Published in final edited form as:

Proteins. 2009 January ; 74(1): 61–71. doi:10.1002/prot.22124.

X-RAY STRUCTURE OF ILL2, AN AUXIN-CONJUGATE AMIDOHYDROLASE FROM *ARABIDOPSIS THALIANA*

Eduard Bitto¹, Craig A. Bingman¹, Lenka Bittova, Norma L. Houston², Rebecca S. Boston², Brian G. Fox^{1,3}, and George N. Phillips Jr.^{1,3,*}

¹ Center for Eukaryotic Structural Genomics, University of Wisconsin-Madison, Madison, WI 53706-1544

³ Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706-1544

² Department of Plant Biology, North Carolina State University, Raleigh, NC 27695

Abstract

The plant hormone indole-3-acetic acid (IAA) is the most abundant natural auxin involved in many aspects of plant development and growth. The IAA levels in plants are modulated by a specific group of amidohydrolases from the peptidase M20D family that release the active hormone from its conjugated storage forms. Here we describe the X-ray crystal structure of IAA-amino acid hydrolase IAA-leucine resistant-like gene 2 (ILL2) from *Arabidopsis thaliana* at 2.0 Å resolution. ILL2 preferentially hydrolyses the auxin-amino acid conjugate *N*-(indol-3-acetyl)-alanine. The overall structure of ILL2 is reminiscent of dinuclear metallopeptidases from the M20 peptidase family. The structure consists of two domains, a larger catalytic domain with 3-layer αβ sandwich architecture and aminopeptidase topology and a smaller satellite domain with 2-layer αβ sandwich architecture and alpha-beta plaits topology. The metal coordinating residues in the active site of ILL2 include a conserved cysteine that clearly distinguishes this protein from previously structurally characterized members of the M20 peptidase family. Modeling of *N*-(indol-3-acetyl)-alanine into the active site of ILL2 suggests that Leu175 serves as a key determinant for the amino acid side chain specificity of this enzyme. Furthermore, a hydrophobic pocket nearby the catalytic dimetal center likely recognizes the indolyl moiety of the substrate. Finally, the active site of ILL2 harbors an absolutely conserved glutamate (Glu172), which is well positioned to act as a general acid-base residue. Overall, the structure of ILL2 suggests that this enzyme likely uses a catalytic mechanism that follows the paradigm established for the other enzymes of the M20 peptidase family.

Keywords

M20D peptidase family; auxin homeostasis; metalloenzyme; X-ray structure

INTRODUCTION

Auxins are ubiquitous plant hormones involved in regulation of many aspects of plant development and growth including germination, root initiation, phototropism, leaf

*Address correspondence to: George N. Phillips, Jr., Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706, Tel. (608) 263-6142; Fax: (608) 263-6142; E-mail: phillips@biochem.wisc.edu.

Frequently used abbreviations: IAA, indole-3-acetic acid; IAA-Ala, *N*-(indole-3-acetyl)-alanine; ILL2, IAA-leucine-resistant-like gene 2 precursor; rmsd; root mean square deviation; VAST; Vector Alignment Search Tool; CPG2, carboxypeptidase G2; GCPII, glutamate carboxypeptidase II

formation, and fruit development.¹ One of the most common natural auxins is indole-3-acetic acid (IAA). This hormone is present in plants as either a physiologically active free acid or as a non-active conjugate in which the carboxyl group of IAA is derivatized with amino acids, peptides or sugars.^{1,2} IAA-conjugates represent a storage pool from which the active hormone can be released by selective classes of enzymes, such as aminoacid amidohydrolases in the case of IAA-amino acid conjugates, thus providing a means for regulation of auxin levels in plants.^{1,3} The most common auxin conjugates in *Arabidopsis thaliana* include IAA-Ala, IAA-Leu, IAA-Asp, IAA-Glu, IAA-glucose, and IAA-peptide.^{3–5} IAA-Asp may serve as an intermediate in the auxin degradation pathway.⁶

The gene locus At5g56660.1 of *A. thaliana* encodes a 48 kDa IAA-leucine-resistant-like gene 2 precursor (ILL2), an IAA-amino acid hydrolase from the amidohydrolase family (also known as the peptidase M20D family) involved in auxin metabolism.⁷ Multiple enzymes from the amidohydrolase family have been identified in *A. thaliana*. Initially, ILR1 and IAR3 (also known as ILL4) were identified through mutant screens based on reduced sensitivity to IAA-conjugates in the root inhibition bioassay.^{8,9} Five additional isoforms ILL1, ILL2, ILL3, ILL5 and ILL6, which share 40–85% sequence identity, have since been described in *A. thaliana* and shown to have orthologs in other dicots as well as monocots and gymnosperms.^{7,10,11} Such widespread occurrence of a multigene family indicates an ancient evolutionary divergence within the plant kingdom. The *A. thaliana* IAA-hydrolases lack close homologs in non-plant higher eukaryotes, although they share a more distant relationship to an array of bacterial hydrolases (30–40% sequence identity).

