Published in final edited form as:

J Struct Funct Genomics. 2012 March ; 13(1): 27–36. doi:10.1007/s10969-012-9127-5.

Crystal structure of a putative isochorismatase hydrolase from *Oleispira antarctica*

Anna M. Goral^{1,3,4,§}, Karolina L. Tkaczuk^{1,3,§}, Maksymilian Chruszcz^{1,3}, Olga Kagan^{2,3}, Alexei Savchenko^{2,3}, and Wladek Minor^{1,3,*}

¹Department of Molecular Physiology and Biological Physics, University of Virginia, 1340 Jefferson Park Avenue, Charlottesville, VA 22908, USA ²Banting and Best Department of Medical Research, 112 College Street, University of Toronto, Toronto, Ontario M5G 1L6, Canada ³Midwest Center for Structural Genomics ⁴Inter-Faculty Interdisciplinary Doctoral Studies in Natural Sciences and Mathematics, University of Warsaw, ul. Żwirki i Wigury 93, 02-089 Warsaw, Poland

Abstract

Isochorismatase-like hydrolases (IHL) constitute a large family of enzymes divided into five structural families (by SCOP). IHLs are crucial for siderophore-mediated ferric iron acquisition by cells. Knowledge of the structural characteristics of these molecules will enhance the understanding of the molecular basis of iron transport, and perhaps resolve which of the mechanisms previously proposed in the literature is the correct one.

We determined the crystal structure of the *apo*-form of a putative isochorismatase hydrolase OaIHL (PDB code: 3LQY) from the antarctic γ -proteobacterium *Oleispira antarctica*, and did comparative sequential and structural analysis of its closest homologs. The characteristic features of all analyzed structures were identified and discussed. We also docked isochorismate to the solved crystal structure by *in silico* methods, to highlight the interactions of the active center with the substrate.

The putative isochorismate hydrolase OaIHL from *Oleispira antarctica* possesses the typical catalytic triad for IHL proteins. Its active center resembles those IHLs with a D-K-C catalytic triad, rather than those variants with a D-K-X triad. OaIHL shares some structural and sequential features with other members of the IHL superfamily. *In silico* docking results showed that despite small differences in active site composition, isochorismate binds to in the structure of OaIHL in a similar mode to its binding in phenazine biosynthesis protein PhzD (PDB code 1NF8).

Keywords

iron uptake; isochorismatase hydrolases; structural comparison of isochorismatases; structural genomics

Introduction

Active Fe³⁺ uptake from the environment is an important process for the growth and proliferation in bacteria. It is achieved by siderophore-mediated ferric iron acquisition [1, 2].

*correspondence to: wladek@iwonka.med.virginia.edu.

§Both authors contributed equally to the project

AMG present address: Faculty of Chemistry, University of Warsaw, 02-089 Warsaw, ul. Pasteura 1, Poland;

Siderophores secreted outside the cell are structurally diverse small molecules having chelating abilities [3]. There are three main families of siderophores: catecholate, hydroxamate and carboxylate. The best known and characterized siderophore is enterobactin [4–6], a cyclic trimeric lactone with three bidentate catecholates.

After the ferric iron is chelated [7], the enterobactin is transported through the outer membrane by the 80 kDa O-2b receptor [8]. After transport into the cell, there are two main theories about the mechanism of release of iron from the siderophore: either hydrolysis of the triserine macrocycle [9], or reduction of Fe^{3+} to its ferrous state (Fe^{2+}), which has lower affinity to enterobactin [10].

One of the precursor molecules of many siderophores is isochorismate. Isochorismate belongs to the shikimic acid metabolic pathway, which converts glucose into shikimate and subsequently chorismic acid. Chorismic acid is an intermediate product in the biosynthesis of many biologically relevant molecules like phenylalanine, tyrosine, tryptophan, phenolate, ubiquinones, vitamin K, catecholate siderophores among others; thus, it is often called the 'parent of many aromatic compounds' [11].

The enzyme isochorismatase (EC 3.3.2.1) catalyzes the hydrolysis of isochorismate into 2,3-dihydro-2,3-dihydroxybenzoate and pyruvate (Figure 1). The enzyme belongs to the IHL (isochorismatase-like hydrolases) superfamily, which has been well characterized in pathogenic organisms [12]. We present the crystal structure of the putative isochorismatase from *Oleispira antarctica* (OaIHL) solved at 1.75 Å resolution. Its function was assigned on the basis of the theoretical analysis, therefore the functional role of the target protein still remains to be confirmed experimentally.

The *Oleispira antarctica* [13] is an antarctic bacteria identified in Arctic coastal marine environment. Psychrophilic marine microorganisms possess a lot of interesting properties, such as the breakdown of oil via hydrocarbon degradation [13–15], and thus could be used for industrial tasks like pollution control. Moreover, their ability to survive at relatively low temperatures is a useful adaptation for potential industrial applications that can be further explored.

