Structural, biochemical, and evolutionary characterization of glyoxylate/hydroxypyruvate reductases shows their division into two distinct subfamilies

Jan Kutner[†,‡,§] *, Ivan G. Shabalin*[†,‡] *, Dorota Matelska*[†,‡,¶] *, Katarzyna B. Handing*[‡] *, Olga Gasiorowska*[‡] *, Piotr Sroka*[‡] *, Maria W. Gorna*[§] *, Krzysztof Ginalski*[*,¶] *, Krzysztof Wozniak*[*,§] *, and Wladek Minor*[*,‡,#]

[‡]Department of Molecular Physiology and Biological Physics, University of Virginia, 1340 Jefferson Park Avenue, Charlottesville, VA 22908, United States

[§]Biological and Chemical Research Centre, Department of Chemistry, University of Warsaw, 101 Zwirki i Wigury, 02-089 Warsaw, Poland

[¶]Laboratory of Bioinformatics and Systems Biology, Centre of New Technologies, University of Warsaw, 93 Zwirki i Wigury, 02-089 Warsaw, Poland

[#]Department of Chemistry, University of Warsaw, 1 Ludwika Pasteura, 02-093 Warsaw, Poland

ABSTRACT

The D-2-hydroxyacid dehydrogenase (2HADH) family illustrates a complex evolutionary history with multiple lateral gene transfers, gene duplications, and losses. As a result, the exact functional annotation of individual members can be extrapolated to a very limited extent. Here, we revise the previous simplified view on the classification of the 2HADH family; specifically, we show that the previously delineated glyoxylate/hydroxypyruvate reductase (GHPR) subfamily consists of two evolutionary separated GHRA and GHRB subfamilies. We compare two representatives of these subfamilies from *Sinorhizobium meliloti* (*Sm*GhrA and *Sm*GhrB), employing a combination of biochemical, structural, and bioinformatics approaches. Our kinetic results show that both enzymes reduce several 2-ketocarboxylic acids with overlapping, but not equivalent, substrate preferences. *Sm*GhrA and *Sm*GhrB show highest activity with glyoxylate and hydroxypyruvate, respectively; in addition, only *Sm*GhrB reduces 2-keto-D-gluconate, and only *Sm*GhrA reduces pyruvate (with low efficiency). We present nine crystal structures of both enzymes in *apo*-forms and in complexes with cofactors and substrates/substrate analogs. In particular, we determined a crystal structure of *Sm*GhrB with 2-keto-D-gluconate, which is the biggest substrate crystallized with a 2HADH member. The structures reveal significant differences between *Sm*GhrA and *Sm*GhrB, both in the overall structure and within the substrate-binding pocket, offering insight into the molecular basis for the observed substrate preferences and subfamily differences. In addition, we provide an overview of all GHRA and GHRB structures complexed with a ligand in the active site.

INTRODUCTION

D-2-hydroxyacid dehydrogenases (2HADHs) catalyze the reversible NAD(P)H-dependent stereospecific reduction of 2-ketocarboxylic acids to the corresponding (R)-2-hydroxy acids (Figure 1). Under physiological conditions, these enzymes preferentially catalyze the reaction in one direction using either NAD(P)H or $NAD(P)^+$ as a cofactor. Due to the diversity of the accepted substrates, 2HADHs are involved in various cellular processes. For example, vancomycin resistance protein H is associated with antibiotic resistance,[1] phosphonate dehydrogenase is involved in the utilization of phosphite as an exogenous phosphorus source,[2] D-lactate dehydrogenase catalyzes the last step of anaerobic glycolysis,[3] and plant hydroxypyruvate dehydrogenases are involved in the photorespiratory cycle.[4] Human 2HADH enzymes play an important role in metabolic pathways; a mutation in the human gene encoding glyoxylate reductase/hydroxypyruvate reductase (GRHPR) is a genetic basis for primary hyperoxaluria type II, a rare inherited disorder of glyoxylate metabolism.[5]

2HADHs have a number of practical applications beyond their multiple cellular functions. Enzymatic reduction by 2HADHs produces optically active (R)-2-hydroxy acids with exceptionally high stereoselectivity, and the optically pure products have potential for use as "green" synthons and precursors for chiral compounds in pharmaceutical and biotechnology industrial applications.[6-8] Therefore, obtaining in-depth insights into the substrate specificities, catalytic kinetics, and molecular mechanism of these enzymes is of great interest.

Prior efforts have been made to systematically classify 2HADHs and their characteristics. Based on sequence similarity, it was proposed almost three decades ago that 2HADHs constitute a protein family distinct from the L-specific dehydrogenases.[9] Further biochemical studies have characterized more members of this family, which resulted in initial attempts to divide them into

evolutionarily related subfamilies sharing similar substrate specificity.[10, 11] The proposed division does not, however, fully explain differences in substrate specificity and cofactor preference for closely related members within the same subfamily.

The most striking examples are found in the previously designated glyoxylate/hydroxypyruvate reductase (GHPR) subfamily.[10] GHPR comprises both reductases and dehydrogenases with relatively broad substrate specificities, spanning all three domains of life. In the classification by Fauvart *et al.*,[10] the subfamily included yeast mandelate dehydrogenase,[12] human and archaeal glyoxylate/hydroxypyruvate reductases,[13] bacterial phosphonate dehydrogenases,[14] plant and fungal hydroxyphenylpyruvate reductases,[11, 15] and bacterial enzymes that reduce a broad range of substrates.[10, 16] In addition to the direction of the reaction and accepted substrates, the cofactor preferences also differ among the GHPR members. For example, GHPR from *Escherichia coli* (*Ec*GhrA) reduces glyoxylate and hydroxypyruvate, preferentially using NADPH as a cofactor.[16] Yet its homolog from *Rhizobium etli*, *Re*GxrA, was reported to reduce the same substrates but with only NADH as the cofactor.[10] These differences suggest that even within the same subfamily, the functional annotation cannot be extrapolated. Although crystal structures for 17 GHPRs are reported in the PDB, most of them are without the cofactor and/or substrate bound,[13, 17-19] leaving the underlying substrate and cofactor discrimination mechanisms elusive.

To revise the simplified view on the previously delineated GHPR family, we chose two enzymes from *Sinorhizobium meliloti* 1021, an α-proteobacterium whose genome encodes a large number of 2HADH paralogs, none of which have been studied to date. *S. meliloti* is a free-living soil Gram-negative bacterium, capable of nitrogen fixation and symbioses with leguminous plants, thus playing a vital role in agriculture.[20, 21]

In this paper, we present a phylogenetic analysis that suggests a division of GHPRs into two distinct subfamilies. In an attempt to understand the functions of and the differences between the two novel subfamilies, we used biochemical, structural, and bioinformatics approaches to comparatively characterize a representative enzyme from each of the subfamilies from *S. meliloti*.

MATERIALS AND METHODS

Reagents and Software Used. All reagents used in this work with the respective commercial sources and catalog numbers, as well as all software with RRID identifiers, are listed in Table S4.

Phylogenetic Analysis. Sequences of proteins assigned to the "D-isomer specific 2-hydroxyacid dehydrogenase family" confirmed by biochemical tests were collected from UniProt.[22] To extend this initial set, these collected sequences were used as queries in BLAST searches in selected proteomes in UniProt (*Homo sapiens*, *Mus musculus*, *Saccharomyces cerevisiae*, *Escherichia coli*, *Agrobacterium radiobacter*, *Cyanidioschyzon merolae*, *Arabidopsis thaliana*, *Solanum lycopersicum*, *Zea mays*, *Bartonella grahamii*, *Brucella suis*, *Medicago truncatula*, *Plasmodium falciparum*, *Pyrococcus furiosus*, *Rhizobium etli*, *Rhizobium grahamii*, *Sinorhizobium americani*, *Sinorhizobium meliloti*, *Rhizobium fredii*, and *Xanthobacter autotrophicus*). Hits with E-value < 10^{-20} were included in the final set. Sequence alignment for the region encompassing both catalytic and substrate-binding domains was generated using MAFFT 7[23] with the "linsi" option. Poorly aligned regions (with more than 20% gaps) were removed with trimAl 1.4.[24] The maximum-likelihood tree was generated with FastTree2[25] (JTT+CAT model and Shimodaira-Hasegawa test for estimation of support values) and visualized with Archaeopteryx.[26]

In order to independently evaluate the consistency of the obtained tree topology, alternative trees were built and compared in the context of the GHRA and GHRB separation as follows: a) based

on the alignment prior to the trimming; b) based on the alignment with columns containing more than 80% of gaps removed; c) based on the alignment of the catalytic domain only (corresponding to positions 3–97 and 287–319 in *Sm*GhrA); and d) using the WAG+CAT model. The alternative trees showed similar topologies with clear separation between the GHRA and GHRB clades.

