetic parameters were obtained by monitoring synthesis of (*R*)-MAN by an HPLC at 254 nm.

determination and mechanism of function. It is already known that the cleavage of cyanohydrins by other HNLs is initiated by the deprotonation of the hydroxyl group by a general base [33]. As shown in Fig. 5, the architectures of the active sites of several HNLs hitherto known are largely different. A His residue is commonly used as a general base in the HNLs belonging to GMC oxidoreductase (Fig. 5A) [34,35], α/β-hydrolases (Fig. 5B, 5C) [36,37], dimeric $\alpha + \beta$ barrel folds (Fig. 5D) [18], and cupin superfamily [3,38]. In the α/β-hydrolase HNL family, Ser deprotonated by His is used as a general base (Fig. 5B) [39]. In the Bet v1 superfamily HNL (*Dt*HNL), Tyr deprotonated by a water bridges between the OH-group of Tyr and the guanidinium group of Arg (Fig. 5E), which is an unusual case among the known HNLs [9]. In the case of ChuaHNL, the hydroxyl group of (*R*)-MAN interacts with R38 and K117. Considering with a lower pKa of Lys than Arg (10.5 and 12.5, respectively), K117 would act as a base in the cleavage reaction of cyanohydrins. The deprotonated state of Lys is unstable due to its high pKa, this pKa should be perturbed and decreased to the reactive pH 4–5 of ChuaHNL (Fig. 5F). The hydrophobic environment of the active site could stabilize the deprotonated state of K117, resulting in the pKa decrease. It has been reported that the pKa of Lys is decreased in a hydrophobic environment of various enzymes by excluding water molecules which donate protons and Lys can act as a base which abstracts a proton from substrates [40,41]. As observed in several enzymes, it is possible that R38 could act as a base [42]. The weak cleavage activity of K117R variant ($k_{cat} = 0.44 \pm 0.1$ s^{-1}) suggests that Arg may act as a weak base. The proton abstracted by K117 is donated to the released cyanide ion to produce a hydrogen cyanide. However, the crystal structure of the complex formed with a cyanide ion indicated that

the cyanide ion interacts with Y40 and R38 (Fig. 3C). Therefore, R38 acts the proton donor for the produced cyanide ion to form hydrogen cyanide (HCN). This implies that in the catalytic pocket, hydrogen of R38 is more prone to dissociate than HCN.

The interaction between the nitrile group and the positively charged residues is found in other HNLs, such as His and FAD in GMC oxidoreductase family, Lys in α/β-hydrolase family, and His in cupin superfamily. In ChuaHNL, the nitrile group of cyanohydrin interacts with R38 and Y103. Mutation in these residues caused enzyme inactivation, indicating the importance of the alignment of the nitrile group during the enzymatic reaction.

From the crystal structures, docking simulation, mutational analysis, and kinetic studies, we propose the catalytic mechanism of ChuaHNL (Fig. 6). By the desolvation effect in the hydrophobic active site, the deprotonated state of K117 and the protonated state of D56 are stabilized. As a result, the proton is shared by K117 and D56 (Fig. 6A, the resting state). Upon the binding of a cyanohydrin, K117 abstracts a proton from cyanohydrin and an electron is received by the carbon atom of the cyanide group (Fig. 6B). The pKa of R38 is decreased to less than that of cyanide ion (pKa: 10.5) by the desolvation effect in the hydrophobic active site. The released cyanide ion abstracts a proton from R38 and a hydrogen cyanide is produced (Fig. 6D). Aldehyde and cyanide ions are released from the active site. The proton bound on D56 is transferred to R38 and the active site is returned to the resting state (Fig. 6D). Thus, in the proposed catalytic reaction, K117 acts as a general base and R38 acts as a general acid.

During the synthesis reaction of cyanohydrins, the reaction could occur as suggested for other HNLs. The proton of the bound hydrogen cyanide was

abstracted by R38 and the negatively charged carbon of cyanide ion acts as a nucleophile and attacks the carbon of the aldehyde group of the substrate (Fig. 6C). The hydrogen bound on K117 was abstracted to the aldehyde group of the substrate concomitantly with the electron binding to the hydrogen (Fig. 6B). The synthesized cyanohydrin is released out of the active site of ChuaHNL (Fig. 6A). Therefore,

ChuaHNL utilizes a general acid/base catalysis mechanism during the cleavage and synthesis reactions.