The *A. thaliana* isoforms ILL1, IAR3, and ILL5 meet stringent bioinformatics criteria for localization in the endoplasmic reticulum (ER) including having a predicted amino terminal signal sequence and carboxy terminal His/Lys-Asp-Glu-Leu tetrapeptide that serves as an ER retrieval tag.^{9,12,13} The importance of ER localization is not clear, however, as phylogenetic trees of 66 orthologs across the plant kingdom resolved two separate monocot clades, one with members possessing ER retrieval sequences and one lacking them.¹¹ The remaining *A. thaliana* isoforms either lack ER retrieval sequence motifs (ILL3, Lys-Ser-Gly-Asp; ILL6, Asp-Lys-His-Ser) or have variant motifs for which there is evidence of ER retention in fungal and animal systems but not yet in plants (ILR1, Lys-Ser-Glu-Leu; ILL2, His-Glu-Glu-Leu).^{14,15}

Several *A. thaliana* amidohydrolases have been characterized biochemically. In enzymatic assays with purified glutathione S-transferase (GST)-amidohydrolase fusion proteins, ILL2 showed the highest overall hydrolytic activity among these enzymes with a clear preference for IAA-Ala.¹⁰ In contrast, ILL1, which shares 87% identity with ILL2, had the lowest activity toward all the substrates tested. ILR1 preferred IAA-Leu and bulky aromatic conjugates IAA-Tyr and IAA-Phe. Although IAR3 also hydrolyzed IAA-aminoacid conjugates with small side chains, it acted as a relatively poor catalyst. A disparate study has shown that this isoform can hydrolyze amino acid conjugates of the plant hormone jasmonic acid, suggesting a role in wound repair response.^{16,17} The hydrolysis rates for each IAA-amino acid conjugate correlated with the results of a root inhibition assay, suggesting that the hydrolase activity is directly linked to biological activity of IAA-conjugates.⁷ The gene expression of ILL1, ILL2, IAR3, and ILR1 has been examined in Northern and promoter- β -glucuronidase assays.⁷ ILL2 and ILR1 showed overlapping but distinct pattern of expression in *A. thaliana* seedlings and adult plants. Interestingly, unlike the triple *ilr1-iar3-ill2* mutant, single *ill2* mutation had only limited effect on the inhibition of the root elongation in the presence of various IAA-conjugates suggesting redundancy of function with the other two *A. thaliana* amidohydrolase isoforms.⁷

Here we report the X-ray crystal structure of *A. thaliana* ILL2 (UniProt code ILL2_ARATH18) at 2.0 Å resolution and discuss the structural basis for the substrate specificity of this enzyme. The structure of ILL2 represents a sequence-to-structure target with <30% sequence identity to any structure in Protein Data Bank19 at the time of selection and deposition determined under the National Institutes of Health Protein Structure Initiative.

MATERIALS AND METHODS

Prediction of Signal Peptide Residues

The ILL2 amino acid sequence encoded by At5g56660.1 was analyzed with web-based programs to predict subcellular localization. Secretory pathway targeting was predicted by TargetP20 (<http://www.cbs.dtu.dk/services/TargetP/>) and confirmed with iPSORT21 (<http://hc.ims.u-tokyo.ac.jp/iPSORT/>) and SignalP12 (<http://www.cbs.dtu.dk/services/SignalP/>). The predictions for ILL2 and homologous amidohydrolases from *A. thaliana* indicated a secretory pathway localization as well as a highly reliable cleavage score (>0.8 with TargetP and SignalP) for all members except ILL5_ARATH. The putative signal peptide of ILL2 corresponded to amino terminal residues 1–22 based on these analyses.

Expression and Purification of ILL2

The gene At5g56660.1 encoding mature ILL2 (residues 23–439) was cloned and both the unlabeled and selenomethionine (SeMet)-labeled proteins were expressed and purified following the standard Center for Eukaryotic Structural Genomics pipeline protocols for cloning²², protein expression²³, protein purification²⁴, and overall bioinformatics management²⁵. In short, the cDNA encoding ILL2 was cloned into a pDONR221 plasmid (Invitrogen, Carlsbad, CA) and for expression purposes transferred into a pVP13 plasmid, a custom vector derived from pQE80 (Qiagen, Valencia, CA). The predicted signal peptide spanning residues 1–22 of ILL2 was eliminated at the cloning stage. Proteins were expressed in *Escherichia coli* B834 p(lacI+RARE) cells using 2 L of self-inducing media. Upon sonication of the harvested cells, the protein in the supernatant was purified via immobilized nickel affinity chromatography, and TEV protease was used to cleave the affinity/solubility tag consisting of His⁶-maltose binding protein. After tag capture by subtractive nickel affinity chromatography, and a final desalting step, the proteins were concentrated to 10 mg/mL and dialyzed against the protein buffer (50 mM NaCl, 5 mM TRIS pH 8.0). Protein aliquots were then drop frozen in liquid nitrogen and stored at 193 K. Protein purifications resulted in 9.0 mg of native and 7.5 mg of SeMet-labeled recombinant proteins.

Crystallization and Structure Solution of ILL2

Crystals of ILL2 were grown at 293 K by the hanging drop method from a 10 mg.mL⁻¹ protein solution in the protein buffer mixed with an equal amount of well solution containing 14% (w/v) polyethylene glycol 1500 (PEG 1.5K), 80 mM MgSO₄, 100 mM *N*-cyclohexyl-2-aminoethanesulfonic acid (CHES) pH 9.0. Crystals were cryoprotected at 293 K by soaking in well solutions containing 7%, 14%, and 20% (v/v) ethylene glycol and were flash-frozen in a stream of cryogenic nitrogen gas at 100 K. X-ray diffraction data for phasing were collected near the selenium K absorption edge (12,660 eV) from SeMet-labeled ILL2 crystals at Structural Biology Center (SBC-CAT) 19-BM beamline at the Advanced Photon Source at Argonne National Laboratory. X-ray diffraction data used for refinement were collected at 12,665 eV from native ILL2 crystals at the same beamline. The diffraction images were integrated and scaled using HKL2000.²⁶ The structure was automatically phased and density-modified using Solve/Resolve using single wavelength

anomalous diffraction methodology.^{27,28} The initial model was built using Arp/Warp.²⁹ The structure was completed with multiple cycles of iterative manual building in Xfit³⁰ and refinement in REFMAC⁵³¹. All refinement steps were monitored using an R_{free} value based on 5.1% independent reflections. The stereochemical quality of the final model was assessed using MOLPROBITY.³² The substrate docking to the active site was performed using Coot³³ and the PyMOL³⁴ sculpting module. The figures were prepared using PyMOL.