Materials and Methods

Protein cloning, expression, purification and crystallization

Selenomethionine (Se-Met) substituted protein was produced using standard MSCG protocols as described by Zhang et al. [16]. Crystals were grown by vapor diffusion in hanging drops at room temperature. The reservoir buffer (25 % w/v PEG 3350 25%, 0.2 M NaCl, 0.1 M Na citrate, pH = 5.6) was mixed with an equal volume of protein solution (5 mg·mL⁻¹). Crystals were placed in a cryoprotectant solution (25% w/v PEG3350, 0.2M NaCl, 0.1M Na citrate, 5% glycerol, 10% PEG400, pH=5.6) for 2 minutes and then cooled in a stream of cooled nitrogen vapor.

Data collection, structure determination and refinement

Data collection was performed at 100K on the 19-ID beamline [17] of the Structural Biology Center at the Advanced Photon Source and processed with the HKL-3000 program package at a wavelength of 0.9793 Å [18]. The structure was solved using single-wavelength anomalous diffraction (SAD). Table 1 presents a summary of the data collection and refinement statistics. The phasing and initial model building was performed with HKL-3000 coupled with SHELXD/C/E [19], MLPHARE [20], DM [21], RESOLVE [22], CCP4 [21], ARP/wARP [23], and COOT [24]. The model was further completed and refined by iterations of manual model building with COOT and maximum-likelihood refinement with

REFMAC [25]. 8 TLS groups were used during refinement, which were determined with the TLSMD server [26]. The structure was validated by MOLPROBITY [27] and ADIT [28]. Atomic coordinates and structure factors are deposited in the Protein Data Bank (PDB) [29], entry code 3LQY.

Structure analysis

Sequence searches were performed using the BLAST family of algorithms against nonredundant protein sequence databases using the sequence of OaIHL (PDB: 3LQY) as a search query [30]. Additional comparisons were done by HHpred [31, 32] against the PDB and PFAM databases. Structural analyses were performed with DALI [33] against the PDB database. The multiple sequence alignment (MSA) was generated using Muscle [34]. Pairwise sequence identity was calculated with BLAST2SEQ [30]. Secondary structure was assigned using DSSP [35]. Oligomerization state was analyzed by the PISA server [36]. Mapping of sequence conservation within the superfamily on the structure was done with Consurf [37]. The sequence logos were created with WebLogo [38, 39], based on the MSA shown in Figure 4.

Docking studies

Isochorismic acid was docked to OaIHL using the Molsoft ICM program (version 3.6–1). Solution clustering was also performed using ICM. The best solution was selected on the basis of the best predicted energy of binding (−25.9 kcal) and the highest ICM “naft” quality parameter (3) values (naft is a measure describing the number of visits after the last improvement of energy) [40].

Results

Overall structure

The OaIHL protein consists of 190 residues. The protein crystallizes in a hexagonal $P3_22$ space group with one polypeptide chain in the asymmetric unit (data collection and refinement parameters are shown in Table 1). It belongs to the class of ‘alpha-beta-alpha sandwich’ structures according to both the SCOP [41] and CATCH [42] protein structural classification criteria. It adapts the isochorismatase-like hydrolase fold, characterized by a twisted 6-strand parallel β -sheet (the order of β -strands in the sheet is 3-2-1-4-5-6 relative to the primary sequence) flanked by 6 α -helices (in a Rossmann fold) and an additional 2-strand antiparallel β -sheet (residues 152–154 and 157–159, which correspond to strands 6 and 7 in Fig. 2). This antiparallel β -sheet fragment may exhibit interesting stabilization properties, and is a unique feature among proteins in the IHL superfamily. Two glycerol molecules, which originate from the cryoprotectant solution, are ordered in the structure.. The overall structure of OaIHL is shown in Fig. 2.

The PISA server [36] predicts that OaIHL forms a dimer in solution, based on analysis of the crystal packing what is in perfect agreement with the gel filtration results. The 2-strand antiparallel β -sheet appears to serve as the potential dimerization site (Fig. 3A and B). The solvent-accessible surface area of the monomer is 9220 Å², and the decrease in the solvent-accessible area of the monomer upon dimerization is 1240 Å².

Structural relatives and characteristics of IHL proteins

All proteins similar in sequence to OaIHL (as identified by position-specific BLAST) in the PDB are described as putative isochorismatases. In a search of the Pfam protein family database (PfamA_24.0), OaIHL was classified as a member of family PF00857 which groups together isochorismatases (P-value = 0 and E-value = 8.5×10^{-44} , with a probability score of 100% and an overall sequence identity score of 36%). The HHpred searches also

identified OaIHL as belonging to the isochorismatase family: the best 10 hits (P-value and E-values of 0, probability score of 100%) contain 6 isochorismatases and 4 other proteins belonging to the same enzyme superfamily. Two of them have experimentally confirmed function, others are annotated as members of aforementioned superfamily based on bioinformatics predictions. The most similar proteins with structure are listed in Tables 2 and 3. Similar results were found using DALI to search for similar structures; the 10 highest-scoring hits (P-value and E-values of 0, probability score of 100%) there are 7 isochorismatases and 3 other proteins from the IHL superfamily.

Included in the 10 highest-scoring BLAST hits in the PDB is the PhzD protein from *Pseudomonas aeruginosa* (PDB code 1NF9). PhzD is an enzyme, as determined by kinetic studies [43, 44], that hydrolyzes 2-amino-2-deoxyisochorismate into *trans*-2,3-dihydro-3-hydroxyanthranilic acid and pyruvate, which is the first step in the synthesis of phenazine. An inactive mutant of PhzD was subsequently cocrystallized with an isochorismate substrate (PDB code 1NF8).