The lipocalin protein family is known as a large group of small proteins distributed among three kingdom of life, with the molecular masses mostly of 18 to 20 kDa, and their primary sequence diversity is large [26]. Lakshmi et al report that the sequence identities of over 80% of typical lipocalins are in the range as
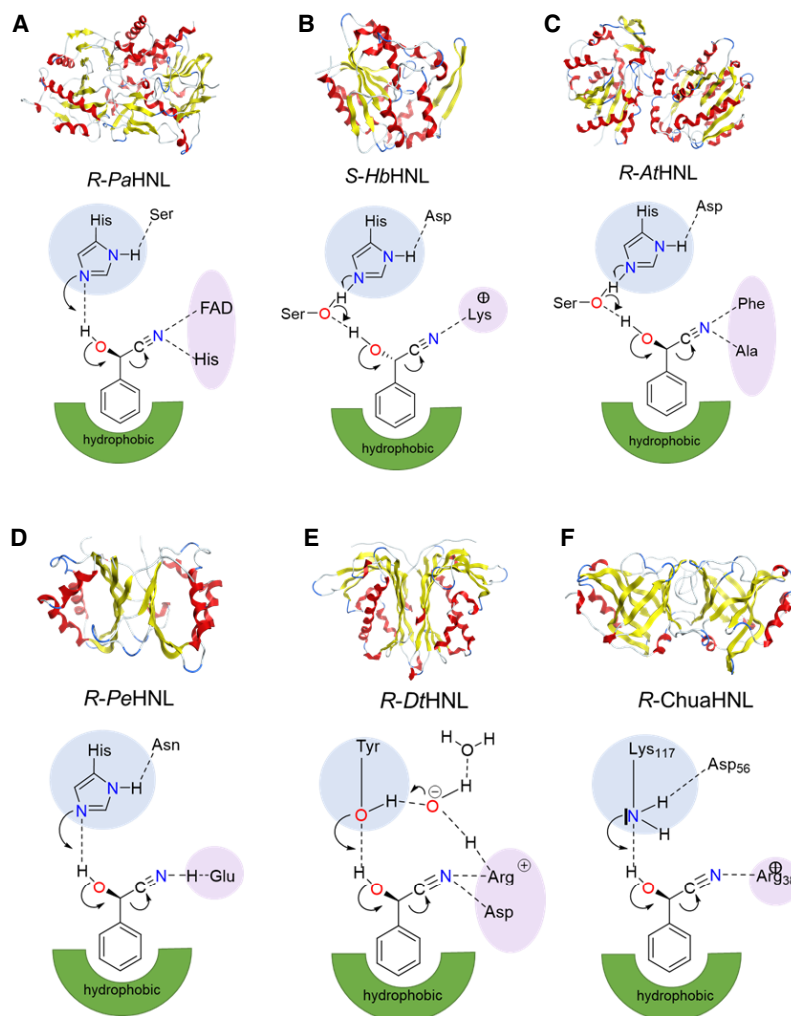

Fig. 5. A general base catalysis model of HNLs. (A) Structure and proposed mechanism of GMC oxidoreductase (*Pa*HNL). His residue acts as a general base. (B) Structure and proposed mechanism of α/β-hydrolase (*S-Hb*HNL). Ser deprotonated by His acts as a general base. (C) Structure and proposed mechanism of α/β-hydrolase (*R-At*HNL). Ser deprotonated by His acts as a general base. (D) Structure and proposed mechanism of HNL with dimeric α + β barrels (*R-Pe*HNL). His residue acts as a general base. (E) Structure and proposed mechanism of Bet v1 superfamily HNL (*R-Dt*HNL). Tyr deprotonated by a water bridging between the guanidinium group of Arg acts as a general base. (F) Structure and proposed mechanism of ChuaHNL. The hydroxyl group of (*R*)-MAN interacts with R38 and K117. Lys acts as a general base in the cleavage reaction of cyanohydrins. In all of the models, the first step of the catalysis is deprotonation from the hydroxyl group of cyanohydrins. Except *Dt*HNL and ChuaHNL, His residue acts as a base to protrude hydrogen cyanide from mandelonitrile, whereas *Dt*HNL and ChuaHNL use Tyr and Lys, respectively, as a general base. The negatively charged nitrile groups interact with basic residue except *Pe*HNL. *Pa*HNL; *Prunus amygdalus*. *Hb*HNL; *Hevea brasiliensis*. *At*HNL; *Arabidopsis thaliana*. *Pe*HNL; *Passiflora edulis*. *Dt*HNL; *Davallia tyermannii*. The protein structures were designed using the program PyMOL [61].

low as 1–20% [43]. On the other hand, their three-dimensional structures are highly conserved. It is accepted that the 3D-structures are conserved better than the amino acid sequences, which have been traditionally used for sequence-based phylogenetic approaches. They have several disulfide bonds. They form a single eight-stranded hydrogen-bonded antiparallel β-barrel, which forms a binding site. Some of them have abilities to bind small hydrophobic molecules. They exhibit functional diversity, with roles in transport of retinol, olfaction, cryptic coloration, and pheromone. They also appear to function in immunomodulation and the regulation of cell homoeostasis. Violaxanthin de-epoxidase and zeaxanthin epoxidase of plant origins are grouped as lipocalins from their primary structures [68]. Prostaglandin D synthase (EC5.3.99.2) is one of a few characterized enzymes belonging to the lipocalin family [26,44-47]. The enzyme catalyzes the isomerization of 9,11-endoperoxide group of prostaglandin precursor to prostaglandin D_2. Recently, it was reported that the protein has dual functional roles as the enzyme and as an extracellular transporter for diverse lipophilic compounds in the cerebrospinal fluid [48]. So, in our study, the elucidation of the structure and function of ChuaHNL and the future study seeking for ancestral

protein is of much interest based on its genome structure. Many protein frames with similar primary structures to ChuaHNL are found in the genome of *C. hualienensis* (unpublished results), although these proteins have no identity among database. Thus, it is indicated that ChuaHNL might have an ancestral protein which is functioning in the transport or olfactory responses on low molecular weight compounds in the arthropod.