RESULTS AND DISCUSSION

Structure quality

The three-dimensional crystal structure of ILL2 was phased with a single wavelength anomalous diffraction method^{35,36} using a 2.0 Å diffraction dataset collected near the selenium K-edge from selenomethionyl crystals of ILL2. The structure was refined against a 2.0 Å dataset collected from the native ILL2 crystal. The final model has an R factor of 15.7% and an R_{free} of 20.4%. Detailed data collection, phasing, and refinement statistics are summarized in Table I. ILL2 crystallized in space group $P3_221$ with one monomer per asymmetric unit. The final model consists of residues 37–228, 238–274, 283–428 and additional 285 water molecules. Residues 1–36, 229–237, 275–282, and 429–439 of ILL2 were not modeled, as they were either not present (residues 1–22 of the signal peptide) or not resolved in the electron density map. Ile92 adopts an unusual backbone conformation ($\phi = 80.2^\circ, \psi = -63.1^\circ$) in what has traditionally been considered to be a forbidden region of the Ramachandran plot. Inspection of the electron density map for this residue did not reveal any ambiguity in the modeling of this residue (Figure 1a). In fact, while the adopted conformation is unusual it is allowed based on the up-to-date *MOLPROBITY* backbone geometry validation criteria.³²

The X-ray structure of ILL2

The structure of ILL2 consists of two distinct domains. The larger catalytic domain (residues 1–217 & 333–439) consists of a central 8-stranded β -sheet sandwiched between five and two α -helices on either side (Figure 1b, red and cyan). Based on the CATH classification³⁷, the catalytic domain adopts a 3-layer($\alpha\beta\alpha$) sandwich architecture with aminopeptidase topology found in bacterial dinuclear zinc exopeptidases. Two short β -strands connect the catalytic domain to a satellite domain, which spans the residues 218–332. The satellite domain is inserted between β -strand 6 and helix E of the catalytic domain. It adopts a ferredoxin-like fold with 2-layer($\alpha\beta$) sandwich architecture and alpha-beta plaits topology. The domain consists of a 4-stranded antiparallel β -sheet flanked on one side by two α -helices (Figure 1b, green and yellow). The satellite domain serves as a dimerization module in several structural homologs of ILL2 from bacteria.^{38,39} Analysis of the crystal packing of the ILL2 crystals did not, however, reveal a dimer formation.

Structural comparison of ILL2 and bacterial exopeptidases

According to the VAST server⁴⁰, the structure of ILL2 shows a close overall similarity to several members of the M20 family of peptidases, including *Bacillus subtilis* yxP protein (PDB ID 1ysj)⁴¹, *Pseudomonas* sp. carboxypeptidase G2 (CPG2, PDB ID 1cg2)³⁹, and *Salmonella typhimurium* peptidase T (PDB ID 1fno)³⁸. The closest structural homolog of ILL2, with a VAST score of 35.8, is an uncharacterized yxP protein solved by the Midwest Center for Structural Genomics.⁴¹ The structural alignment of these two proteins resulted in a close pairing of 345 residues with a root mean square deviation (rmsd) of 1.6 Å and 36% sequence identity. There are several minor differences in loop conformations between the proteins. Two notable insertions exist in ILL2 spanning residues 387–393 of the catalytic domain and residues 325–330 of the satellite domain. Importantly, the proteins share a structurally conserved active site (Figure 2a).

The second closest structural homolog of ILL2, with a VAST score of 27.5, is *Pseudomonas* sp. CPG2.39 The proteins structurally align for 346 residues with an rmsd of 3.5 Å and 15.6% sequence identity. The overall fold of CPG2 is of the same topology as that of ILL2. However, while the structures are similar the structural homology is clearly more distant than that in the case of ILL2 and yxeP (compare divergence of protein backbone traces in Figures 2a and 2b). Significant structural differences between ILL2 and CPG2 exist in multiple surface loops. Also, coordination of the dimetal center by CPG2 differs from that seen in yxeP and ILL2 as detailed below.