Searches against the CATH database showed that the domain structure most similar to OaIHL is a Yecd domain (PDB code 1J2R, CATH ID: 3.40.50.850.2.1.1.1.1), which is 33% identical by sequence to OaIHL. By comparison, PhzD (CATH ID: 3.40.50.850.1.1.1.1.1) is 21% identical by sequence to OaIHL.

Structurally the two proteins (OaIHL and PhzD) differ from each other only by insertions and deletions. Unlike OaIHL, PhzD contains two additional short α -helices in the neighborhood of the active site and lacks the β -hairpin between the 5th and 6th β -strands in the main 6-stranded β -sheet (residues 152–154 and 157–159 in OaIHL). Other members of the superfamily retain the β -hairpin, such as the monomeric cysteine hydrolase PSPPH_2384 from *Pseudomonas syringae* pv. *phaseolicola* (PDB code 3IRV) and the putative hydrolase CV_1320 of the isochorismatase family from *Chromobacterium violaceum* ATCC 12472 (PDB code 3MCW) as listed in Table 2.

All IHLs proteins adopt similar overall topology and can be divided into two groups in terms of oligomeric formation. The first one comprises proteins forming monomers, this group includes PH0999 (PDB code: 1IM5/1ILW), Pnc1p (PDB code:2H0R) and MTPncA (PDB code:3PL1/3GBC). All other proteins fall into the second group, which form dimers or tetramers.

There are two distinct modes of dimerization observed in IHL family members. The first dimerization mode is exemplified by the OaIHL protein (shown in Fig. 3A and B), where the dimer is created by interactions of the α 8 helix with the small antiparallel β -sheet (strands β 6 and β 7). Bxe_A0706 (PDB code:3OQP) shares exactly the same assembly type though in this case the dimer is additionally stabilized by the α 9 helix. The dimerization mode of YecD (PDB code:1J2R) and DVU0033 (PDB code:3HU5) proteins is very similar to OaIHL except for the missing β -hairpin, which in both proteins is mimicked by the hairpin-like loop. The dimers and tetramers of Ta0454 (PDB code:3EEF) and EntB (PDB code:2FQ1) are variations of OaIHL dimers. They only differ in hairpin-like inserts, which are composed of α - β -loop, α -loop or β -loop motifs.

The second dimerization mode is the one characteristic for the structure of the EF3090 protein (PDB code:2A67), as presented in Fig. 3C and D. Here the dimer interface is formed by α 3 and α 4 from one monomer and α 3' and α 4' from the other monomer. The helix pairs are oriented perpendicular with respect to one another. There is no hairpin-like structure in this type of assembly. The same dimerization mode is adapted by the Amet_4583 protein (PDB code:3HB7) which also lacks the hairpin.

The dimer of nicotinamidase/pyrazinamidase *AbPncA* (PDB code:2WT9) is formed by 2 β -strands and an α -helix (β 6- β 8- α 2) unlike *OaIHL*. Although *AbPncA* contains a β -hairpin insertion like *OaIHL*, the hairpin is located in a different place (between strands β 2 and β 3 in the main 6-strand sheet) and does not participate in dimerization. A similar scenario occurs in the case of the CV_1320 protein (which according to the author of the deposit crystallized as a dimer; PDB code: 3MCV), where the α - β hairpin-like structure does not take part in dimerization but rather coordinates the binding of the substrate/ligand molecule. The dimerization contacts in the presumed CV_1320 and Bxe_A0706 are predominately electrostatic in nature. In contrast, in the predicted dimer of the *OaIHL* protein, hydrophobic or nonpolar residues (Gln163, Val164, Ala167, Phe168 and Ala171) in helix α 7 interact with residues Leu152 and Phe154 (β 6) and Ile157 and Val159 (β 7) during oligomerization. This suggests stabilization through hydrophobic interactions.

Most members of the IHL superfamily contain an insert which is another structural characteristic of these proteins, when a small antiparallel β -sheet is inserted after strand β 5. This insert is seen in the case of *OaIHL*, CV_1320 and Bxe_A0706, and proteins containing it we call type I (Table 3). In *AbPncA* and *Pnc1p*, the same β -sheet is inserted between strands β 2 and β 3 (relative to *OaIHL*), and proteins with this topology we call type II.

Active center

So far 22 nonredundant structures were deposited in the PDB of members of the isochorismatase-like superfamily (IHL). 11 of them were classified by CATH, 9 by SCOP and 16 by the PFAM database [45]. Table 3 lists structures of homologs of *OaIHL* classified by SCOP. SCOP divides the IHL superfamily into six subgroups by catalyzed reaction. These are phenazine biosynthesis protein (PhzD), pyrazinamidase/nicotinamidase, 2,3-dihydro-2,3-dihydroxybenzoate synthetase (isochorismatase), YcaC octameric hydrolases of unknown function, ribonuclease MAR1 and *N*-carbamoylsarcosine amidohydrolase. Some of the IHL proteins have not been biochemically characterized, so they are described as having an unknown or uncertain biological function.