In our previous study, ChuaHNL was found to be synthesized and localized in the defensive secretory gland(s) in the paraterga of the millipede by a histochemical localization experiments [5,49]. The millipede has a storage chamber of neat (*R*)-MAN synthesized by the inner cells of the chamber, upstream of the way out of the products HCN gas and benzaldehyde through the hole called ozopore. It is noteworthy that (*R*)-MAN is pooled neat in the storage chamber without glycosylation, as we could not detect the glycosylated (*R*)-MAN in the millipede [28]. In the next chamber connected to the reaction chamber, ChuaHNL is specifically biosynthesized and exuded at the ozopore to outside through a duct of the ozopore from the reaction chamber. Thus, because the neat substrate would be supplied in a higher concentration from the storage chamber to the next chamber
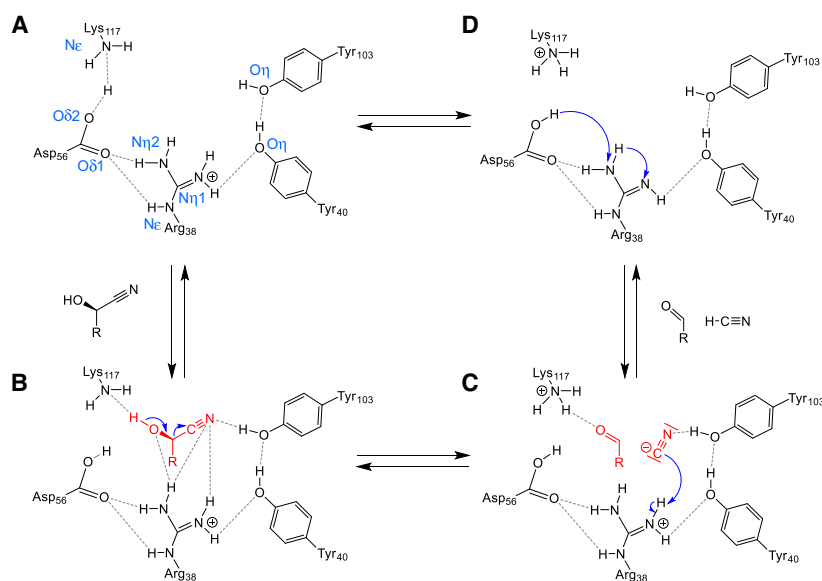

Fig. 6. The proposed catalytic cycle of ChuaHNL. (A) The resting state. The deprotonated state of Lys117 is stabilized by the desolvation effect in the hydrophobic active site. The residues are labeled and colored in cyan. (B) The binding of cyanohydrin in the active site. The bound cyanohydrin is colored in red. Lys117 forms a hydrogen bond with the cyanohydrin and abstracts a proton from the cyanohydrin. The electron released from the hydrogen is received by the carbon atom of the nitrile group. (C) The formation of hydrogen cyanide. The released cyanide ion abstracts a proton from Arg38. Subsequently, aldehyde and cyanide ions are released out of the active site. (D) The recovering state. The proton bound on Asp56 is transferred to Arg38 and the active site is returned to the resting state

containing HNL and the product HCN gas and benzaldehyde are pushed out instantly through the ozopore to outside the paraterga by the action of the animal, higher K_m value (8.7 mM) for (R)-MAN and the extremely high specific activity for (R)-MAN cleavage (406 U·mg^{-1}) shown by ChuaHNL are quite reasonable, not like most of other plant HNLs which usually act in lower substrate concentrations after the cyanogenic glycosides are hydrolyzed by β-glucosidase reaction [50]. The V_{max} values of the HNLs from plant origins are compared: they usually show less than 100 U·mg^{-1} in the degradation of MAN and sometime more than 100 U·mg^{-1} in MAN synthesis. Km values for MAN and benzaldehyde are around 1.0 mM and 10 mM or less, respectively [2,23].

A phylogenetic analysis was performed for the ChuaHNL, with the four similar structures obtained by DALI search and 42 kinds of typical lipocalins having diverse primary sequences and the 3D structures [43]. Fig. S2 shows that the primary structure of ChuaHNL is close to rat epididymal retinoic acid binding protein (PDB ID: 1EPA) [52,63]. In Fig. S3, structural superimposition and a comparison were made on the size and shape of the binding cavity of ChuaHNL with the lipocalin family proteins, having high Z-scores by DALI search, which are shown in Fig. 2: human retinol binding protein 4 (RBP4) and rat epididymal-specific retinoic acid binding protein, lipocalin 5 (LCN_RAT). In Fig. S4, structural superimposition and comparison of catalytic pockets in ChuaHNL and prostaglandin D synthase were made. Although the 3D structure of ChuaHNL showed similarities with these proteins, the direction and the size of the entrances for the substrates, and the depths of the cavities are quite different, show in the unique structure of ChuaHNL.