Protein Expression and Purification. The genes encoding *Sm*GhrA (SMc02828) and *Sm*GhrB (SMc04462) were cloned into the pSGC-His expression vectors by the New York Structural Genomics Research Consortium (NYSGRC targets 011884 and 012132, respectively). The *E. coli* BL21-CodonPlus (DE3)-RIPL cells, transformed with the pSGC-His expression vector coding the protein of interest, were grown at 37 °C in Luria Broth expression media. After the culture had reached an OD_{600} of 0.6–1.0, the temperature was decreased to 16 °C, protein expression was induced with 0.5 mM IPTG, and cell growth continued for 18 additional hours. Cells were harvested by centrifugation at 35,000 rpm at 4 °C for 40 min (JA-10, Beckman Coulter), and pellets were stored at –80 °C for further use. Pellets were thawed on ice and resuspended in buffer A [20 mM HEPES (pH 7.5), 500 mM NaCl, 15 mM NaN_3, 0.5 mM TCEP, 10% glycerol, and 5 mM imidazole] with freshly added 250 units of benzonase endonuclease and a tablet of protease inhibitor cocktail. The pellets were disrupted by passing the cell suspension three times through a homogenizer at 15,000–17,000 psi (EmulsiFlex-C3, Avestin Inc.). Cell lysates were centrifuged at 35,000 rpm (45-Ti, Beckman Coulter) at 4 °C for 40 min.

The supernatants were loaded onto a gravity chromatography column containing 5 mL Ni-NTA resin calibrated with buffer A and then washed with buffer A supplemented with 20 mM imidazole. Afterwards, the proteins were eluted with buffer A, with imidazole concentration increased to 350 mM. The eluted proteins were further purified by size-exclusion chromatography (ÄKTA FPLC with a Hi-Load 16/600 SuperdexTM 200 pg column, GE Healthcare) and eluted with

crystallization buffer B [20 mM HEPES (pH 7.5), 150 mM NaCl, 10% glycerol, 15 mM NaN_3, and 0.5 mM TCEP]. Protein purity in the resulting fractions was confirmed by SDS-PAGE. The combined fractions containing *Sm*GhrA or *Sm*GhrB protein were concentrated (10 kDa Amicon Ultra-15 Centrifugal Filter Units, EMD Millipore) to 11.7 mg/mL and 13.2 mg/mL, respectively, and stored in buffer B.

For the activity assays only, the His-tag was removed by a five-fold addition of 0.2 mg/ml rTEV protease to the samples eluted from the affinity column. The rTEV cleavage was performed at 16 ^{o}C overnight and confirmed by SDS-PAGE. The protease reaction products were passed through a second Ni-NTA column pre-equilibrated with buffer A. The flowthrough from the second Ni-NTA column was applied on the size-exclusion column connected to the ÄKTA FPLC system, eluted with buffer B, and concentrated to 1 mg/ml.

Protein Thermal Shift Assay. The thermal stability of proteins was measured with RT-PCR (C1000, Bio-Rad) in a 96-well PCR plate (Bio-Rad). Each well contained 12 µl of 20x Sypro Orange dye in buffer B and 12 µl of 0.2 mg/mL of the protein, optionally supplemented with 10 mM cofactor and/or 10 mM ligand of interest. The used method consisted of three steps: i) sample pre-incubation for 5 min at 20 °C; ii) temperature increase by 0.5 °C from 20 °C until 80 °C within the total time of 30 min.; and iii) temperature decrease to 4 °C.

Substrate Screening. Enzymes with NADPH were incubated for 10 minutes in a 96-well half-area clear polystyrene plate (Corning) prior to adding 20 µL of the substrates to initiate the reaction. The final assay mixture contained 0.01 µg/µL *Sm*GhrA or 0.004 µg/µL *Sm*GhrB protein, 0.4 mM NADPH, and substrates of various concentrations (Tables S1 and S2), in a final volume of 120 µL of kinetic buffer (150 mM NaCl, 150 mM Tris-HCl pH 7.5) at 25 °C. The reactions were monitored using the absorbance of NADPH at 340 nm. The absorbance measurements were

carried out in a PHERAstar FS (BMG LABTECH Omega) or an Infinite M200 PRO (Tecan) microplate reader equipped with UV absorbance spectrophotometer at 340 nm. The reaction was measured for 750 s with the absorption reading every 25 s.

Steady-state Kinetics. The initial reaction velocity was examined by varying the concentration of one substrate while keeping the NADPH concentration constant. Enzyme *Sm*GhrA (0.01 μg/μL) or *Sm*GhrB (0.004 μg/μL) with 0.4 mM NADPH was incubated for 10 min in a 96-well half-area clear polystyrene plate (Corning) in filtered and degassed kinetic buffer (150 mM NaCl, 150 mM Tris-HCl pH 7.5) at 25 °C. To initiate the reaction, 20 μL of the respective substrate stock was added. The absorbance was measured at 340 nm every 25 s for 750 s in total. Each measurement was performed in triplicate. The initial reaction velocities were fitted by the nonlinear least-squares method to both a) the Michaelis-Menten equation $v = V_{max} \times s / (K_m + s)$, where v is the reaction velocity, V_{max} is maximum reaction velocity, s is substrate concentration, and K_m is the Michaelis constant and b) the Hill equation $v = V_{max} \times s^n / (K_{0.5}{}^n + s^n)$, where n is the Hill coefficient. Because K_m is a special case of $K_{0.5}$ (n=1), K_m and $K_{0.5}$ are generally expressed as $K_{0.5}$ throughout the paper. The obtained fittings were compared with the F-test (with the null hypothesis that the simpler Michaelis-Menten model is correct), using a cutoff of $p = 0.01$. The kinetic parameters were calculated using Prism software (GraphPad).

Protein Crystallization and Data Collection. All complexes of *Sm*GhrA and *Sm*GhrB with ligands and the *apo*-forms were crystallized with the sitting-drop vapor-diffusion method using the recombinant protein preparations with His-tag present. 100 mM NADPH tetrasodium salt and $NADP^+$ disodium salt were prepared in 100 mM HEPES buffer at pH 7.0. The 1 M solutions of oxalic acid, glycolic acid, and 2-keto-D-gluconic acid were prepared in the same buffer. After dissolution, the pH of the oxalic acid and 2-keto-D-gluconic acid solutions was adjusted to 7.0 by

the addition of sodium hydroxide. If applicable, the respective ligand stock solution(s) was added to the protein, and the resulting mix was incubated for 30 min at 16 °C. Subsequently, the protein-ligand solution was supplemented with either a) 1/40 (v/v) of 2 mg/mL chymotrypsin solution to perform limited proteolysis or b) 1/15 (v/v) of 1 mg/mL rTEV solution to cleave the His-tag and incubated for 1 hour at 16 ˚C. The crystallization experiments were set up using a Mosquito crystallization robot (TTP Labtech). 0.2 μL of the protein solution was mixed with 0.2 μL of the crystallization screen solution and equilibrated against 1.5 M NaCl in 3-drop, 96-well crystallization plates (Hampton Research). The details of protein preparation and crystallization are summarized in Tables 2 and 3.

Crystals suitable for X-ray experiments were harvested and cryo-protected using one of the following approaches: 1) dipping in the precipitant mix supplemented with 25% glycerol or 20% ethylene glycol; 2) bathing in paratone-N oil for 2-3 minutes; 3) drying for 5-15 minutes over 1 M NaCl solution in a closed vial; or 4) drying by waving the harvested crystal in the air for 1 minute. Subsequently, the crystals were flash-cooled in liquid nitrogen. Diffraction data were collected from single crystals at 100 K at the LS-CAT 21-ID-G or at the SBC-CAT 19-ID beamlines at the Advanced Photon Source, Argonne National Laboratory (Argonne, IL). The details of cryo-protection and data collection are summarized in Tables 2 and 3.