Three of the groups, namely the pyrazinamidase/nicotinamidase, ribonuclease and *N*-carbamoylsarcosine amidohydrolase proteins, share an Asp-Lys-Cys catalytic triad, while the others lack the nucleophilic cysteine (for example, Asp38-Lys122-Gly155 in PhzD) [43]. *OaIHL* highly resembles those enzymes where the catalytic triad acts via the cysteine nucleophile. Its catalytic triad comprises the following residues: Asp13, Lys91 and Cys123 (Table 2, Figs. 4 and 5).

The potential active site is located at the C-termini of α 1, β 1, β 2, and β 3, and the N-terminus of α 6. Other residues involved in the putative active site are His58 (β 2) and Gln15 (β 1) as shown in Fig. 4. As shown in Table 2, in the vicinity of the putative site IHL proteins also contain a strongly conserved threonine, a polar/charged amino acid (both located on the second β -strand of the main β -sheet) and a largely conserved glutamine (the loop between β 1 and α 1). However, the role of these semiconserved residues is not fully understood – it is most likely that they stabilize the substrate molecules via hydrogen bonds. In 13 of the 22 proteins (Table 2) both glutamine and threonine are conserved. Neither are conserved in two of the proteins (Amet_4583 and Ta0454). The potential active site of the *OaIHL*, CV_1320, Bxe_A0706 and EF3090 proteins consist of the same conserved residues. We postulate that aforementioned proteins belong to the same subgroup within IHL superfamily.

The structure of EntB (PDB code: 2FQ1) is also very interesting, as it contains an additional globular α -helical domain and a long α -helix just above the active center. Among IHL proteins only *OaIHL*, Bxe_A0706, CV_1320, EF3090, *LdMAR1*, 1YZV, Amet_4583 and *CeMAR1* do not have a loop or helix/helices localized there. Unlike *OaIHL*, the PhzD

protein contains two additional helices ($\alpha 5$: Thr83–Gly88 and $\alpha 6$: Leu90–Gly95) with hydrophobic and aromatic residues (especially Leu90 and Trp94) above the cavity where the substrate was found in the crystal structure. The additional helices may take part in stabilization of the benzoate ring of isochorismatase (and other isochorismate-like molecules) and serve as a “cover” for this cavity that could possibly slow down the release of the product molecule, as shown in Fig. 6A.

In all structurally characterized proteins of the IHL superfamily, the invariant glycine residue precedes the fully conserved *cis*-peptide (in OaIHL Gly119 and Ala120–SeMet121 respectively, Figs. 4 and 5). As proposed by the Caruthers and co-workers [12] in the comparative study of the protozoan isochorismatase superfamily members (PDB codes 1X9G, 1XN4, and 1YZV) this *cis*-peptide probably plays an important role in the correct positioning of the substrate binding residues [12]. In general, the majority of peptide bonds are arranged in *trans* conformation [46]. Only proline, because of its molecular topology, is more likely to adopt a *cis* conformation. According to Weiss [47], non-proline peptide bonds adopt *cis* conformations in 0.03% of all cases, while peptide bonds containing proline adopt *cis* conformations in 5.2% of all cases. MacArthur [48] reported 5.5% and Stevart [49] 6.5% for proline *cis*-peptide and 0.05% for non-proline bonds. However, the characteristic *cis*-peptide between different amino acids is the most conserved feature in IHL proteins of known structure. Even the cysteine from the catalytic triad is not always conserved. Therefore the role of the *cis*-peptide bond must be of crucial relevance to the spatial arrangement of the catalytic site.

Interestingly proteins belonging to the *N*-carbamylsarcosine amidase (Ta0454-3EEF, Amet_4583-3HB7), phenazine biosynthesis protein PhzD (PhzD-2FQ1) and almost all pyrazinamidase/nicotinamidases (PH999-1IM5/1ILW, Pnc1p-2H0R, AbPncA-2WT9, SpNic-3O94, MTPncA-3PL1/3GBC) groups were solved with a metal ion bound near the active center. In the majority of cases it was a divalent metal ion Zn^{2+} , Mg^{2+} or Fe^{2+} . As described by Luo and co-workers [50], the Zn^{2+} in Ta0454 plays a significant structural and catalytic role for this enzyme.

Docking studies

Isochorismate was docked to the crystal structure of OaIHL. The best docking solution shown on Fig. 6B is very similar to the conformation of isochorismate found in the crystal structure of the PhzD protein (Figure 6A) which supports that this *in silico* result may be a correct solution. As shown in Fig. 6B the ligand is located in a highly conserved pocket and interacts with residues in the putative active center of the enzyme.