The significance of the current research is that the enzyme shows a totally novel primary structure as well as a new 3D structure, and the new key residues in the catalysis, as compared with the known HNLs. The addition of our arthropod HNL with the high similarity of the three-dimensional structure to the lipocalin family, well-studied proteins with variety of functions in mammals, indicates that we are expanding the evolutionary relationship of HNLs to a new family of proteins since these enzymes would be good examples of convergent evolution. HNLs are from various protein families which catalyze the same reaction in nature. One of the most important HNLs is originated from almond (R-PaHNL) which looks like an oxidase activity, as evidenced also by its 3D structure [14]. Therefore, it is noteworthy that the new HNL from *C. hualienensis*

catalyzes this reaction while it evolutionally forms a new scaffold.

Materials and Methods

Materials and chemicals

The invasive millipedes *Chamberlinius hualienesis* were manually collected in the Japanese cedar forest located in Kyushu, Japan, and were stored at −80°C until use for protein purification. Benzaldehyde (redistilled, 99.5%), (R)-MAN, and (rac)-MAN were purchased from Sigma-Aldrich (St. Louis, MO, USA). All chemicals used for protein crystallization were purchased from Hampton Research (Aliso Viejo, CA, USA). Other chemicals were used in the experiments were purchased from chemical sources and used without further purification.

Enzyme assays

The enzyme activities for (R)-MAN synthesis and MAN cleavage were assayed by HPLC and spectrophotometric methods, respectively. The HNL activity during (R)-MAN synthesis was measured as reported previously [5]. In brief, the reaction (total volume = 1 mL) was initiated by adding potassium cyanide (KCN; 100 mM) in citrate buffer (400 mM, pH 4.2) containing benzaldehyde (50 mM; 40 μL of 1250 mM benzaldehyde dissolved in dimethyl sulfoxide (DMSO)) and enzyme solution (0.5-5 U·mL^{-1}). The mixture was incubated at 22 °C for 5 min, and then, aliquots (100 μL) of reactant were mixed with 9-fold volume of *n*-hexane:2-propanol (85:15) mixture. Finally, the organic phase was analyzed using an HPLC instrument (UFLC Prominence Liquid Chromatography LC-20AD, Shimadzu, Kyoto, Japan) equipped with a chiral column (CHIRAL-CEL OJ-H, Daicel, Osaka, Japan) (i.d. 4.6 mm × 250 mm length, 5 μm particle size). One unit of synthesis activity was defined as the amount of the enzyme that synthesizes 1 μmol of (R)-MAN from benzaldehyde and KCN per min.

For MAN cleavage activity assay, the reaction was started by adding enzyme sample to a reaction mixture (1 mL) consisting of racemic mandelonitrile or (R)-MAN (2 mM) in citrate buffer (100 mM, pH 5.0). The reaction velocity of benzaldehyde formation was followed by monitoring absorbance at 280 nm for one min (extinction coefficient of benzaldehyde = 1,352 L mol^{-1} cm^{-1}) using a UV-2600 UV-Vis spectrophotometer (Shimadzu, Kyoto, Japan).

Natural ChuaHNL preparation and crystallography

The native ChuaHNL protein was purified by the protocol reported previously [5]. In brief, first, a homogenate of millipedes solubilized in buffer-A (20 mM potassium

phosphate, pH 7.0), was fractionated by an ammonium sulfate precipitation procedure. Next, the native ChuaHNL protein was purified using TOYOPEARL Butyl-650M (Tosoh, Tokyo, Japan), TOYOPEARL DEAE-650M (Tosoh, Tokyo, Japan), Q Sepharose FF (GE Healthcare, Chicago, IL, USA), and Superdex 75 10/300 GL (GE Healthcare, Chicago, IL, USA). Finally, the active fractions were collected and subjected to sodium dodecyl sulfate/polyacrylamide gel electrophoresis (SDS–PAGE) [51] to evaluate their purity. For protein crystallization, the buffer was replaced by 50 mM citrate buffer (pH 5.0) and then was concentrated up to 10 mg ml^{-1} by using an Amicon Ultra Centrifugal Filter Unites NMWL, 10 kDa (Merck Millipore, Billerica, MA, USA). Protein concentration was measured using a Quick Start Protein Assay (Bio-Rad Laboratories, Hercules, CA, USA) using bovine serum albumin (Sigma-Aldrich, St. Louis, MO, USA) as a standard.

First, the crystallization conditions were screened using CrystalScreen I and II (Hampton Research) and bipyramidal crystals were obtained using the following conditions: 0.2 M ammonium sulfate, 0.1 M sodium acetate (pH 4.6), and 30 % (w/v) polyethylene glycol (PEG) monomethyl ether 2,000. Finally, the best crystals were obtained using a hanging drop vapor diffusion method at 20 ˚C for 3 days in the following conditions: 0.2 M ammonium sulfate, 0.1 M sodium acetate (pH 5.0), 28-32 % (w/v) PEG monomethyl ether 2,000, and 0.3 M NDSB-195.