For some complexes, both limited proteolysis with chymotrypsin and His-tag removal with rTEV resulted in crystals that diffracted to ~2 Å resolution and led to high-quality structures. In these cases, the best dataset was chosen to represent each complex. In one case (PDB ID: 5j23), NADPH was presumably hydrolyzed in the crystallization mixture due to the acidic environment created by the glycolic acid (Table 3), and the nicotinamide moiety was eliminated from NADPH, producing 2′-phospho-ADPR,[27] which was discovered during the structure refinement. Oxalate

and 2-keto-D-gluconic acid were explicitly added to the protein samples, but other ligands in the active site (sulfate, malonate, and citrate) were components of the commercial crystallization cocktails used for screening. All experimental details were tracked by the LabDB database.[28, 29]

Data Processing and Structure Determination. Data reduction and scaling for all structures were performed with HKL-3000.[30, 31] The structures were determined by molecular replacement in HKL-3000 integrated with MOLREP[32] and auxiliary programs from the CCP4 package.[33] The first structure of *Sm*GhrA (complexed with $NADP^+$ and sulfate, PDB ID: 4weq) was solved using the 4n18 PDB deposit as the template; the first structure of *Sm*GhrB (complexed with NADPH and oxalate, PDB ID: 5v7g) was solved using the 3baz PDB deposit as the template.

Model Building and Refinement. The initial model building was performed with Buccaneer[34] followed by optimization of side-chain conformations with Fitmunk[35] as implemented in HKL-3000. After the initial model building and refinement, the atomic models for the first structures were used as templates for solving all other structures by molecular replacement. The structures were refined in HKL-3000 using REFMAC5[36] in the restrained mode with isotropic ADPs and hydrogen atoms in riding positions. TLS groups were introduced for all structures in the later stages of refinement as determined by the TLS Motion Determination Server.[37] Water molecules were not included in the TLS groups. Automatic local NCS was applied throughout refinement for all structures containing more than one subunit in the asymmetric unit. The Coot software[38] was used for the visualization of electron density maps, manual inspection, and correction of the atomic models. Multiple tools integrated into Coot, standalone version of MolProbity,[39] and the PDB validation tools[40] were used for structure quality assessment. The structures of *Sm*GhrA have no Ramachandran plot outliers; some structures of *Sm*GhrB have one outlier in each chain (Asp94) located next to the binding site for the pyrophosphate moiety of the cofactor. These residues are

well-defined on the electron density maps. PDB IDs and statistics for data collection, refinement, and validation are summarized in Tables 2 and 3. The diffraction images are available on the Integrated Resource for Reproducibility in Macromolecular Crystallography website (http://proteindiffraction.org/).[41] The ligands in the active site and the electron density maps, including omit maps, can be inspected interactively using Molstack[42] at http://molstack.bioreproducibility.org/c/FdBO/. The reviewers of this manuscript had access to the coordinates, diffraction data, and electron density maps for all the structures through the PDB.

Accession Codes. UniProt accession IDs: Q92T34 (*Sm*GhrA, SMc02828), Q92LZ4 (*Sm*GhrB, SMc04462). *Sm*GhrA PDB IDs: 4weq, 4z0p, 5unn. *Sm*GhrB PDB IDs: 5v7n, 5v7g, 5v6q, 5j23, 5uog, 5v72.

RESULTS

Maximum-likelihood Phylogenetic Analysis. To cluster 2HADHs into functionally related groups, we performed in-depth phylogenetic analyses. According to the previously reported neighbor-joining phylogenetic tree,[10] the 2HADH family can be divided into seven functionally related subfamilies, one of which is the GHPR subfamily. Contrary to this division, our maximum-likelihood tree suggests that sequences previously assigned to the subfamily, in fact, belong to two evolutionary separated clades (Figure 2). The underlying tree topology appears consistently in trees obtained with modified approaches (see Materials and Methods for details).

The larger clade, which we call GHRB, has 16 biochemically characterized members. GHRB shows a broad taxonomic distribution and is consequently highly divergent, including enzymes with diverse substrate and cofactor specificities, such as yeast mandelate dehydrogenase (Uniprot ID: Q7LLW9_RHOGR), 2-ketogluconate reductase (Uniprot ID: 2KGR_GLUOX), and bacterial (Uniprot ID: GHRB_ECOLI), human (Uniprot ID: GRHPR_HUMAN), and plant (Uniprot ID:

HPR1_ARATH, HPR2_ARATH, HPR3_ARATH) glyoxylate/hydroxypyruvate reductases.[14, 19, 43-45] There are 32 GHRB structures reported in the PDB, including six presented herein. Twelve structures have a ligand in the active site (Table 4). Several papers describing some of the GHRB structures have previously been published.[18, 19, 46, 47]

The second clade, which we call GHRA, is bacteria-specific and comprises only two members that have been studied biochemically:[10, 16] *E. coli Ec*GhrA (UniProt ID: GHRA_ECOLI) and *R. etli Re*GxrA (UniProt ID: C1JH53_RHIET). Thirteen structures of the GHRA subfamily have been determined so far, including three reported herein. Six structures have a ligand bound in the active site (Table 4), but no papers describing the GHRA structures have been published.

In all reconstructed trees, one of the sister clades of GHRA includes a 2HADH from *Haloferax mediterranei* (UniProt ID: DDH_HALMT), which reduces 2-ketocarboxylic acids with an unbranched chain of 4-5 carbon atoms.[48] The consistent placement of the two clades (GHRA and ddh) in relation to GHRB supports the validity of the tree and the polyphyletic origin of GHRA and GHRB enzymes.

Protein Purification and Substrate Screening. To dissect differences between the GHRA and GHRB subfamilies, we selected *Sm*GhrA (locus tag SMc02828) and *Sm*GhrB (SMc04462) as their representatives, respectively, for further experimental studies. We heterologously expressed and purified *Sm*GhrA and *Sm*GhrB as described in Materials and Methods. Using NADP(H) as a cofactor, we performed an activity screening with 24 selected compounds at concentrations expected to be saturating for physiological substrates (Tables S1 and S2). The selected compounds included seventeen 2-ketocarboxylic acids and seven 2-hydroxycarboxylic acids; consequently, both reductase and dehydrogenase activities were tested. The enzyme activity was observed for at

least one enzyme with thirteen 2-ketocarboxylic acids. No detectable activity with reduced substrates was observed (Table S2), suggesting that these enzymes act only as reductases.

Steady-state Kinetics for Selected Substrates. To explore detailed kinetic characteristics of both enzymes, we performed steady-state kinetic assays with selected compounds that resulted in the most prominent activity in the preliminary screening and with pyruvate, which showed poor activity in the screening yet is a substrate for some GHRB enzymes (Table S1). After determining the initial rates of the reaction, we attempted to fit the data to both hyperbolic and sigmoidal curves, by calculating parameters of the Michaelis-Menten and Hill equations with the nonlinear least-squares method. Using an ANOVA F-test at $p = 0.01$, we assessed which model better fit the data. The kinetic parameters (K_m or $K_{0.5}$, generally expressed as $K_{0.5}$, turnover number k_{cat}, and $k_{cat}/K_{0.5}$) were calculated only for substrates with relatively high affinity to the enzyme (i.e., with $K_{0.5}$ values falling within the tested concentration ranges, Table 1, Figure S2).

For each enzyme, we were able to obtain catalytic parameters for four substrates: glyoxylate, hydroxypyruvate, phenylpyruvate, and pyruvate for *Sm*GhrA and glyoxylate, hydroxypyruvate, phenylpyruvate, and 2-keto-D-gluconate for *Sm*GhrB (Table 1, Figure S2). Glyoxylate and hydroxypyruvate, which are known substrates for other GHRA representatives (e.g., *Ec*GhrA and *Re*GxrA) and most GHRB members, are two of the most efficient substrates for both *Sm*GhrA and *Sm*GhrB. *Sm*GhrA exhibits its highest catalytic efficiency ($k_{cat}/K_{0.5}$) with glyoxylate due to its high k_{cat}, albeit with slightly lower affinity (as indicated by the respective K_m or $K_{0.5}$ values). *Sm*GhrB shows lower catalytic effectiveness with glyoxylate due to a poor $K_{0.5}$ for this substrate. Conversely, *Sm*GhrB exhibits its highest catalytic effectiveness with hydroxypyruvate, with an unusually high k_{cat} and ordinary $K_{0.5}$. The enzymes show similar catalytic parameters with respect to phenylpyruvate. The main difference observed between the enzymes is that only *Sm*GhrA

reduces pyruvate, with K_m = 9.7 mM, whereas only *Sm*GhrB reduces 2-keto-D-gluconate, with K_m = 11 mM. The broad substrate specificities, similar catalytic efficiencies, and rather low affinities for even the most prominent substrates suggest that both studied proteins are enzymatic generalists with, to some extent, redundant functionalities.