The active site of ILL2

Structural homologs of ILL2 bind two metal ions in the active site. Available biochemical evidence suggests that hydrolysis of IAA-conjugates by ILL2 and other *A. thaliana* amidohydrolases is also metal-dependent. It has been established that these enzymes prefer Mn^{2+} as a cofactor and ILR1 could also utilize Cu^{2+} .¹⁰ The crystallization conditions used to grow crystals of ILL2 did not include metal ions. The resulting apoILL2 structure revealed two water molecules in the active site. However, as shown in Figure 2a, the active sites of apoILL2 and yxeP, which possesses a dimetal center, are closely similar. The catalytic domains of these proteins superpose with an all-atom rmsd of 0.5 Å. Therefore, we have identified five residues likely involved in metal coordination in ILL2, namely Cys137, Glu173, and His397 that coordinate metal 1, and Cys137, His139, and His197 that coordinate metal 2. In addition, a bridging water molecule likely completes the coordination of the metals; however, this feature was not unambiguously resolved in yxeP due to limited resolution. As expected, all the coordinating residues are fully conserved in all known *A. thaliana* amidohydrolases as well as other related proteins (Figure 3). It has been reported that mutations of His405 of bacterial IAA-Asp hydrolase, which corresponds to His397 of ILL2, lead to a complete loss of the enzymatic activity.⁴² This observation is consistent with a crucial role of this histidine in coordinating a metal ion in the active site. Several additional loss-of-function mutations have been described in ILL2 homologs from *A. thaliana* (Table II). Mutations E69K and G139D in ILR1 likely compromise the geometry of the dimetal binding site. We note there are slight differences in conformations of putative ligand residues between yxeP and apoILL2, presumably due to a lack of stabilizing interactions with metal ions in ILL2. The differences include alternate conformations of Cys137 and a minor displacement and rotation of His397 (Figure 2a). Importantly, Cys98 of yxeP, a residue analogous to Cys137 of ILL2, coordinates both metals in the active site of yxeP. The coordination of metals by this bridging cysteine residue seems to be a unique feature of ILL2 and yxeP, both members of the peptidase M20D family, that clearly distinguishes these enzymes from other structurally characterized enzymes from the M20 family, in which a conserved aspartate coordinates metals in a bidentate mode (Asp141 in CPG2, Figure 2b). We queried sequence databases by the FFAS03 server⁴³ (using PSI-BLAST NR85S multiple sequence alignment option) and found over 700 bacterial and plant proteins that show >30% sequence identity to ILL2 and contain cysteine at the position corresponding to Cys137 of ILL2. A large portion of identified proteins is currently functionally unannotated. Some of the annotated proteins from the list include hippurate hydrolases, a range of *N*-acyl-L-amino acid amidohydrolases, and a single *N*-acetyldiaminopimelate deacetylase. There are additional differences in the metal coordination between ILL2 and CPG2 (taken here as an representative of M20 family of peptidases). First, Glu200 in CPG2 replaces His197 in ILL2. Second, “Metal 2” is in CPG2 coordinated by His112, which originates from a different secondary structure element than its functional equivalent His139 of ILL2 (Figure 2b, double-arrow). His112 of CPG2 is the carboxy terminal residue of the third β-strand, while His139 of ILL2 marks the beginning of the helix C of the catalytic domain. In summary, it is clear that the proteins from the M20D

peptidase family can be distinguished from the other M20 peptidases based on interesting remodeling of the active site that changed its preference from zinc(II) to manganese(II).

The substrate binding model

In order to understand how ILL2 binds and selectively recognizes its preferred substrate, IAA-Ala, we modeled this compound into the active site of ILL2. We based the initial placement and conformation of IAA-Ala on the structure of transition-state-like inhibitor (S)-2-(4-iodobenzylphosphonomethyl)-pentanedioic acid (GPI-18431) bound to the active site of glutamate carboxypeptidase II (GCPII, PDB ID 2c6c).⁴⁴ The catalytic domains of ILL2 and GCPII share a similar fold (with a VAST score of 16.7, rmsd of 2.9 Å for 221 aligned residues with 13.6% sequence identity) as well as the active site with a dimetal center. Both the substrate of ILL2, IAA-Ala, and the transition-state-like inhibitor of GCPII, GPI-18431, have similar size and share some common structural features, including an amino terminal aromatic group and free carboxylate group at the carboxy terminus. The carbonyl oxygen of the scissile amide bond of IAA-Ala was positioned and constrained to reside within 2 Å from Metal 1. In addition, the dimetal center of ILL2, which was modeled based on that of yxeP, was restrained. Two conformational modifications were necessary to position IAA-Ala into the active site cavity of ILL2. First, a sizeable hydrophobic cavity presented itself for the indole ring IAA-Ala (Figure 4a and 4b). The indole ring was therefore steered into this cavity. Second, the sidechain of Ser367 of ILL2 clashed with the terminal carboxylate of IAA-Ala. Inspection of the electron density map of apo-ILL2 revealed that the rotamer of Ser367 was not well defined and a different rotamer could relieve the clash with the modeled substrate. In fact, the new rotamer Ser367 may be stabilized by hydrogen bond(s) to Arg293. To reach the optimal conformation, IAA-Ala as well as protein residues and water molecules in 8 Å radius from the carbonyl oxygen of IAA-Ala were allowed to relax under van der Waals constraints. The resulting model of the active site with the bound substrate is presented in Figure 4a. Several key insights can be made from this model.

First, the indole ring of IAA-Ala is likely stabilized by binding within a hydrophobic cavity formed by residues Leu215, Val364, Leu383, Phe381, and Phe214. These residues remain hydrophobic in *A. thaliana* amidopeptidases, with the exception of Glu363 of ILL1 that corresponds to Val364 of ILL2 (Figure 3).

Second, Glu172 is well positioned to act as a general acid-base residue during the catalysis. This residue can act as a general base to facilitate deprotonation of a water molecule coordinated to Metal 2 and thus assist in a nucleophilic attack on the amide carbonyl of IAA-amino acid conjugate. The resulting negatively charged tetrahedral intermediate is likely stabilized by electrostatic interaction with manganese(II) dimetal center. Next, the protonated Glu172 can act as a general acid donating a proton to the amide nitrogen of IAA-amino acid conjugate in order to facilitate the fissure of the scissile bond by forming a better leaving group. Glu172 is strictly conserved in all *A. thaliana* amidohydrolases and their homologs from the M20D peptidase family. Importantly, it is homologous to glutamates implicated as general acid-base residues in other M20 peptidases (for example Glu175 of CPG2, or Glu424 of GCPII). The structural similarity of ILL2 and the enzymes of the M20 peptidase family as well as conservation of the key catalytic residue and the similar chemical nature of substrates (e.g. being amides) of these enzymes suggest that ILL2 uses a catalytic mechanism similar to that proposed for the peptidases from M20 family. This mechanism is in turn a variation of a mechanism established for prototypical zinc proteases such as carboxypeptidase A.⁴⁵⁻⁴⁶

Residues Glu368, Arg110, Gln169, and Asp369 are highly conserved among proteins that show >30% identity to ILL2. This set of residues seems to be involved in a hydrogen bond

network nearby the active site next to the putative general acid-base residue Glu172. Glu368 is found in a salt bridge with the buried Arg110 and is therefore likely present in its basic form. We speculate that a function of Glu368 and Asp369 may be to increase the pKa value of Glu172.