Conclusion and discussion of the catalytic mechanism

Structural and bioinformatics analyses show that OaIHL from *Oleispira antarctica*, deposited in the PDB with code 3LQY, is a isochorismatase hydrolase. Structural comparisons with closest homologs showed that OaIHL contains the D–K–C catalytic triad, which corresponds to the conserved active site of members of the isochorismatase-like hydrolase (IHL) superfamily. Docking results (Figure 3, 4, 5) confirm that predicted triad is likely to be a catalytic triad. Among the structural relatives we identified a few proteins *i.e.* CV_1320, PSPPH_2384 and EF3090 having the same semiconserved residues glutamine and histidine in the presumed active site also observed in OaIHL. Interestingly unlike best characterized PhzD OaIHL does not contain a structural active site “cover” which perhaps facilitates the catalytic reaction. Also hypothetical mechanism of the conserved catalytic triad can be proposed: glutamic acid protonates the C3' methylene on the pyruvyl sidechain and lysine coordinates ether oxygen which could result in a more favorable ether bond cleavage. Cysteine may either act as a basic species during the deprotonation of the transient

hemiketal or it may attack the C2' atom as a nucleophile. Following the studies made by Walsh and co-workers [11] the lack of the ^{18}O incorporation into 2,3-dihydro-2,3-dihydroxybenzoate (diDHBA) indicates that the water molecule binds to the pyruvyl C2' rather than to the benzoate ring.

This protein also shares the evolutionarily preserved *cis*-peptide which potentially positions the substrate binding residues in a proper orientation for the reaction to take place. However, OaIHL has not yet been investigated from biochemical point of view by experimental methods, thus an exact function must be confirmed.

Acknowledgments

The authors thank Andrzej Joachimiak and the members of the Structural Biology Center at the Advanced Photon Source and the Midwest Center for Structural Genomics for help and discussions. The authors also thank Matthew Zimmerman for critically reading the manuscript. The work described in the paper was supported by NIH PSI Grant GM074942. The results shown in this report are derived from work performed at Argonne National Laboratory, at the Structural Biology Center of the Advanced Photon Source. Argonne is operated by the University of Chicago Argonne, LLC, for the US Department of Energy, Office of Biological and Environmental Research under Contract DE-AC02-06CH11357.

List of abbreviations used

OaIHL	putative isochorismatase hydrolase from <i>Oleispira antarctica</i>
IHL	isochorismatase-like hydrolase
ISC	isochorismate/isochorismic acid
Å	angstrom
SCOP	structural classification of proteins
PDB	Protein Data Bank
RMSD	root mean square deviation
MSA	multiple sequence alignment

References

1. Kadi N, Challis GL. Chapter 17. Siderophore biosynthesis a substrate specificity assay for nonribosomal peptide synthetase-independent siderophore synthetases involving trapping of acyladenylate intermediates with hydroxylamine. *Methods Enzymol.* 2009; 458:431–457. [PubMed: 19374993]
2. Neilands JB. A brief history of iron metabolism. *Biol Met.* 1991; 4(1):1–6. [PubMed: 1854584]
3. Raymond KN, Muller G, Matzanke BF. Complexation of iron by siderophores – a review of their solution and structural chemistry and biological function. *Topics in Current Chemistry.* 1984; (123): 49–102.
4. O'Brien IG, Gibson F. The structure of enterochelin and related 2,3-dihydroxy-N-benzoylserine conjugates from *Escherichia coli*. *Biochimica et biophysica acta.* 1970; 215(2):393–402. [PubMed: 4926450]
5. Pollack JR, Neilands JB. Enterobactin, an iron transport compound from *Salmonella typhimurium*. *Biochem Biophys Res Commun.* 1970; 38(5):989–992. [PubMed: 4908541]
6. Raymond KN, Dertz EA, Kim SS. Enterobactin: an archetype for microbial iron transport. *Proceedings of the National Academy of Sciences of the United States of America.* 2003; 100(7): 3584–3588. [PubMed: 12655062]
7. Harris WR, Carrano CJ, Cooper SR, Sofen SR, Avdeef AE, McArdle JV, Raymond KN. Coordination chemistry of microbial iron transport compounds. 19. Stability-constants and