Interestingly, all but three steady-state reactions exhibit standard Michaelis-Menten behavior. The Hill equation better fits the observed data ($p < 0.01$) for both *Sm*GhrA and *Sm*GhrB with glyoxylate (n=1.83 and 1.48, respectively) and for *Sm*GhrB with hydroxypyruvate (n=1.74). This observation suggests that the efficient substrates of both enzymes also positively regulate their activities.

Due to the high absorption of 4-hydroxyphenylpyruvate at 340 nm, we were not able to reliably measure initial velocities of the reduction reaction. Preliminary measurements of fluorescence at 340 nm suggested that 4-hydroxyphenylpyruvate may be an efficient substrate for both *Sm*GhrA and *Sm*GhrB (data not shown).

Cofactor Specificity. The preference towards NADH or NADPH was tested using one of the most prominent substrates for both enzymes, hydroxypyruvate, at a saturating concentration (66.7 mM) and a cofactor at 0.4 mM concentration. Under the conditions tested, both enzymes exhibited 2–3% activity with NADH as compared to NADPH (data not shown). This observation is supported by the results of thermal shift assays: NAD^+ and NADH did not stabilize the enzymes, while $NADP^+$ and NADPH resulted in increases of melting temperatures by 2.0–5.5 ˚C (Table S3). Steady-state kinetic assays showed that *Sm*GhrA has a six-fold smaller K_m for the preferential cofactor NADPH than *Sm*GhrB (Table 1, Figure S1), suggesting stronger cofactor binding by *Sm*GhrA.

Overall Fold and Core Catalytic Residues of *Sm*GhrA and *Sm*GhrB. To gain insight into structural differences between the two enzymes, we determined three structures for *Sm*GhrA and six structures for *Sm*GhrB at 1.7–2.4 Å resolution, both in *apo*-forms and in various complexes (Tables 2 and 3). The refined models have good overall geometry with a very low percentage of rotamer outliers and low MolProbity clashscore.

In the solved crystal structures, both enzymes are symmetrical homodimers composed of 319 and 322 amino acids (Figure 3). PISA server analysis[49] showed the average dimer interface area for the cofactor-bound forms to be 2183 Å^2 for *Sm*GhrA and 2268 Å^2 for *Sm*GhrB; this, together with the size-exclusion chromatography results (data not shown), strongly suggests that the proteins also exist as homodimers in solution. The monomers comprise two globular α/β/α domains: a coenzyme-binding domain (residues 96-286 in *Sm*GhrA; 95-281 in *Sm*GhrB) and a catalytic domain (1-95 and 287-319 in *Sm*GhrA; 1-94 and 282-322 in *Sm*GhrB). The coenzyme-binding domain of the classical NAD(P)H-binding Rossmann fold acts as a dimerization domain. The substrate and cofactor bind in the cleft between those two domains, causing the cleft closure (Figure 3). The cofactor is bound mostly to the coenzyme domain in a manner canonical for Rossmann-fold enzymes (Figure 3 and 4), as described previously.[50]

The core catalytic residues typical for 2HADHs—Arg234/230 (binding and charge compensation; residues refer to *Sm*GhrA/*Sm*GhrB, respectively), His282/277 (binding and proton transfer), Glu263/259 (maintenance of the protonated state of His282/277), and Ala74/Val72-Gly75/73 (binding of the carboxyl group via main-chain)[51]—are spatially conserved in the *Sm*GhrA and *Sm*GhrB structures (Figure 5).

Structures of *Sm*GhrA. Cofactor-bound forms of *Sm*GhrA (complexes with $NADP^+$ and sulfate, NADPH and oxalate) have only one protein chain in the asymmetric unit; the dimer in

these structures is formed via the crystallographic two-fold axis (Table S2). With the exception of three N-terminal residues, all residues, including the C-terminal residue Tyr319, were modeled into the electron density maps. The cofactor-bound structures of *Sm*GhrA are highly similar to each other, with Cα atom root-mean-square deviation (RMSD) of 0.21 Å and maximum Cα deviation of 0.69 Å. In addition to binding by the coenzyme-binding domain, the cofactor has two strong interactions with the catalytic domain. The OH group of the C-terminal residue Tyr319 forms a hydrogen bond with the 2′-phosphate group, and the positively charged side-chain of Arg92 forms at least two hydrogen bonds with the pyrophosphate moiety and serves as the charge compensator (Figure 4).

Superposition of the two cofactor-bound structures shows that despite the high similarity of the two cofactor-bound complexes, there are small adjustments in the active site depending on the ligand. In the complex with sulfate, the side-chain of Trp53 moves down by 0.50 Å, and that of Arg234 moves away by 0.40 Å in order to accommodate the sulfate, which is a "thicker" ligand than oxalate (Figure 5A,B).

The structure with oxalate displays a three-point fixation of this substrate analog, where all three oxygen atoms that are equivalent to the core of the substrate (carboxyl and keto groups) are bound by the protein (Figure 5A). This mode of binding is similar to that in the *Sm*GhrB complex with $NADP^+$ and 2-keto-D-gluconate (Figure 5E) and in the *Hs*GhrB complex with $NADP^+$ and hydroxypyruvate (Figure 5F). These structures suggest that the position of the actual substrate in *Sm*GhrA should be similar to those in *Sm*GhrB and *Hs*GhrB. The carboxyl group of the substrate will be bound by the hydrogen bonds with the main-chain nitrogen atoms of Gly73 and Gly75 from the catalytic domain, as well as with the guanidine group of Arg234 from the coenzyme-binding domain. The variable radical of the 2-ketocarboxylic acid will be located at the position

of the unbound oxygen atom. The keto group of the substrate will be bound by the hydrogen bonds with Arg234 and His282 from the coenzyme-binding domain. Thus, the substrate serves as a linker between the domains.

The *apo*-form of *Sm*GhrA has two protein chains in the asymmetric unit. The catalytic domain is highly disordered: it was possible to build only ~30% of the catalytic domain residues in chain A and just a few residues in chain B (Figure 3A). The catalytic domain occupies several different conformations in each subunit (which can be easily accommodated by the crystal lattice), and only a major conformation is visible in chain A. Although the *apo*-form could not be fully reconstructed, a comparison to the cofactor-bound form shows a major rotation of the catalytic domain upon NADPH binding (~30˚, Figure 3A,C and Figure S3A). At the same time, there are no significant changes observed in the coenzyme-binding domain upon NADP(H) binding, suggesting that the domain is preorganized for cofactor binding.

Interestingly, in the structures of *Sm*GhrA complexes with $NADP^+$ and sulfate and NADPH and oxalate, the same small ligands were found in the interdomain interface as in the active site (sulfate and oxalate, Figure 3C). The binding of these ligands might cause the potential allosteric regulation, suggested by the sigmoidal kinetic behavior in the reduction of glyoxylate by *Sm*GhrA.

Structures of *Sm*GhrB. All structures of *Sm*GhrB have four protein chains in the asymmetric unit, forming two dimers (Table 2). Unlike *Sm*GhrA, almost all residues, except for the first two N-terminal and the last three C-terminal residues, showed electron density and were modeled in all *Sm*GhrB structures.

The structures of *Sm*GhrB complexed with (1) $NADP^+$ and 2-keto-D-gluconate, (2) NADPH and oxalate, (3) $NADP^+$ and malonate, and (4) 2′-phospho-ADPR are very structurally similar to each other. The complex containing 2′-phospho-ADPR has the same closed conformation as all three

ternary complexes, suggesting that the major domain movement can occur upon cofactor binding even without a substrate present. The pairwise Cα atom superposition of all enzyme subunits from these structures shows RMSD values in the range of 0.19–0.41 Å, with a maximum Cα deviation of 1.20 Å, suggesting substantial rigidity of these complexes. In contrast to *Sm*GhrA, the cofactor is bound solely by the coenzyme-binding domain (Figure 4C,D).