For the phase determination, a native ChuaHNL crystal was soaked in the reservoir solution containing 0.5 M sodium iodide at 20 ˚C for 30 min prior to freezing and data collection. However, for the native and recombinant structures complexed with acetate, the obtained crystals were directly used for the data collection. For the ligand-free form, a native protein crystal was soaked in soaking solution-1 (32 % (w/v) PEG monomethyl ether 2,000, 0.3 M NDSB-195, 50 mM bis-tris-propane, and 50 mM citric acid; pH 4.5) at 20 ˚C for 20 min to remove acetate. For complexes with thiocyanate, iodoacetate, and cyanide ion, crystals of the ligand-free form were soaked in soaking solution-1 containing either 80 mM potassium thiocyanate, 10 mM sodium iodoacetate, or 0.5 M potassium cyanide, respectively. All crystals were cryoprotected with perfluoroether and the data sets were collected at 100 K under the nitrogen stream. The X-ray diffraction data for single anomalous dispersion (SAD) phasing were collected using an in-house X-ray generator and an imaging plate (MicroMAX-007 and R-AXIS VII, Rigaku, Tokyo, Japan). Other data were collected using a silicon pixel detector (Pilatus 2M-F, DECTRIS, Baden-Daettwil, Switzerland) and a CCD detector (Quantum 315r, Area Detector Systems, Poway, CA, USA) at Photon Factory beamline BL-1A and BL-5A (Tsukuba, Japan).

All data sets were integrated using XDS [53] and scaled using SCALA [54]. The initial phase was determined using SHELX suite [55]. All models were corrected using COOT [56] and refined using either REFMAC5 [57] or PHENIX [58] programs.

Structural Analysis and Docking Simulation

All structures reported in this paper were validated using MolProbity [59], and there was no residue in disallowed regions of the Ramachandran plot. Statistics for phase determination and those for data collections and structure refinements are listed in Table 1. The structural analysis was performed using PISA [60]. The protein structures were designed using the program PyMOL (http://www.pymol.org) [61]. (*R*)-MAN was docked with ChuaHNL by using the Molecular Operating Environment program (MOE; version 2016.8, Montreal, Canada).

Microbial strains, media, and culture condition

Escherichia coli DH5α (Takara Bio, Otsu, Japan) was inoculated in Luria–Bertani (LB) medium (1% tryptone, 0.5% yeast extract, and 0.5% NaCl) containing zeocin (25 µg·mL^{-1}) for plasmid propagation. *P. pastoris* cells harboring PpPDI/GS115 was used as host for the recombinant His-ChuaHNL expression as described previously [27]. The yeast strain was precultured in a buffered minimal glycerol (BMG) medium (1.34 % yeast nitrogen base without amino acids, 4×10^{-5} % biotin, 100 mM potassium phosphate buffer (pH 7.0), and 1 % glycerol) and then the yeast cells were transferred to induction BMM medium (BMG in which 1 % glycerol was replaced by 1 % methanol). The culture was grown at 28 ˚C under aerobic conditions with reciprocal shaking (150 rpm), and the growth of the yeast was monitored by measuring the optical density at 600 nm wavelength.

Construction of His-ChuaHNL variants for expression in *P. pastoris*

The site-directed mutagenesis was performed using a Quick-Change Lightning Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA). The primer pairs are listed in Table S1 and pPICZαA-His-ori-ChuaHNL vector [27] was used as template. The PCR products were treated with *Dpn*I (TaKaRa) at 37 ˚C for 1 h and then transformed to *E. coli* DH5α strain and selected on a low-salt LB agar plate with Zeocin (25 µg·mL^{-1}). The mutant plasmids were isolated by a plasmid miniprep kit (Sigma-Aldrich), and the amino acid exchanges were confirmed by DNA sequencing analysis. The constructed vector pPICZαA-His-ori-ChuaHNL mutant was then linearized by digesting with *Sac*I and electroporated into *P. pastoris* PpPDI/GS115, according to our previous report [27].

Expression and purification of His-ChuaHNL in *P. pastoris* system

The coexpression strain, *P. pastoris* ori-His-ChuaHNL/PpPDI/GS115 mutant, was expressed and purified as described previously [27]. Initially, after six days of an induction BMM medium cultivation, the pH of the supernatant was adjusted to 7.5 by addition of 2 M sodium hydroxide. Next, the His-tagged protein was purified using Ni Sepharose 6 fast flow column (GE Healthcare, Chicago, IL, USA), Mono Q 5/50 GL (GE Healthcare, Chicago, IL, USA). Finally, the purified His-ChuaHNL mutants were dialyzed into 20 mM KPB buffer (pH 7.0) and analyzed the purity by SDS–PAGE.

Kinetic parameters analysis

Kinetic parameters for (*R*)-MAN synthesis from benzaldehyde and KCN of purified wild-type and mutants of His-ChuaHNL were analyzed in citrate buffer (400 mM pH 4.2), according to the method described above, at various concentrations (0.5-50 mM) of benzaldehyde (followed by mixing with total 40 μL of 12.5-1250 mM benzaldehyde dissolved in DMSO). For the cleavage reaction of (*R*)-MAN, the initial velocity was determined over a substrate concentration of 0.1-5 mM for (*R*)-MAN in citrate buffer (400 mM pH 5.0), according to the method described above. The kinetic parameters of the enzyme were calculated by nonlinear least-square curve fitting against plots of the initial velocity vs substrate concentrations using the Michaelis–Menten equation.