- electrochemical behavior of ferric enterobactin and model complexes. *Journal of the American Chemical Society*. 1979; (101):6097–6104.
8. Ichihara S, Mizushima S. Identification of an outer membrane protein responsible for the binding of the Fe-enterochelin complex to *Escherichia coli* cells. *J Biochem*. 1978; 83(1):137–140. [PubMed: 146705]
  9. Langman L, Rosenber H, Young IG, Frost GE, Gibson F. Enterochelin system of iron transport in *Escherichia coli*: mutations affecting ferric-enterochelin esterase. *Journal of Bacteriology*. 1972; (112):1142–&. [PubMed: 4565531]
  10. Lee CW, Ecker DJ, Raymond KN. The pH-Dependent Reduction of Ferric Enterobactin Probed by Electrochemical Methods and Its Implications for microbial iron transport. *Journal of the American Chemical Society*. 1985; (107):6920–6923.
  11. Walsh CT, Liu J, Rusnak F, Sakaitani M. Molecular Studies on Enzymes in Chorismate Metabolism and the Enterobactin Biosynthetic Pathway. *Chemical Reviews*. 1990; (90):1105–1129.
  12. Caruthers J, Zucker F, Worthey E, Myler PJ, Buckner F, Van Voorhuis W, Mehlin C, Boni E, Feist T, Luft J, et al. Crystal structures and proposed structural/functional classification of three protozoan proteins from the isochorismatase superfamily. *Protein Sci*. 2005; 14(11):2887–2894. [PubMed: 16199669]
  13. Yakimov MM, Giuliano L, Gentile G, Crisafi E, Chernikova TN, Abraham WR, Lunsdorf H, Timmis KN, Golyshin PN. *Oleispira antarctica* gen. nov., sp. nov., a novel hydrocarbonoclastic marine bacterium isolated from Antarctic coastal sea water. *International journal of systematic and evolutionary microbiology*. 2003; 53(Pt 3):779–785. [PubMed: 12807200]
  14. Head IM, Jones DM, Roling WF. Marine microorganisms make a meal of oil. *Nat Rev Microbiol*. 2006; 4(3):173–182. [PubMed: 16489346]
  15. Yakimov MM, Timmis KN, Golyshin PN. Obligate oil-degrading marine bacteria. *Curr Opin Biotechnol*. 2007; 18(3):257–266. [PubMed: 17493798]
  16. Zhang RG, Skarina T, Katz JE, Beasley S, Khachatryan A, Vyas S, Arrowsmith CH, Clarke S, Edwards A, Joachimiak A, et al. Structure of *Thermotoga maritima* stationary phase survival protein SurE: a novel acid phosphatase. *Structure*. 2001; 9(11):1095–1106. [PubMed: 11709173]
  17. Rosenbaum G, Alkire RW, Evans G, Rotella FJ, Lazarski K, Zhang RG, Ginell SL, Duke N, Naday I, Lazarz J, et al. The Structural Biology Center 19ID undulator beamline: facility specifications and protein crystallographic results. *J Synchrotron Radiat*. 2006; 13(Pt 1):30–45. [PubMed: 16371706]
  18. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A*. 1997; (276):307–326.
  19. Sheldrick GM. A short history of SHELX. *Acta crystallographica Section A, Foundations of crystallography*. 2008; 64(Pt 1):112–122.
  20. Otwinowski, Z., editor. MLPHARE: Isomorphous replacement and anomalous scattering- Proceedings of the CCP4 SERC, Daresbury Laboratory; 1991.
  21. Cowtan K. 'dm': An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*. 1994; (31):34–38.
  22. Terwilliger TC. Automated structure solution, density modification and model building. *Acta crystallographica Section D, Biological crystallography*. 2002; 58(Pt 11):1937–1940.
  23. 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]
  24. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Cryst D*. 2004; 60:2126–2132.
  25. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica Section D, Biological crystallography*. 1997; 53(Pt 3):240–255.
  26. Painter J, Merritt EA. TLSMD web server for the generation of multi-group TLS models. *Journal of Applied Crystallography*. 2006; 39(1):109–111.

27. Lovell SC, Davis IW, Adrendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by C alpha geometry: phi,psi and C beta deviation. *Proteins-Structure Function and Genetics*. 2003; (50):437–450.
28. Yang H, Guranovic V, Dutta S, Feng Z, Berman HM, Westbrook JD. Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. *Acta crystallographica Section D, Biological crystallography*. 2004; 60(Pt 10):1833–1839.
29. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr. Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol*. 1977; 112(3):535–542. [PubMed: 875032]
30. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997; 25(17):3389–3402. [PubMed: 9254694]
31. Soding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics*. 2005; 21(7): 951–960. [PubMed: 15531603]
32. Soding J, Remmert M, Biegert A, Lupas AN. HHsenser: exhaustive transitive profile search using HMM-HMM comparison. *Nucleic acids research*. 2006; 34(Web Server issue):W374–378. [PubMed: 16845029]
33. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. *Nucleic Acids Res*. 2010; 38(Web Server issue):W545–549. [PubMed: 20457744]
34. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004; 32(5):1792–1797. [PubMed: 15034147]
35. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983; 22(12):2577–2637. [PubMed: 6667333]
36. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007; 372(3):774–797. [PubMed: 17681537]
37. Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*. 2003; 19(1):163–164. [PubMed: 12499312]
38. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004; 14(6):1188–1190. [PubMed: 15173120]
39. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. *Nucleic acids research*. 1990; 18(20):6097–6100. [PubMed: 2172928]
40. Abagyan RA, Totrov MM. D.A.K.: Icm: A New Method For Protein Modeling and Design: Applications To Docking and Structure Prediction From The Distorted Native Conformation. *Journal of Computational Chemistry*. 1994; (15):488–506.
41. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *Journal of molecular biology*. 1995; 247(4):536–540. [PubMed: 7723011]
42. Greene LH, Lewis TE, Addou S, Cuff A, Dallman T, Dibley M, Redfern O, Pearl F, Nambudiry R, Reid A, et al. The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution. *Nucleic Acids Res*. 2007; 35(Database issue):D291–297. [PubMed: 17135200]
43. Parsons JF, Calabrese K, Eisenstein E, Ladner JE. Structure and mechanism of *Pseudomonas aeruginosa* PhzD, an isochorismatase from the phenazine biosynthetic pathway. *Biochemistry*. 2003; 42(19):5684–5693. [PubMed: 12741825]
44. Kunzler DE, Sasso S, Gamper M, Hilvert D, Kast P. Mechanistic insights into the isochorismate pyruvate lyase activity of the catalytically promiscuous PchB from combinatorial mutagenesis and selection. *J Biol Chem*. 2005; 280(38):32827–32834. [PubMed: 16036918]
45. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, et al. The Pfam protein families database. *Nucleic Acids Res*. 2004; 32(Database issue):D138–141. [PubMed: 14681378]
46. Jabs A, Weiss MS, Hilgenfeld R. Non-proline cis peptide bonds in proteins. *J Mol Biol*. 1999; 286(1):291–304. [PubMed: 9931267]