*Sm*GhrB's active site displays the same three-point fixation of the core of the substrate (carboxyl and keto groups) as *Sm*GhrA and *Hs*GhrB, with the same types of residues involved for both the substrate and substrate analogs (Figure 5). The substrate analogs (oxalate and malonate) are bound only via these three points in the same way as in the *Sm*GhrA structures. Remarkably, 2-keto-D-gluconate—the biggest substrate bound in any of the reported 2HADH structures—forms two additional hydrogen bonds. First, the hydroxyl group on the C3 carbon of 2-keto-D-gluconate is bound by the side-chain of Ser280 in the same manner that hydroxypyruvate is bound in the *Hs*GhrB structure (Figure 5F). Second, the hydroxyl group on C5 of this substrate is bound by the side-chain of the catalytic residue Arg230.

Two structures of the *apo*-form of *Sm*GhrB were solved: the unliganded form and the complex with citrate (Table 3). The complex with citrate has the ligand bound in the active site in one of the four subunits. The citrate molecule is located in the place of the nicotinamide moiety of the cofactor, suggesting that it might be a competitive inhibitor of NADPH binding. These cofactor-free structures show that *Sm*GhrB's catalytic domain has some degree of flexibility with respect to its positioning (Figure S3B); this rotation, however, is much smaller than that observed in S*m*GhrA (Figure 3 and Figure S3). Furthermore, *Sm*GhrB displays much higher flexibility of the cofactor-binding pocket in the cofactor-free forms: the loop comprising residues 201–209 occupies different conformations in different complexes and even different subunits, and the loop

comprising residues 228-233 has a different conformation in subunit A of the citrate complex structure (Figure S3B). The catalytic domain is highly similar in all *Sm*GhrB structures, suggesting that it moves as a rigid body upon cofactor binding.

DISCUSSION

In this work, we show a clear separation of enzymes previously known as GHPR into two evolutionary separated subfamilies. In contrast to the previous studies, our maximum-likelihood phylogenetic analysis shows that most GHPRs form two separate clades, GHRA and GHRB. Our kinetic studies of two representative proteins from GHRA and GHRB of *S. meliloti* show broad substrate versatility of these enzymes with similar substrate preferences, albeit with clear differences in specificities for some substrates (e.g., 2-keto-D-gluconate and pyruvate). *Sm*GhrA and *Sm*GhrB exhibit greatest catalytic efficiency (as shown with $k_{cat}/K_{0.5}$) with glyoxylate and hydroxypyruvate, respectively, and none of the enzymes catalyze the reverse reaction, proving that these enzymes are indeed glyoxylate/hydroxypyruvate reductases.

We present nine crystal structures of two hereto unstudied enzymes with a series of substrates/substrate analogs and cofactors along with the enzymes' *apo*-forms. In particular, we determined the crystal structure of *Sm*GhrB with 2-keto-D-gluconate, which is the biggest substrate (by the number of atoms) co-crystallized with any 2HADH. Moreover, the complex with 2-keto-D-gluconate is just the second reported ternary complex of GHRA/GHRB with an actual substrate (not a substrate analog, Table 4). This structure represents a complex with the substrate for the direct reaction (reduction), and a previously reported structure (PDB ID: 2gcg) represents a complex with a substrate for the reverse reaction (oxidation). Apart from these two structures, the PDB contains 16 structures of GHRA/GHRB enzymes with a ligand bound in the active site (including those reported herein). These structures, however, have ligands that are either substrate

analogs (which may bind differently, as shown by PDB ID: 5bqf) or wrongly modelled substrates (PDB ID: 5aow and 5aov, Table 4). To resolve one case of an incorrectly modelled substrate in the active site, we took the liberty of using the deposited structure factors to re-refine and re-deposit the structure 5aow. Our re-refinement (PDB ID: 6bii) shows that the actual ligand is malonate ion, which was present at 1.7 M concentration in the crystallization solution (http://molstack.bioreproducibility.org/c/FdBO/). The structure of *Sm*GhrB complexed with 2-keto-D-gluconate provides further experimental support for the mode of substrate binding observed previously for the complex of *Hs*GhrB with substrate hydroxypyruvate (Figure 5F).[18]

Superposition of the *Sm*GhrA and *Sm*GhrB structures shows substantial structural differences between the overall folds of these enzymes (Figure S4). Most of the differences are within the catalytic domains; coenzyme-binding domains are much more conserved structurally, as supported by the sequence alignment (Figure S5).

In line with differences in catalytic behavior of the two enzymes, substantial differences are also observed within the substrate-binding pocket (Figure 5). The *Sm*GhrA structures reveal a smaller binding site due to the bulky and hydrophobic Trp53 residue. Consequently, 2-keto-D-gluconate—a large compound—is reduced only by *Sm*GhrB and not by *Sm*GhrA. In GHRA, Trp53 is highly conserved, suggesting that a smaller substrate-binding site is a feature of this subfamily. In GHRB, Trp53 corresponds to Ser50 in *Sm*GhrB and Leu59 in *Hs*GhrB, which are located in a highly variable sequence region (Figure S5). This variability suggests higher diversity in substrate specificities in the GHRB subfamily.

Despite high sequence variability, GHRB enzymes possess an active site feature that is characteristic of most members of the subfamily: residue Ser280/Ser296 (*Sm*GhrB/*Hs*GhrB numbering), which binds hydroxypyruvate via its hydroxyl group (Figure 5). This residue is likely

responsible for making hydroxypyruvate the most prominent substrate for *Sm*GhrB, which is a typical case for GHRB.[16] Moreover, it likely accounts for the inactivity of *Sm*GhrB with pyruvate, due to potential clashes between serine's side-chain and pyruvate's methyl group. In *Sm*GhrA, Ser280 corresponds to the well-conserved Ala285, suggesting that weak activity with pyruvate can be preserved within the GHRA subfamily. *Sm*GhrA substrate-binding site also has a unique feature: the bulky Met102 and Leu150 residues (Val100 and Ile153 in *Sm*GhrB) on the side opposite the nicotinamide moiety of the cofactor. These residues are highly conserved in GHRA (Figure S5) and are likely involved in increasing the catalytic efficiency of these enzymes by pushing the cofactor toward the substrate.

*Sm*GhrA and *Sm*GhrB display strong preferences for NADP(H) over NAD(H), as shown by the thermal shift and kinetic data. As revealed by the crystal structures, the preference is enabled by the S(T)RS(T)XR(K) motif (Figure 4B,D). Ser169 in *Sm*GhrA and Thr172 in *Sm*GhrB are neutrally charged and not bulky, allowing for the accommodation of the 2′-phosphate group. Two positively charged arginine residues near the negatively charged 2′-phosphate moiety (Arg170 and Arg173 in *Sm*GhrA; Arg173 and Arg176 in *Sm*GhrB) assist NADP(H) binding via charge neutralization and hydrogen bonds. In addition, Ser171 in *Sm*GhrA and Thr174 in *Sm*GhrB form hydrogen bonds with the phosphate group.

The multiple sequence alignment suggests that this preference for NADPH is highly ubiquitous for the GHRA subfamily due to the conservation of the S(T)RS(T)XR(K) motif (Figure 4B,D and Figure S5). This cofactor preference agrees with the previously reported data for *Ec*GhrA[16] but not for *Re*GxrA, which was shown to reduce substrates only with NADH.[10] The sequence of *Re*GxrA (UniProt ID: C1JH53_RHIET), however, has the characteristic SRTRK motif (Figure 4A). Additionally, crystal structures of a similar GHRA enzyme from *R. etli* with 84% identity to

*Re*GxrA have bound NADP(H) (PDB IDs: 5bqf, 4xcv, and 5tds), which is inconsistent with the original publication.[10] The same motif is also common for the GHRB subfamily; however, several GHRB enzymes have Thr172 of *Sm*GhrB substituted by Asp—a typical mutation responsible for specificity for NAD(H) versus NADP(H).[52] This phenomenon is always accompanied by mutations of both aforementioned arginine residues, further pinpointing the specificity toward NADH for these GHRB members. Thus, our kinetic, structural, and sequence data suggest that GHRA enzymes are highly specific for NADPH, whereas the GHRB subfamily includes members that are specific for either NADH or NADPH.