Acknowledgements

Author contributions

FM, AI, and YA planned experiments; FM, AI, AN, ZZ, MD, SS, TY, and SN performed experiments; FM, AI, AN, TY, and SN analyzed data; and FM, AI, AN, and YA wrote the paper.

Conflicts of interest

The authors declare no conflict of interest.

References

1 Gruber-Khadjawi M, Fechter MN & Griengl H (2011) Cleavage and formation of cyanohydrin. In Enzyme Catalysis in Organic Synthesis (Drauz K, Gröger H. & Griengl H, eds), pp. 947–990.

2 Dadashipour M & Asano Y (2011) Hydroxynitrile lyases: insights into biochemistry, discovery, and engineering. *ACS Catal* **1**, 1121–1149.

3 Hajnal I, Lyskowski A, Hanefeld U, Gruber K, Schwab H & Steiner K (2013) Biochemical and structural characterization of a novel bacterial manganese-dependent hydroxynitrile lyase. *FEBS J* **280**, 5815–5828.

4 Wiedner R, Gruber-Khadjawi M, Schwab H & Steiner K (2014) Discovery of a novel (*R*)-selective bacterial hydroxynitrile lyase from *Acidobacterium capsulatum*. *Comput Struct Biotechnol J* **10**, 58–62.

5 Dadashipour M, Ishida Y, Yamamoto K & Asano Y (2015) Discovery and molecular and biocatalytic properties of hydroxynitrile lyase from an invasive millipede, *Chamberlinius hualienensis*. *Proc Natl Acad Sci* **112**, 10605–10610.

6 Yamaguchi T, Nuylert A, Ina A, Tanabe T & Asano Y (2018) Hydroxynitrile lyases from cyanogenic millipedes: molecular cloning, heterologous expression, and whole-cell biocatalysis for the production of (*R*)-mandelonitrile. *Sci Rep* **8**, 3051.

7 Asano Y, Tamura K, Doi N, Ueatrongchit T, H-kittikun A & Ohmiya T (2005) Screening for new hydroxynitrilases from plants. *Biosci Biotechnol Biochem* **69**, 2349–2357.

8 Tomescu MS, Davids D, DuPlessis M, Darnhofer B, Birner-Gruenberger R, Archer R, Schwendenwein D, Thallinger G, Winkler M & Rumbold K (2020) High-throughput in-field bioprospecting for cyanogenic plants and hydroxynitrile lyases. *Biocatal Biotransform* **38**, 234–240.

9 Lanfranchi E, Pavkov-Keller T, Koehler E-M, Diepold M, Steiner K, Darnhofer B, Hartler J, Van Den Bergh T, Joosten H-J & Gruber-Khadjawi M

(2017) Enzyme discovery beyond homology: a unique hydroxynitrile lyase in the Bet v1 superfamily. *Sci Rep* **7**, 46738.

10 Osbourn AE (1996) Preformed antimicrobial compounds and plant defense against fungal attack. *Plant cell* **8**, 1821–1831.

11 Bracco P, Busch H, von Langermann J & Hanefeld U (2016) Enantioselective synthesis of cyanohydrins catalysed by hydroxynitrile lyases - a review. *Org Biomol Chem* **14**, 6375–6389.

12 Wöhler F & von Liebig JF (1832) Ueber die Bildung des Bittermandelöls. *Ann Pharm* **22**, 1–24.

13 Vennesland B, Conn EE, Knowles CJ, Westley J, Wissing F Eds. Cyanide in Biology, 1st ed. (1981); Academic Press, London. https://openlibrary.org/books/OL3789773M/Cyanide_in_biology

14 Dreveny I, Gruber K, Glieder A, Thompson A & Kratky C (2001) The hydroxynitrile lyase from almond: A lyase that looks like an oxidoreductase. *Structure* **9**, 803–815.

15 Guterl JK, Andexer JN, Sehl T, von Langermann J, Frindi-Wosch I, Rosenkranz T, Fitter J, Gruber K, Kragl U, Eggert T & *et al.* (2009) Uneven twins: comparison of two enantiocomplementary hydroxynitrile lyases with alpha/beta-hydrolase fold. *J Biotechnol* **141**, 166–173.

16 Nakano S, Dadashipour M & Asano Y (2014) Structural and functional analysis of hydroxynitrile lyase from *Baliospermum montanum* with crystal structure, molecular dynamics and enzyme kinetics. *Biochim Biophys Acta Proteins Proteom* **1844**, 2059–2067.

17 Lauble H, Miehlich B, Forster S, Wajant H & Effenberger F (2002) Crystal structure of hydroxynitrile lyase from *Sorghum bicolor* in complex with the inhibitor benzoic acid: a novel cyanogenic enzyme. *Biochemistry* **41**, 12043–12050.

18 Lauble H, Miehlich B, Förster S, Kobler C, Wajant H & Effenberger F (2002) Structure determinants of substrate specificity of hydroxynitrile lyase from *Manihot esculenta*. *Protein Sci* **11**, 65–71.