47. Weiss MS, Jabs A, Hilgenfeld R. Peptide bonds revisited. *Nat Struct Biol.* 1998; 5(8):676. [PubMed: 9699627]
48. MacArthur MW, Thornton JM. Influence of proline residues on protein conformation. *Journal of molecular biology.* 1991; 218(2):397–412. [PubMed: 2010917]
49. Stewart DE, Sarkar A, Wampler JE. Occurrence and role of cis peptide bonds in protein structures. *Journal of molecular biology.* 1990; 214(1):253–260. [PubMed: 2370664]
50. Luo HB, Zheng H, Zimmerman MD, Chruszcz M, Skarina T, Egorova O, Savchenko A, Edwards AM, Minor W. Crystal structure and molecular modeling study of N-carbamoylsarcosine amidase Ta0454 from *Thermoplasma acidophilum*. *Journal of structural biology.* 2010; 169(3):304–311. [PubMed: 19932181]

Figure 1.
Scheme of the reaction catalyzed by EC 3.3.2.1 enzymes.

Figure 2. Structure of the putative isochorismatase hydrolase from *Oleispira antarctica* (PDB code 3LQY) shown in cross-eyed stereo. The N and C termini, as well as the secondary structure elements, are labeled.

Figure 3. Predicted oligomerization and ligand docking. Cartoon (A) and surface representations (B) show the presumed OaIHL dimers with isochorismate docked to the crystal structure. Cartoon (C) and surface representations (D) of another possible variant of dimerization present within the IHL superfamily: the structure of the EF3090 protein from *Enterococcus faecalis* v583. Each monomer is shown in a different color (gray or teal). The ligands in A and B are shown in sphere representation and colored by CPK atom colors.

Figure 4.

The putative active center of OaIHL with docked isochorismate shown in sphere representation and colored by atom type. The presumed catalytic triad of OaIHL (D13, K91 and C123) is marked in red, other residues forming the pocket are shown in magenta. Residues in the putative active site are correlated to logos that describe the sequence conservation within the superfamily. Each residue seen in a given position in the alignment are stacked above the position, with the size of each letter indicating the relative frequency of that amino acid in the family.

Figure 5. Multiple sequence alignment (MSA) of the *Oleispira antarctica* OaIHL (PDB: 3LQY) putative isochorismatase hydrolase and the closest structural homologues. Secondary structure elements corresponding to OaIHL structure are shown above the alignment—arrows represent β -strands while tubes correspond to α -helices. Invariant and strongly conserved residues are highlighted in grey. The numbers shown above the MSA represent the numbering of the residues in the OaIHL sequence. Digits shown in parenthesis pinpoint the number of amino acid residues removed from the MSA and which are not shown in the figure. The catalytic triad is marked with red asterisks, and the other residues in the putative active center are marked with blue asterisks. The threshold of shading is 70% similarity.

Figure 6.

(A) Superposition of PhzD with ligand (yellow) and OaIHL (gray). The active center “cover” is shown with an arrow. (B) Sequence conservation mapped onto the OaIHL structure. The coloring scheme is shown on the figure, where 1 denotes the most variable residues and 9 the most conserved. Docked ligand is shown in sphere representation, colored by atom type.

Table 1

Data collection and refinement statistics.

Organism	<i>Oleispira Antarctica</i> PDB code: 3LQY
Crystal	
Space group	P3 ₂ 2
Unit cell	a = b = 41.6 Å, c = 171.2 Å $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$
Solvent content (%)	41.2
No. of molecules in asymmetric unit	1
Data collection	
Diffraction protocol	SAD
Wavelength (Å)	0.9793
Resolution (Å)	50.00–1.75
Highest resolution shell (Å)	1.79–1.75
Unique reflections	18174
Redundancy	8.7 (6.3)
Completeness (%)	98.8 (91.3)
<i>I</i> / σ (<i>I</i>)	48.5 (2.5)
R _{merge} (%)	9.0 (54.0)
Refinement	
R _{work} (%)	21.2 (35.4)
R _{free} (%)	24.3 (37.9)
RMSD for bond length (Å)	0.016
RMSD for bond angles (Å)	1.6
Number of protein atoms	1410
Number of water molecules	126
Ramachandran plot	
Most favored regions (%)	97.9
Additional allowed regions (%)	2.1

Data for the highest resolution shell are given in parentheses.

Ramachandran statistics were calculated with *MOLPROBITY*.R_{merge} was calculated for merged Friedel pairs.