The *apo*-form protein structures suggest that the interdomain cleft opening is at least twice wider in *Sm*GhrA than in *Sm*GhrB (Figure 3), which is a general trend for GHRA and GHRB enzymes. Higher *Sm*GhrA flexibility is associated with the catalytic domain involvement in the cofactor binding via residues Arg92 and Tyr319 (Figure 4A), which may assist with the closure of the wide-open interdomain cleft prior to catalysis. The residues are highly conserved within the GHRA subfamily, suggesting that this mode of the cofactor binding is common for the GHRA enzymes (Figure S5). In addition, the results of steady-state kinetics suggest stronger cofactor binding by *Sm*GhrA as compared to *Sm*GhrB (Table 1), which may also be caused by the involvement of the catalytic domain in the cofactor binding. The stronger cofactor binding may be a feature of the GHRA subfamily, but this hypothesis is yet to be proven by comparison of binding data. On the contrary, the catalytic domain is not involved in the cofactor binding in *Sm*GhrB, and the coenzyme-binding domain residue Arg152, which binds the pyrophosphate moiety of the cofactor in *Sm*GhrB instead of Arg92 in *Sm*GhrA (Figure 4C), is highly conserved within the GHRB subfamily.

The broad and overlapping substrate specificities of the GHRA and GHRB subfamilies suggest that their functional specializations might have arisen from acquired differences in gene expression and subsequent adaptation to specific roles in metabolic pathways. The genome of *S. meliloti* 1021 encodes 16 2HADH members, including two GHRA and seven GHRB members, some of which result from very recent duplications (Figure 2). *Sm*GhrA is a highly similar ortholog of the previously characterized *Re*GxrA protein (61% sequence identity). Their corresponding genes are potentially co-expressed with genes encoding the Yej transporter, a member of the ABC superfamily of transporters (Figure S6). The gene encoding *Sm*GhrB, similarly to its close orthologs from Rhizobiales, is located downstream to *pckR*, and together, they presumably form an operon (Figure S6).[53] PckR is a LacI-like transcription factor predicted to be involved in the regulation of central carbohydrate metabolism.[54] Thus, the genomic data suggest that *Sm*GhrA and *Sm*GhrB are expressed under different conditions, adopting different metabolic or signaling roles in the cell.

Cumulatively, the division of GHPR into GHRA and GHRB subfamilies is strongly supported by the conserved structural features of these enzymes that can be linked to the observed differences in substrate preferences. Thus, our work advances the systematic classification of the 2HADH family, presents broad kinetic and structural characterizations of the two selected members, and constitutes the first systematic effort to investigate differences between the GHRA and GHRB subfamilies. Further studies are required to understand why GHRA and GHRB have remained generalists over the course of evolution and how their functions differ *in vivo*.

FIGURES

Figure 1. Reaction catalyzed by D-2-hydroxyacid dehydrogenases.

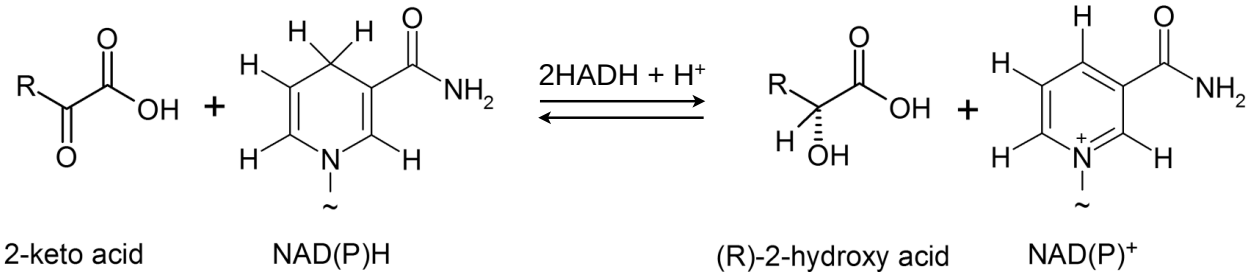


Figure 2. Maximum-likelihood evolutionary tree of the 2HADH superfamily. The unrooted tree was constructed with FastTree2, based on multiple alignment of sequences comprising both cofactor-binding and catalytic domains. The branch labels correspond to locus tags of proteins from *S. meliloti* ("SM" prefix) and UniProt accessions of proteins with studied substrate specificities. Orange dots correspond to proteins with solved crystal structures. Proteins studied in this work, from the GHRA clade: *Sm*GhrA (SMc02828) and from GHRB clade: *Sm*GhrB (SMc04462), are indicated with arrows. Sequence clusters that represent separate enzymatic subgroups are shaded. The scale bar represents the number of estimated changes per position.

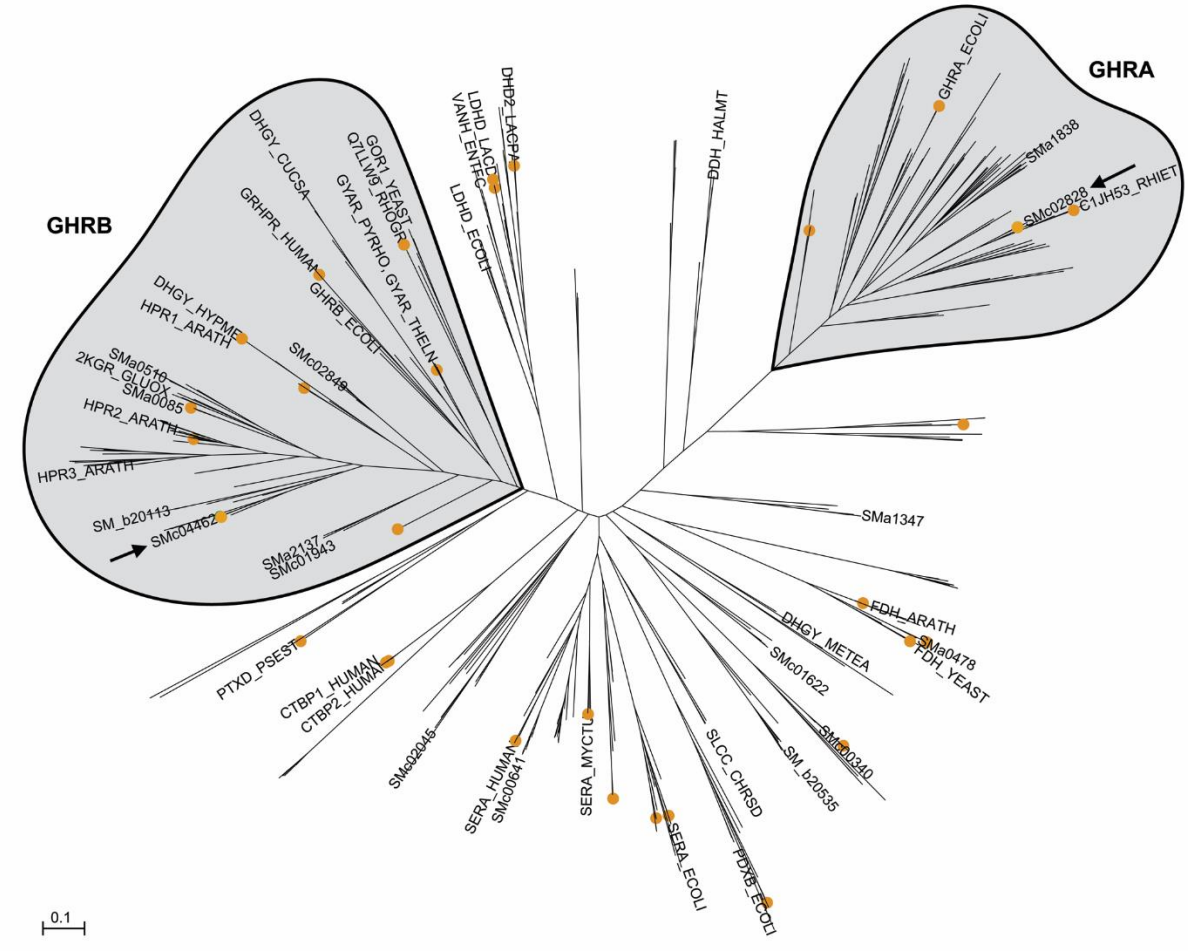


Figure 3. *Sm*GhrA and *Sm*GhrB: overall structure and catalytic domain movement upon the cofactor binding. (A, B) *Apo*-forms of *Sm*GhrA and *Sm*GhrB, (C, D) *Sm*GhrA and *Sm*GhrB in complex with NADPH and oxalate. Dimers of *Sm*GhrA and *Sm*GhrB are shown along the twofold intermolecular axis. The coenzyme-binding and catalytic domains are shown in blue and magenta, respectively. The NADPH and oxalate molecules are represented as cylinder models with carbon atoms in green, oxygen in red, and phosphorus in orange. Arrows in the upper panel show the catalytic domain movement upon the cofactor binding.