Third, Leu175 seems to be responsible for selectivity against IAA-amino acid conjugates with amino acid side chains bulkier than alanine or serine (Figure 4b). In fact, Leu175 is not conserved in *A. thaliana* amidohydrolases, it is only present in ILL2 and its closest homolog ILL1. In ILR1, this residue is replaced by Tyr176, which could stabilize the aromatic side chains of the preferred substrates of this amidohydrolase isoform (IAA-Phe or IAA-Tyr) by a stacking interaction. The remaining amidohydrolases from *A. thaliana* contain glycine in the structurally equivalent position. In the case of CPG2, Arg324 has been suggested to be responsible for recognition and stabilization of the glutamate side chain of the peptide substrates.⁴⁴ When the active sites of CPG2 and ILL2 were overlaid and compared, the guanidyl moiety of Arg324 of CPG2 was found about 3 Å (in the direction away from the active site) from the side chain of Leu175 of ILL2, further supporting the hypothesis about the role Leu175. The wheat IAR3 homolog (*TaIAR3*, UniProt Code Q66VR4_WHEAT) has been shown to be specific for N-indole-3-butyryl-alanine or -glycine and N-indole-3-propionyl-alanine.⁴⁷ Judging from the multiple sequence alignment in Figure 3, all important residues of the active site remain conserved in *TaIAR3*. Glycine 168 of *TaIAR3* corresponds to the putative selectivity filter residue Leu175 of ILL2. The only other notable difference is a single residue insertion (Thr375 of *TaIAR3*) located in the vicinity of the residues forming the hydrophobic cavity for the indole ring. It is possible that these two modifications within the active site of *TaIAR3* contribute to the ability this enzyme to productively bind and hydrolyze auxin derivatives with longer fatty acid chains.

Fourth, the terminal carboxyl group of IAA-Ala is likely stabilized by a pair of hydrogen bonds to the main chain of Ser367. However, based on our model this group is not recognized by additional protein residues. This is in a contrast to GCPII, where the analogous group is stabilized by electrostatic interaction with Arg210.⁴⁴ The only basic residue in the vicinity of the active site of ILL2 is Arg293. This residue is strictly conserved in all *A. thaliana* amidopeptidases and also in other members of the M20 peptidase family (Arg288 of CPG2, and Arg280 of carboxypeptidase T) suggesting an important role for this residue. However, the distance between the guanidinium moiety of Arg293 and the terminal carboxyl of modeled IAA-Ala is about 7 Å. Therefore, the enzyme would have to undergo a significant conformational change upon substrate binding to allow for stabilization of the terminal carboxyl of IAA-Ala. Instead, it seems more likely that Arg293 stabilizes the overall active site architecture by hydrogen bonding to the mainchain carbonyls of Gly366 and Ala216.

Accession Numbers

The atomic coordinates and structure factors for ILL2 have been deposited in the Protein Data Bank¹⁹ under accession code 1xmb.

Acknowledgments

We acknowledge the financial support from NIH National Institute for General Medical Sciences grant P50 GM64598 and U54 GM074901 (J. L. Markley, PI) and U.S. Department of Energy grant no. DE-FG02-00ER150065 to R.S.B. N.L.H was supported by a graduate fellowship from NSF-IGERT grant no. 9987555. Use of the Argonne National Laboratory Structural Biology Center beamlines at the Advanced Photon Source, was supported by the U. S. Department of Energy, Office of Energy Research, under Contract No. W-31-109-ENG-38. Special thanks go to all members of the CESG team involved in this work, especially Russell L. Wrobel, Ronnie O. Frederick, Craig S. Newman, Eric A. Steffen, Todd L. Kimball, Sandy Thao, John P. Kunert, Brendan T. Burns,

Janet E. McCombs, David J. Aceti, Frank C. Vojtik, Andrew C. Olson, Megan A. Ritters, Hassan K. Sreenath, and David W. Smith.