Table 2

Characteristics of homologous IHL proteins.

organism	protein/locus tag	PDB ID	Catalytic triad	Additional structural motif	Cis-peptide	Conserved residues	Polar/charged residue	Oligomerization in the crystal
<i>Oleispira antarctica</i>	OaIHL	3LQY	D13-K92-C125	β-type I	A120-M121	NA, Q15	H58	dimer
<i>Desulfotribrio vulgaris</i>	DVU0033	3HU5	D15-K106-C139	-	T134-Q135	NA, Q17	R58	tetramer
<i>Chromobacterium violaceum</i>	CV_1320	3MCW	D18-K92-S125	β-type I	V121-S122	NA, Q20	H59	dimer
<i>Enterococcus faecalis</i> v583	EF3090	2A67	D10-K80-C113	-	V108-Q109	NA, Q12	H51	dimer
<i>Burkholderia xenovorans</i>	Bxe_A0706	3OQP	D12-K88-C121	β-type I	Y116-M117	NA, Q14	N55	dimer
<i>Streptomyces avermitilis</i>	SAV_1388	3KL2	E49-K148-C181	-	F176-L177	NA, Q51	NA	tetramer
<i>Escherichia coli</i>	YcaC/c1034	1YAC	D19-R84-C118	-	V113-V114	T57, Q21	S59	octamer
<i>Escherichia coli</i>	YecD	1J2R	D26-K113-G145	-	I140-S141	NA, Q28	NA	dimer
<i>Thermoplasma acidophilum</i>	Ta0454	3EEF	D9-K90-C123	-	L118-D119	NA, NA	D50	dimer/tetramer
<i>Arthrobacter</i> sp.	AAB23138	1NBA	D51-K144-C177	(α-helix)	A172-T173	T92, NA	N94	tetramer
<i>Alkaliphilus metalliredigens</i>	Amet_4583	3HB7	D10-K96-C129	-	V124-W125	NA, NA	E56	dimer/tetramer
<i>Caenorhabditis elegans</i>	CeMAR1/F35G2.2	2B34	D21-K84-C114	-	I109-E110	T57, Q23	Q59	tetramer
<i>Leishmania donovani</i>	LdMAR1/LinJ10.1740	1X9G/IXN4	D19-K82-C112	-	I107-E108	T57, Q21	H59	tetramer
<i>Trypanosoma cruzi</i>	EntB/Z0737	1YZV	D19-K83-C115	-	F110-E111	T58, Q21	Q60	tetramer
<i>Escherichia coli</i>	PhzD	2FQ1	D37-K123-G156	(α-domain)	V151-Y152	T77, Q39	Q79	dimer
<i>Pseudomonas aeruginosa</i>	PhzD	1NF9/INF8	D38-K122-G155	-	V150-Y151	T76, Q40	Q78	dimer
<i>Pseudomonas syringae</i>	PSPPH_2384	3IRV	D56-K146-C179	loop-type I	T174-V175	NA, Q58	H99	dimer
<i>Pyrococcus horikoshii</i>	PH0999	1IM5/ILW	D10-K94-C133	-	V128-A129	T50, Q12	D52	monomer
<i>Saccharomyces cerevisiae</i>	Pnc1p/YGL037C	2H0R	D8-K122-C167	β-type II	V162-A163	T49, Q10	D51	monomer
<i>Streptococcus pneumoniae</i>	SpNlc/SP_1583	3O94	D9-K103-S136	-	V131-L132	T51, NA	D53	tetramer
<i>Acinetobacter baumannii</i>	AbPncA/ABAYE0059	2WT9	D16-K114-C159	β-type II	I154-A155	T52, Q18	D54	dimer
<i>Mycobacterium tuberculosis</i>	MTPhcA/Rv2043c	3PL1/3GBC	D8-K96-C138	-	I133-A134	T47, Q10	D49	monomer

Structures shaded in different colors correspond to different subfamilies identified in the SCOP database. In gray are pyrazinamidase/nicotinamidases, in yellow are phenazine biosynthesis proteins (PhzD), in light blue are ribonucleases, in pink are N-carbamylsarcosine amidohydrolases, in blue is the hypothetical protein YecD, and in purple is the octameric hydrolase of unknown function (YcaC). “NA” stands for “not available”; / - two possibilities for the oligomerization state are shown as predicted by PISA.

Table 3

CATH and SCOP classification of IHL proteins.

PDB ID	Scop Domain	CATH ID	Release date	reference
1NF9/1NF8	Phenazine biosynthesis protein PhzD	3.40.50.850.1.1.1.1.1	2003	Parsons <i>et al.</i> , 2003
1J2R	Hypothetical protein YecD	3.40.50.850.2.1.1.1.1	2004	NA
1YAC	YcaC	3.40.50.850.3.1.1.1.1	1999	Colovos <i>et al.</i> , 1998
1IM5/1ILW	Pyrazinamidase/nicotinamidase	3.40.50.850.4.1.1.1.1	2001	Du <i>et al.</i> , 2001
1X9G	Ribonuclease MAR1	3.40.50.850.5.1.1.1.1	2004	Caruters <i>et al.</i> , 2005
1XN4	Ribonuclease MAR2	3.40.50.850.5.1.1.2.1	2005	Caruters <i>et al.</i> , 2005
1NBA	N-carbamoylsarcosine amidohydrolase	3.40.50.850.6.1.1.1.1	1994	Romao <i>et al.</i> , 1992 Zajc <i>et al.</i> , 1996
2FQ1	Isochorismatase	3.40.50.850.1.2.1.1.1	2006	Drake <i>et al.</i> , 2006
1YZV	Hypothetical protein	3.40.50.850.5.1.1.1.1	2005	Caruters <i>et al.</i> , 2005