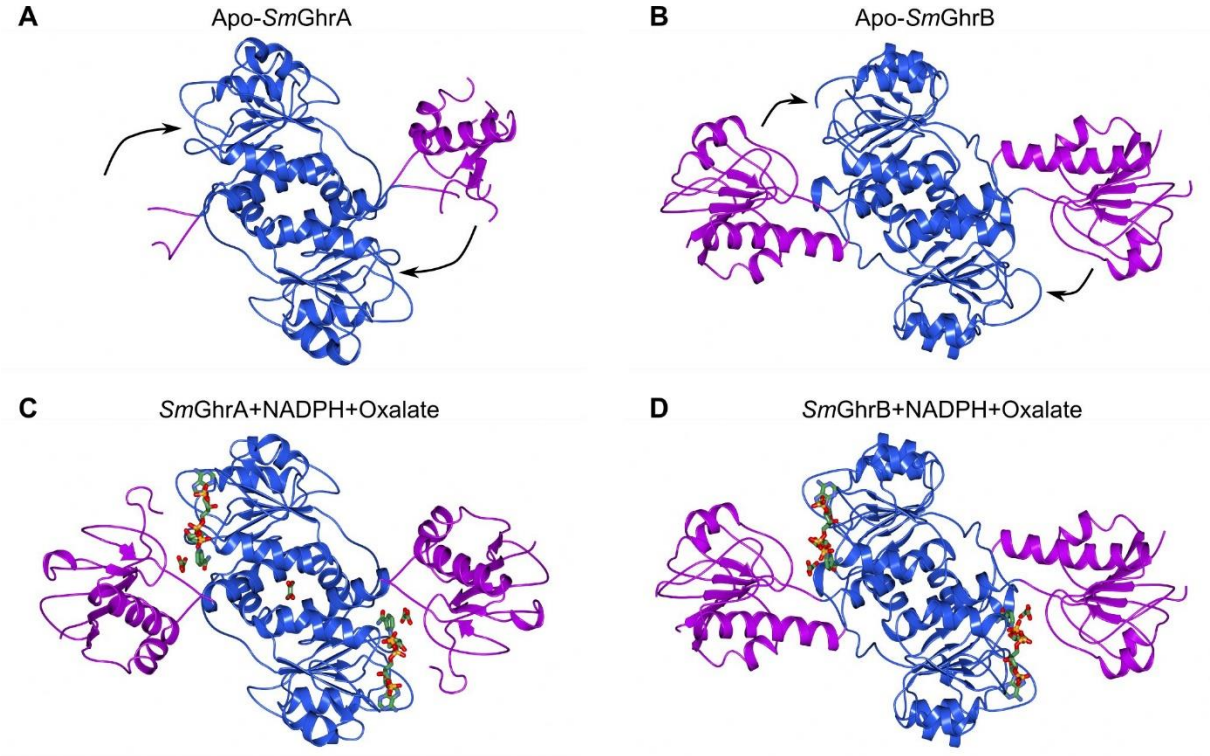


Figure 4. Comparison of the cofactor-binding environment in *Sm*GhrA and *Sm*GhrB. (A, C) Residues forming hydrogen bonds with the cofactor. (B, D) 2′-phosphate binding site formed by the S(T)RS(T)XR(K) motif (non-conserved residue X is not shown). The cofactor is shown in stick representation; oxygen atoms are depicted in red, carbon in green, nitrogen in blue, and phosphorus in orange. Amino acid residues involved in the cofactor binding by hydrogen bonding are shown in a similar color schema with carbon depicted in gray. Hydrogen bonds are indicated as red dashed lines. Residues from the coenzyme-binding domain are labeled in black, and residues from the catalytic domain are labeled in magenta.

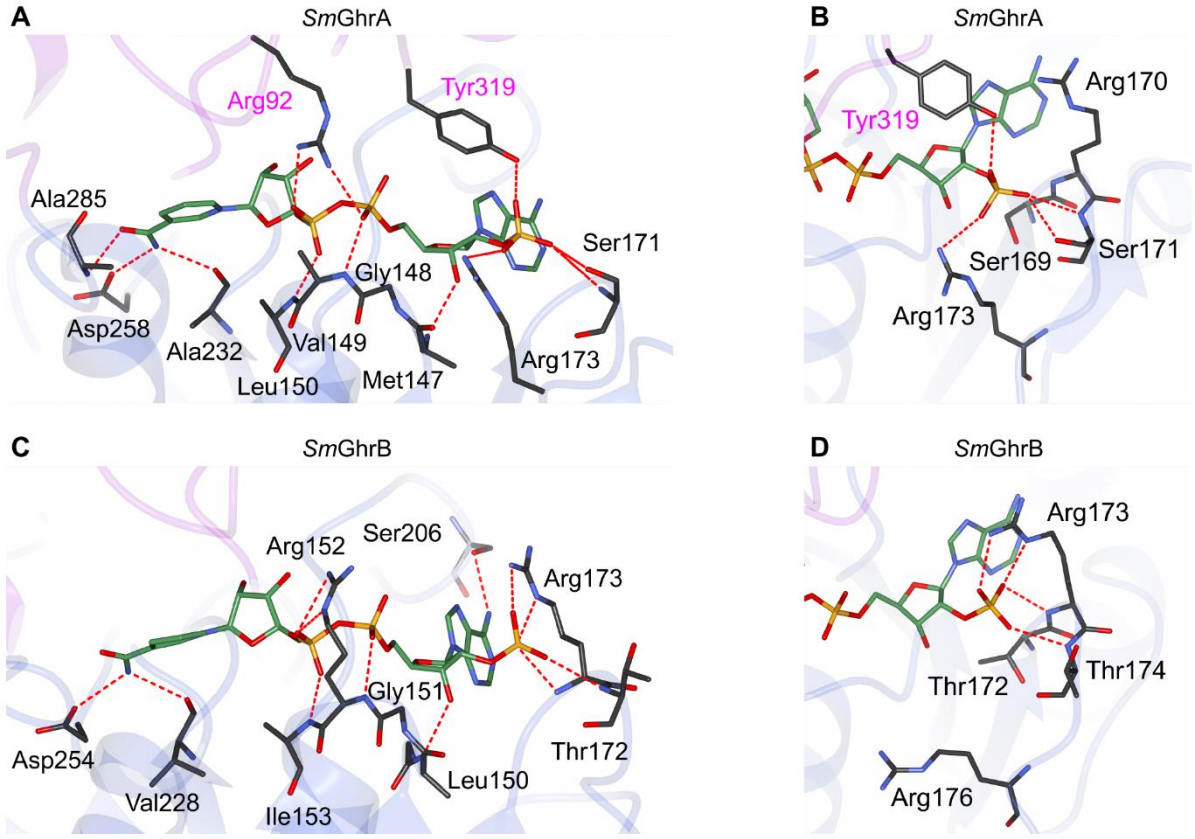
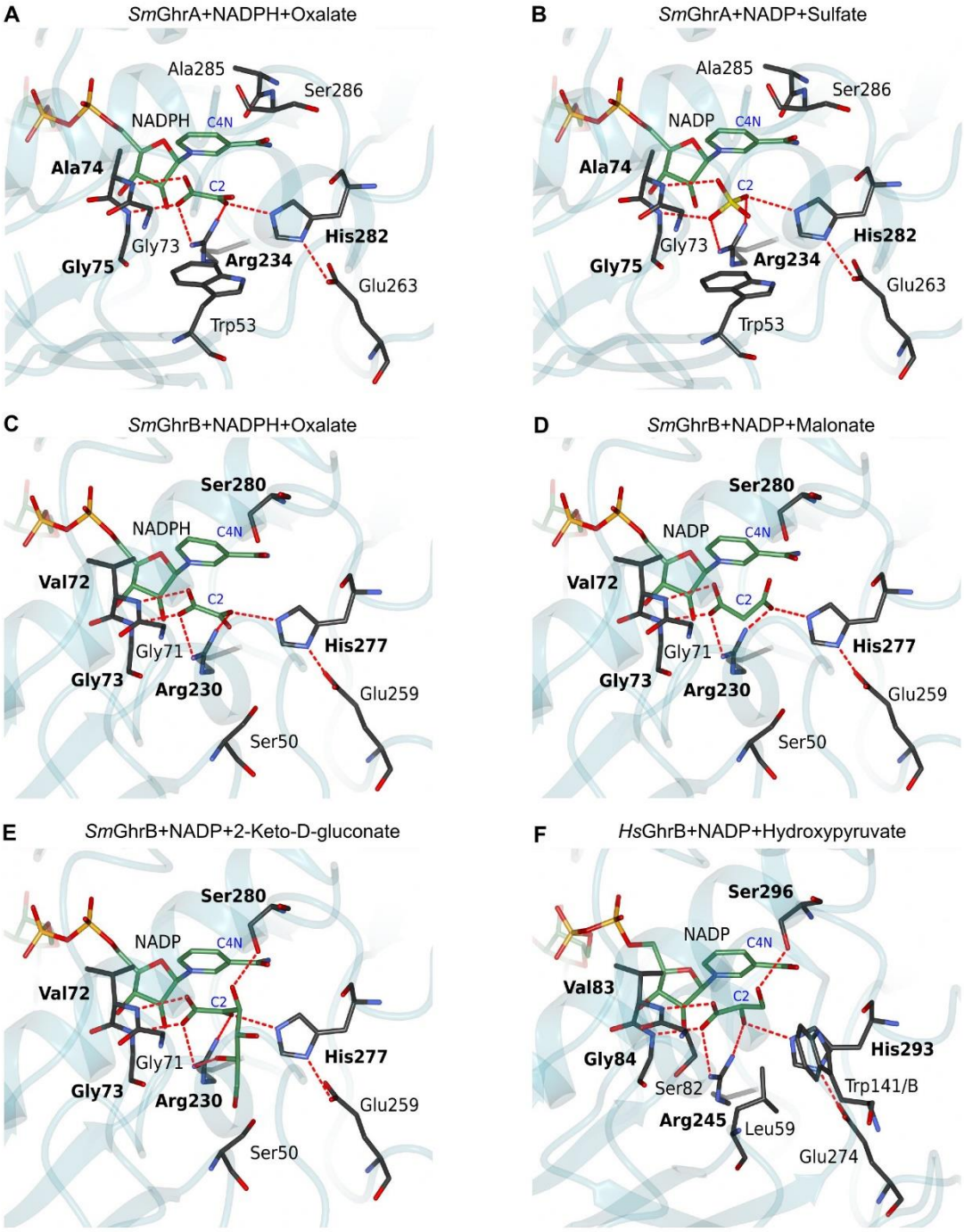