19 Trummler K & Wajant H (1997) Molecular cloning of acetone cyanohydrin lyase from flax (*Linum usitatissimum*). Definition of a novel class of hydroxynitrile lyases. *J Biol Chem* **272**, 4770–4774.

20 Motojima F, Nuylert A & Asano Y (2018) The crystal structure and catalytic mechanism of hydroxynitrile lyase from passion fruit, *Passiflora edulis*. *FEBS J* **285**, 313–324.

21 Dreveny I, Andryushkova AS, Glieder A, Gruber K & Kratky C (2009) 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**, 3370–3377.

22 Padhi SK, Fujii R, Legatt GA, Fossum SL, Berchtold R & Kazlauskas RJ (2010) Switching from an esterase to a hydroxynitrile lyase mechanism requires only two amino acid substitutions. *Chem Biol* **17**, 863–871.

23 Isobe K, Kitagawa A, Kanamori K, Kashiwagi N, Matsui D, Yamaguchi T, Fuhshuku K, Semba H & Asano Y (2018) Characterization of a novel hydroxynitrile lyase from *Nandina domestica* Thunb. *Biosci Biotech Biochem* **82**, 1760–1769.

24 Holm L & Rosenström P (2010) DALI server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545–W549.

25 Mueller GA, Benjamin DC & Rule GS (1998) Tertiary structure of the major house dust mite allergen Der p 2: sequential and structural homologies. *Biochemistry* **37**, 12707–12714.

26 Flower DR (1996) The lipocalin protein family: structure and function. *Biochem J* **318**, 1–14.

27 Zhai Z-Y, Nuylert A, Isobe K & Asano Y (2019) Effects of codon optimization and glycosylation on the high-level production of hydroxynitrile lyase from *Chamberlinius hualienensis* in *Pichia pastoris*. *J Ind Microbiol Biotechnol* **46**, 887–898.

28 Yamaguchi T, Kuwahara Y & Asano Y (2017) A novel cytochrome P450, CYP3201B1, is involved in (*R*)-mandelonitrile biosynthesis in a cyanogenic millipede. *FEBS Open Bio* **7**, 335–347.

29 Kuwahara Y, Tanabe T & Asano Y (2017) Defensive allomone components of an unidentified myrmecophilous millipede in the genus *Ampelodesmus* ("hagayasude" in Japanese) [Polydesmoida: Purgodesmidae]. *Jpn J Environ Entmol Zool* **28**, 55–62.

30 Kuwahara Y, Yamaguchi T, Ichiki Y, Tanabe T & Asano Y (2017) Hydrogen peroxide as a new defensive compound in "benzoyl cyanide" producing polydesmid millipedes. *Sci Nat* **104**, 19.

31 Ishida Y, Kuwahara Y, Dadashipour M, Ina A, Yamaguchi T, Morita M, Ichiki Y & Asano Y (2016) A sacrificial millipede altruistically protects its swarm using a drone blood enzyme, mandelonitrile oxidase. *Sci Rep* **6**, 26998.

32 Kuwahara Y, Shimizu N & Tanabe T (2011) Release of hydrogen cyanide via a post-secretion Schotten-Baumann reaction in defensive fluids of polydesmoid millipedes. *J Chem Ecol* **37**, 232–238.

33 Gruber K & Kratky C (2004) Biopolymers for biocatalysis: Structure and catalytic mechanism of hydroxynitrile lyases. *J Polym Sci Part A: Polym Chem* **42**, 479–486.

34 Cheng I-P & Poulton JE (1993) Cloning of cDNA of *Prunus serotina* (*R*)-(+)-mandelonitrile lyase and identification of a putative FAD-binding site. *Plant Cell Physiol* **34**, 1139–1143.

35 Hu Z & Poulton JE (1997) Sequencing, genomic organization, and preliminary promoter analysis of a

black cherry (*R*)-(+)-mandelonitrile lyase gene. *Plant Physiol* **115**, 1359–1369.

36 Hasslacher M, Schall M, Hayn M, Griengl H, Kohlwein SD & Schwab H (1996) 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* **271**, 5884–5891.

37 Wajant H & Effenberger F (1996) Hydroxynitrile lyases of higher plants. *Biol Chem* **377**, 611–617.

38 Vertregt F, Torrelo G, Trunk S, Wiltsche H, Hagen WR, Hanefeld U & Steiner K (2016) EPR study of substrate binding to Mn (II) in hydroxynitrile lyase from *Granulicella tundricola*. *ACS Catal* **6**, 5081–5085.

39 Stranzl GR, Gruber K, Steinkellner G, Zangger K, Schwab H & Kratky C (2004) Observation of a short, strong hydrogen bond in the active site of hydroxynitrile lyase from *Hevea brasiliensis* explains a large p*K*a shift of the catalytic base induced by the reaction intermediate. *J Biol Chem* **279**, 3699–3707.

40 García-Moreno BE, Dwyer JJ, Gittis AG, Lattman EE, Spencer DS & Stites WE (1997) Experimental measurement of the effective dielectric in the hydrophobic core of a protein. *Biophys Chem* **64**, 211–224.