References

1. Woodward AW, Bartel B. Auxin: regulation, action, and interaction. *Annals of Botany*. 2005; 95(5): 707–735. [PubMed: 15749753]
2. Bartel B. Auxin Biosynthesis. *Annu Rev Plant Physiol Plant Mol Biol*. 1997; 48:51–66. [PubMed: 15012256]
3. Ljung K, Hull AK, Kowalczyk M, Marchant A, Celenza J, Cohen JD, Sandberg G. Biosynthesis, conjugation, catabolism and homeostasis of indole-3-acetic acid in *Arabidopsis thaliana*. *Plant Mol Biol*. 2002; 49(3–4):249–272. [PubMed: 12036253]
4. Kowalczyk M, Sandberg G. Quantitative analysis of indole-3-acetic acid metabolites in *Arabidopsis*. *Plant Physiology*. 2001; 127(4):1845–1853. [PubMed: 11743128]
5. Tan X, Calderon-Villalobos LI, Sharon M, Zheng C, Robinson CV, Estelle M, Zheng N. Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature*. 2007; 446(7136):640–645. [PubMed: 17410169]
6. Tuominen H, Ostin A, Sandberg G, Sundberg B. A Novel Metabolic Pathway for Indole-3-Acetic Acid in Apical Shoots of *Populus tremula* (L.) x *Populus tremuloides* (Michx). *Plant Physiology*. 1994; 106(4):1511–1520. [PubMed: 12232425]
7. Rampey RA, LeClere S, Kowalczyk M, Ljung K, Sandberg G, Bartel B. A family of auxin-conjugate hydrolases that contributes to free indole-3-acetic acid levels during *Arabidopsis* germination. *Plant Physiology*. 2004; 135(2):978–988. [PubMed: 15155875]
8. Bartel B, Fink GR. ILR1, an amidohydrolase that releases active indole-3-acetic acid from conjugates. *Science* (New York, NY. 1995; 268(5218):1745–1748.
9. Davies RT, Goetz DH, Lasswell J, Anderson MN, Bartel B. IAR3 encodes an auxin conjugate hydrolase from *Arabidopsis*. *The Plant Cell*. 1999; 11(3):365–376. [PubMed: 10072397]
10. LeClere S, Tellez R, Rampey RA, Matsuda SP, Bartel B. Characterization of a family of IAA-amino acid conjugate hydrolases from *Arabidopsis*. *The Journal of Biological Chemistry*. 2002; 277(23):20446–20452. [PubMed: 11923288]
11. Campanella JJ, Larko D, Smalley J. A molecular phylogenomic analysis of the ILR1-like family of IAA amidohydrolase genes. *Comparative and Functional Genomics*. 2003; 4(6):584–600. [PubMed: 18629030]
12. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering*. 1997; 10(1):1–6. [PubMed: 9051728]
13. Pelham HRB. THE RETENTION SIGNAL FOR SOLUBLE-PROTEINS OF THE ENDOPLASMIC-RETICULUM. *Trends in Biochemical Sciences*. 1990; 15(12):483–486. [PubMed: 2077689]
14. Derkx PM, Madrid SM. The foldase CYPB is a component of the secretory pathway of *Aspergillus niger* and contains the endoplasmic reticulum retention signal HEEL. *Mol Genet Genomics*. 2001; 266(4):537–545. [PubMed: 11810224]
15. Raykhel I, Alanen H, Salo K, Jurvansuu J, Nguyen VD, Latva-Ranta M, Ruddock L. A molecular specificity code for the three mammalian KDEL receptors. *The Journal of Cell Biology*. 2007; 179(6):1193–1204. [PubMed: 18086916]
16. Bartel B, LeClere S, Magidin M, Zolman BK. Inputs to the active indole-3-acetic acid pool: De novo synthesis, conjugate hydrolysis, and indole-3-butyric acid beta-oxidation. *Journal of Plant Growth Regulation*. 2001; 20(3):198–216.
17. Titarenko E, Rojo E, Leon J, Sanchez-Serrano JJ. Jasmonic acid-dependent and -independent signaling pathways control wound-induced gene activation in *Arabidopsis thaliana*. *Plant Physiology*. 1997; 115(2):817–826. [PubMed: 9342878]
18. Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N, Suzek B.

- The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.* 2006; 34(Database issue):D187–191. [PubMed: 16381842]
19. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res.* 2000; 28(1):235–242. [PubMed: 10592235]
 20. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology.* 2000; 300(4):1005–1016. [PubMed: 10891285]
 21. Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S. Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics (Oxford, England).* 2002; 18(2):298–305.
 22. Thao S, Zhao Q, Kimball T, Steffen E, Blommel PG, Ritters M, Newman CS, Fox BG, Wrobel RL. Results from high-throughput DNA cloning of *Arabidopsis thaliana* target genes using site-specific recombination. *J Struct Funct Genomics.* 2004; 5(4):267–276. [PubMed: 15750721]
 23. Sreenath HK, Bingman CA, Buchan BW, Seder KD, Burns BT, Geetha HV, Jeon WB, Vojtik FC, Aceti DJ, Frederick RO, Phillips GN Jr, Fox BG. Protocols for production of selenomethionine-labeled proteins in 2-L polyethylene terephthalate bottles using auto-induction medium. *Protein Expr Purif.* 2005; 40(2):256–267. [PubMed: 15766867]
 24. Jeon WB, Aceti DJ, Bingman CA, Vojtik FC, Olson AC, Ellefson JM, McCombs JE, Sreenath HK, Blommel PG, Seder KD, Burns BT, Geetha HV, Harms AC, Sabat G, Sussman MR, Fox BG, Phillips GN Jr. High-throughput Purification and Quality Assurance of *Arabidopsis thaliana* Proteins for Eukaryotic Structural Genomics. *J Struct Funct Genomics.* 2005; 6(2-3):143–147. [PubMed: 16211511]
 25. Zolnai Z, Lee PT, Li J, Chapman MR, Newman CS, Phillips GN Jr, Rayment I, Ulrich EL, Volkman BF, Markley JL. Project management system for structural and functional proteomics: Sesame. *J Struct Funct Genomics.* 2003; 4(1):11–23. [PubMed: 12943363]
 26. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol.* 1997; 276:307–326.
 27. Terwilliger TC. Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr.* 2000; 56(Pt 8):965–972. [PubMed: 10944333]
 28. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. *Acta Crystallographica Section D-Biological Crystallography.* 1999; 55:849–861.
 29. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. *Nat Struct Biol.* 1999; 6(5):458–463. [PubMed: 10331874]
 30. McRee DE. XtalView/Xfit—A versatile program for manipulating atomic coordinates and electron density. *J Struct Biol.* 1999; 125(2–3):156–165. [PubMed: 10222271]
 31. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr.* 1997; 53(Pt 3):240–255. [PubMed: 15299926]
 32. Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins.* 2003; 50(3):437–450. [PubMed: 12557186]
 33. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* 2004; 60:2126–2132. [PubMed: 15572765]
 34. DeLano, WL. The PYMOL Molecular Graphic System. San Carlos, CA, USA: DeLano Scientific LLC; 2002.
 35. Rice LM, Earnest TN, Brunger AT. Single-wavelength anomalous diffraction phasing revisited. *Acta Crystallogr D Biol Crystallogr.* 2000; 56(Pt 11):1413–1420. [PubMed: 11053839]
 36. Wang BC. Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* 1985; 115:90–112. [PubMed: 4079800]
 37. Pearl FM, Bennett CF, Bray JE, Harrison AP, Martin N, Shepherd A, Sillitoe I, Thornton J, Orengo CA. The CATH database: an extended protein family resource for structural and functional genomics. *Nucleic Acids Res.* 2003; 31(1):452–455. [PubMed: 12520050]
 38. Hakansson K, Miller CG. Structure of peptidase T from *Salmonella typhimurium*. *European Journal of Biochemistry/FEBS.* 2002; 269(2):443–450. [PubMed: 11856302]

39. Rowsell S, Paupit RA, Tucker AD, Melton RG, Blow DM, Brick P. Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy. *Structure*. 1997; 5(3):337–347. [PubMed: 9083113]
40. Madej T, Gibrat JF, Bryant SH. Threading a database of protein cores. *Proteins*. 1995; 23(3):356–369. [PubMed: 8710828]
41. Minasov G, Shuvalova L, Brunzelle JS, Collart FR, Anderson WF. Structure of Bacillus Subtilis YXEP Protein, a Dinuclear Metal Binding Peptidase from M20 Family. Protein Data Bank (www.rcsb.org/pdb/home). 2005 1ysj. 10.2210/pdb2211ysj/pdb
42. Chou JC, Welch WH, Cohen JD. His-404 and His-405 are essential for enzyme catalytic activities of a bacterial indole-3-acetyl-L-aspartic acid hydrolase. *Plant & Cell Physiology*. 2004; 45(9): 1335–1341. [PubMed: 15509859]
43. Jaroszewski L, Rychlewski L, Li Z, Li W, Godzik A. FFAS03: a server for profile–profile sequence alignments. *Nucleic Acids Res*. 2005; 33(Web Server issue):W284–288. [PubMed: 15980471]
44. Mesters JR, Barinka C, Li W, Tsukamoto T, Majer P, Slusher BS, Konvalinka J, Hilgenfeld R. Structure of glutamate carboxypeptidase II, a drug target in neuronal damage and prostate cancer. *EMBO J*. 2006; 25(6):1375–1384. [PubMed: 16467855]
45. Christianson DW, Lipscomb WN. Carboxypeptidase-A. *Accounts Chem Res*. 1989; 22(2):62–69.
46. Lipscomb WN. Carboxypeptidase A mechanisms. *Proc Natl Acad Sci U S A*. 1980; 77(7):3875–3878. [PubMed: 6933442]
47. Campanella JJ, Olajide AF, Magnus V, Ludwig-Muller J. A novel auxin conjugate hydrolase from wheat with substrate specificity for longer side-chain auxin amide conjugates. *Plant Physiology*. 2004; 135(4):2230–2240. [PubMed: 15299127]

Figure 1.

The X-ray crystal structure of ILL2. (A) Stereo view of a $2mF_o-DF_c$ electron density map of the refined ILL2 model around Ile92, which adopts an unusual backbone conformation. The map is contoured at 1.3σ level. (B) Ribbon diagram of ILL2 in stereo representation shows that this enzyme consists of two distinct domains. Strands and helices in both the catalytic domain (bottom, red and cyan) and the satellite domain (top, green and yellow) are numbered in order of their appearance in respective domains. The catalytic domain of ILL2 consists of the central 8-stranded β -sheet (red) sandwiched between five and two helices on either side (cyan). The satellite domain (top) is inserted between β -strand 6 and helix E of the catalytic domain.

Figure 2.

Comparison of the active sites of ILL2 and its structural homologs. (A) A stereo view the ribbon trace of superimposed ILL2 (red; PDB ID 1xmb) and yxeP from *B. subtilis* (cyan; PDB ID 1ysj). Catalytic domains of ILL2 and yxeP are highly similar and superimpose with an all-atom rmsd of 0.5 Å. The residues involved in coordination of Metal 1 and Metal 2 (green spheres) in yxeP are shown as cyan sticks. The analogous residues present in the apo-ILL2 structure are also shown (red sticks) and annotated by labels. (B) A stereo view of the ribbon trace of superimposed ILL2 (red) and CPG2 (cyan, PDB ID 1cg2). The figure is given in the same orientation as in Figure 2a to allow for easy comparison of all structures. The residues involved in metal coordination are shown in cyan sticks and are annotated for CPG2. Asp141 of CPG2 serves as a bidentate ligand in metal coordination, replacing Cys137 of ILL2. The double arrow points to the functionally related metal ligand that originates from topologically distinct positions in ILL2 (His139) and CPG2 (His112).

Figure 3.

The multiple sequence alignment of selected ILL2 homologs. The multiple sequence alignment of all known IAA-amino acid hydrolases from *Arabidopsis thaliana* (first seven entries; UniProt codes ILR1_ARATH through ILL6_ARATH), IAR3 homolog from wheat (UniProt code Q66VR4_WHEAT), and several bacterial ILL2 homologs including yxP protein from *Bacillus subtilis*, *N*-acyl-L-amino acid amidohydrolase from *Bacillus clausii* (gi|56964568), amino acid amidohydrolase from *Lactobacillus acidophilus* (gi|58337160), *N*-acetyldiaminopimelate deacetylase from *Lactobacillus salivarius* (gi|90961445), hippurate hydrolase from *Geobacillus kaustophilus* (gi|56419585) and hippurate hydrolase from *Bacillus halodurans* (gi|15615231). The ILL2 sequence is shown in bold letters and accompanied by a ruler indicating every 5th (:) and 10th (|) residue. Invariant residues of all the aligned sequences are highlighted in red, other highly conserved residues are colored in blue. The important active site residues are boxed and color-coded as in Figure 4a with metal-coordinating residues in cyan, hydrophobic residues that bind the indole ring of a substrate in orange, a putative general acid-base residue in green, a putative amino acid side chain selectivity filter residue in magenta, and other conserved residues in the vicinity of the active site in gray. Secondary structural elements of ILL2 are shown as cylinders (helices) and arrows (β -strands). The numbering and color-coding scheme of the secondary structural elements corresponds to that in Figure 1b. Purple italics letters highlight the predicted signal peptide sequences of *A. thaliana* amidohydrolases. Bold green letters highlight residues mutated in loss-of-function mutants of *A. thaliana* ILR1 and IAR3.8-9. Further information about these mutations is given in Table II.