41 Ho M-C, Ménétret J-F, Tsuruta H & Allen KN (2009) The origin of the electrostatic perturbation in acetoacetate decarboxylase. *Nature* **459**, 393–397.

42 Schlippe YVG & Hedstrom L (2005) A twisted base? The role of arginine in enzyme-catalyzed proton abstractions. *Arch Biochem Biophys* **433**, 266–278.

43 Lakshmi B, Mishra M, Srinivasan N & Archunan G (2015) Structure-based phylogenetic analysis of the lipocalin superfamily. *PLoS One* **10**, e0135507.

44 Grzyb J, Latowski D & Stzalka K (2006) Lipocalins-a family portrait. *J Plant Physiol* **163**, 895–915.

45 Shimamoto S, Yoshida T, Inui T, Gohda K, Kobayashi Y, Fujimori K, Tsurumura T, Aritake K, Urade Y & Ohkubo T (2007) NMR solution structure of lipocalin-type prostaglandin D synthase: evidence for partial overlapping of catalytic pocket and retinoic acid-binding pocket within the central cavity. *J Biol Chem* **282**, 31373–31379.

46 Kumasaka T, Aritake K, Ago H, Irikura D, Tsurumura T, Yamamoto M, Miyano M, Urade Y & Hayaishi O (2009) Structural basis of the catalytic mechanism operating in open-closed conformers of lipocalin type prostaglandin D synthase. *J Biol Chem* **284**, 22344–22352.

47 Shimamoto S, Yoshida T & Ohkubo T (2011) Ligand recognition mechanism of lipocalin-type prostaglandin D synthase (in Japanese). *Yakugaku Zasshi* **131**, 1575–1581.

48 Elmes MW, Volpe AD, d'Oelsnitz S, Sweeney JM & Kaczocha M (2018) Lipocalin-type prostaglandin D

synthase is a novel phytocannabinoid-binding protein. *Lipids* **53**, 353–360.

49 Eisner T & Meinwald J (1966) Defensive secretions of arthropods. *Science* **153**, 1341–1350.

50 Yamaguchi T & Asano Y (2018) Prunasin production using engineered *Escherichia coli* expressing UGT85A47 from Japanese apricot and UDP-glucose biosynthetic enzyme genes. *Biosci Biotechnol Biochem*. **82**, 2021–2029.

51 Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

52 Newcomer ME, Jones TA, Aqvist J, Sundelin J, Eriksson U, Rask L & Peterson PA (1984) The three-dimensional structure of retinol-binding protein. *EMBO J* **3**, 1451–1454.

53 Kabsch W. (2010) Xds. *Acta Crystallogr Sect D Biol Crystallogr* **66**, 125–132.

54 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr Sect D Biol Crystallogr* **67**, 235–242.

55 Sheldrick GM (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr Sect D Biol Crystallogr* **66**, 479–485.

56 Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr Sect D Biol Crystallogr* **60**, 2126–2132.

57 Murshudov GN, Vagin AA & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr Sect D Biol Crystallogr* **53**, 240–255.

58 Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH & Adams PD (2012) Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallogr Sect D Biol Crystallogr* **68**, 352–367.

59 Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr Sect D Biol Crystallogr* **66**, 12–21.

60 Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774–797.

61 Schrödinger LLC. The PyMOL Molecular Graphics System, Version 2.0.5. https://pymol.org/2/

62 Zanotti G, Ottonello S, Berni R & Monaco HL (1993) Crystal structure of the trigonal form of human plasma retinol-binding protein at 2.5 A resolution. *J Mol Biol* **230**, 613–624.

63 Newcomer ME (1993) Structure of the epididymal retinoic acid binding protein at 2.1 Å resolution. *Structure* **1**, 7–18.

64 Huber R, Schneider M, Mayr I, Müller R, Deutzmann R, Suter F, Zuber H, Falk H & Kayser H (1987) Molecular structure of the bilin binding protein (BBP) from *Pieris brassicae* after refinement at 2.0 Å resolution. *J Mol Biol* **198**, 499–513.

65 Böcskei Z, Groom CR, Flower DR, Wright CE, Phillips SE, Cavaggioni A, Findlay JB & North AC (1992) Pheromone binding to two rodent urinary proteins revealed by X-ray crystallography. *Nature* **360**, 186–188.

66 Pei J & Grishin NV (2014) PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information *Methods Mol. Biol* **1079**, 263–271.

67 Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320–W324.

68 Hieber AD, Bugos RC & Yamamoto HY (2000) Plant lipocalins: violaxanthin de-epoxidase and zeaxanthin epoxidase. *Biochim Biophys Acta* **482** (1–2), 84–91.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. List of primers.

Fig. S1. N-glycans in ChuaHNL structure.

Fig. S2. Phylogenetic analysis of the ChuaHNL (red colored) and 42 lipocalin domains.

Fig. S3. Structural superimposition and cavity binding comparison of ChuaHNL, RBP4 and LCN_RAT.

Fig. S4. Structural superimposition and comparison of catalytic pockets in ChuaHNL and prostaglandin D synthase.