Figure 4.

The active site of ILL2 with a docked substrate. (A) A stereo view of the active site of ILL2 with a modeled substrate *N*-(indole-3-acetyl)-alanine (IAA-Ala; yellow sticks). Important residues of the active site of ILL2 are shown as sticks and are color-coded as follows: metal-coordinating residues (cyan), residues forming a hydrophobic cavity that may bind the indole ring of a substrate (orange), a putative general acid-base residue Glu172 (green) and a putative side-chain selectivity filter residue Leu175 (magenta), other conserved residues (gray) involved in formation of the active site hydrogen bond network. (B) A stereo view of a surface representation of the ILL2 active site with the docked IAA-Ala (yellow sticks). The hydrophobic cavity (orange) accommodates the indole ring of a substrate. The carbonyl group of a scissile peptide bond of the substrate is coordinated by one of the metals in the active site (green spheres) while the C-terminal carboxyl of the substrate points toward the surface of the protein. Leucine 175 (magenta) restricts the size of the amino acid side chain of the substrate that can fit into the active site, providing the basis for the substrate specificity of ILL2.

Table I

Crystal parameters, X-ray data collection, phasing and refinement statistics.

	Refinement-Peak	Phasing-Peak
Crystal parameters		
Space group	P3 ₁ 21	P3 ₁ 21
Unit-cell parameters (Å)	a = b = 75.3 c = 130.9	a = b = 75.4 c = 130.9
Data collection statistics		
Wavelength (Å)	0.97895	0.97932
Energy (eV)	12,665	12,660
Resolution range (Å)	24.21 – 2.00 (2.07–2.00)	24.72 – 2.00 (2.07–2.00)
No. of reflections (measured/unique) ^a	247784/29672	423602/9765
Completeness (%)	99.9 (99.9)	99.9 (100.0)
R _{merge} ^d	0.043 (0.334)	0.097 (0.515)
Redundancy	8.4 (6.3)	14.2 (12.9)
Mean I/sigma(I)	22.0 (3.8)	15.4 (4.3)
Phasing statistics^c		
Mean FOM (centric/acentric)		0.096/0.430
Phasing Power (anomalous)		1.86
Cullis R-factor (anomalous)		0.62
Refinement and model statistics		
Resolution range ^d	24.21–2.00 (2.05–2.00)	
No. of reflections (work/test)	28126/1504	
R _{cryst} ^e	0.157 (0.186)	
R _{free} ^f	0.204 (0.247)	
Rmsd bonds (Å)	0.018	
Rmsd angles (°)	1.587	
ESU from R _{free} (Å)	0.140	
B factor – Wilson plot (Å ²)	28.3	
B factor – overall/protein/waters (Å ²)	31.8/31.0/40.3	
No. of protein molecules/all atoms	1/2895	
No. of waters	285	
Ramachandran Plot by MOLPROBITY (%)		
Favored regions	98.6	
Additional allowed regions	1.4	
Outliers	0.0	
PBD code	1xmb	

^aValues in parentheses are for the highest resolution shell.^b $R_{merge} = \sum_h \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_h \sum_i I_i(h)$, where $I_i(h)$ is the intensity of an individual measurement of the reflection and $\langle I(h) \rangle$ is the mean intensity of the reflection.

^cResolution range for phasing in SHARP was (24.72–2.00) Å.

^dResolution range for refinement was cut (40.5–2.6) Å due to low completeness and signal in the remaining resolution shells.

^e $R_{cryst} = \sum_h |F_{obs}| - |F_{calc}| / \sum_h |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes, respectively.

^f R_{free} was calculated as R_{cryst} using ~5.1% of the randomly selected unique reflections that were omitted from structure refinement.

Table IILoss of function mutations in *A. thaliana* ILL2 homologs.

Mutation	Corresponding residue in ILL2	Is highly conserved residue?	Distance between Ca and metal 1 [Å]	Structural hypothesis about of origins of the defect
E69K ^a	E68	yes	15.0	Conformation of the loop harboring general base Glu172 and metal ligand Glu173 will be altered due to loss of stabilizing hydrogen-bond. Possible salt bridge to Asp114 could further disturb local geometry next to the active site.
G139D ^a	G138	yes	8.0	Steric clash resulting from the mutation will directly compromise positioning of metal coordinating residues Cys137 and His139.
G415E ^a	G411	no	18.2	Mutation places bulky, charged residue into a hydrophobic core of the protein formed by Met145, Ile196, Leu198, Leu384, Val407, and Ile414. Protein likely misfolds.
S206L ^b	S209	no	18.5	Mutation results in a loss of hydrogen bonds to His380 and fully conserved Asn337. Bulkier, hydrophobic residue at this position may disturbs Phe381 which is involved in formation of, the indole-binding hydrophobic pocket.
G224E ^b	G227	yes	41.6	Not clear. Residue is located at the tip of the satellite domain.
G226E ^b	G229	no		Not clear, residue not resolved in the structure of ILL2. Residue is likely located at the tip of the satellite domain.
A230T ^b	A233	no		Not clear, residue not resolved in the structure of ILL2. Residue is likely located at the tip of the satellite domain.

^a Mutation identified in *A.thaliana* ILR1.8^b Mutation identified in *A.thaliana* IAR3.9