Figure 5. Comparison of the catalytic sites of ternary complexes. *Sm*GhrA (A, B), *Sm*GhrB (C, D, E), and *Hs*GhrB (PDB ID: 2gcg) (F). NADP(H), substrates, and substrate analogs are shown in stick representation; oxygen atoms are in red, carbon in green, nitrogen in blue, phosphorus in orange. The catalytic site residues are shown in a similar fashion with carbon depicted in gray. Hydrogen bonds are indicated as red dashed lines; residues involved in the substrate binding via hydrogen bonds are labeled in bold. C4N atom of NADPH (the donor of the hydride ion) and C2 atom of substrates or substrate analogs (representing the acceptor of the hydride ion) are labeled in blue. The electron density maps, including omit maps, can be inspected interactively using Molstack (http://molstack.bioreproducibility.org/c/FdBO/).


A
SmGhrA+NADPH+Oxalate
Ala285
Ser286
NADPH
C4N
Ala74
C2
Gly73
His282
Gly75
Arg234
Glu263
Trp53
B
SmGhrA+NADP+Sulfate
Ala285
Ser286
NADP
C4N
Ala74
C2
Gly73
His282
Gly75
Arg234
Glu263
Trp53
C
SmGhrB+NADPH+Oxalate
Ser280
NADPH
C4N
Val72
C2
Gly71
His277
Gly73
Arg230
Glu259
Ser50
D
SmGhrB+NADP+Malonate
Ser280
NADP
C4N
Val72
C2
Gly71
His277
Gly73
Arg230
Glu259
Ser50
E
SmGhrB+NADP+2-Keto-D-gluconate
Ser280
NADP
C4N
Val72
C2
Gly71
His277
Gly73
Arg230
Glu259
Ser50
F
HsGhrB+NADP+Hydroxypyruvate
Ser296
NADP
C4N
Val83
C2
His293
Gly84
Ser82
Trp141/B
Arg245
Leu59
Glu274

TABLES

Table 1. Kinetic parameters of *Sm*GhrA and *Sm*GhrB for different substrates and cofactors [a]

Substrate	$K_{0.5}$ (mM)		$k_{cat}/K_{0.5}$ (M^{-1} s^{-1})		Kinetic model	
	SmGhrA	SmGhrB	SmGhrA	SmGhrB	SmGhrA	SmGhrB
Glyoxylate	12.3 ± 0.46	80.0 ± 20	$(1.72 \pm 0.034) \times 10^3$	$(8.15 \pm 0.56) \times 10^2$	Hill (n = 1.83)	Hill (n = 1.48)
Pyruvate	9.69 ± 0.80	-	$(5.73 \pm 0.34) \times 10^1$	-	M-M	-
Hydroxypyruvate	9.35 ± 0.86	6.14 ± 0.58	$(1.20 \pm 0.061) \times 10^3$	$(9.31 \pm 0.38) \times 10^3$	M-M	Hill (n = 1.74)
2-Keto-D-gluconate	-	11.0 ± 2.0	-	$(9.35 \pm 0.84) \times 10^2$	-	M-M
Phenylpyruvate	6.58 ± 0.73	5.59 ± 0.92	$(2.67 \pm 0.18) \times 10^2$	$(3.72 \pm 0.38) \times 10^2$	M-M	M-M
NADPH	0.0105 ± 0.0023	0.0590 ± 0.013	$(3.71 \pm 0.59) \times 10^5$	$(7.73 \pm 1.54) \times 10^5$	M-M	M-M

[a] The parameters were obtained for purified enzymes either in a saturating concentration of NADPH (0.4 mM) and different concentration of the substrates or in a saturating concentration of hydroxypyruvate (66.7 mM) and different concentration of NADPH. The reactions for which we observed very weak activity within the substrate concentration ranges (i.e., not allowing for determination of K_m or $K_{0.5}$) are marked with dashes. For the kinetic plots, see Figure S1 and S2.

Table 2. Summary of crystallization conditions and data collection, processing, and refinement statistics for deposited structures of *Sm*GhrA[b]

*Sm*GhrA PDB ID	4weq	4z0p	5unn
Complex	$NADP^+$ and Sulfate	NADPH and Oxalate	*Apo* unliganded
Diffraction images DOI	10.18430/M3D011	10.18430/M3XS3C	10.18430/M35UNN
Crystallization			
Protein stock conc. (mg/ml)	11.7	11.7	11.7
Ligands added	5 mM $NADP^+$	5 mM NADPH 50 mM Oxalic acid pH 7.0	none
Protease added	chymotrypsin	rTEV	rTEV
Precipitant mix	0.1 M Bis-Tris pH=5.5, 25% w/v PEG 3350, 0.2 M Ammonium Sulfate (MCSG Suite 1 #6)	0.2 M Ammonium citrate Tribasic pH=7.0, 20% w/v PEG 3350 (MCSG Suite 2 #28)	0.1 M Sodium citrate pH=5.6, 20% v/v 2-Propanol, 20% w/v PEG 4000 (MCSG Suite 1 #88)
Cryoprotection	25% Glycerol	drying over 1 M Sodium chloride	drying over 1 M Sodium chloride
Data processing			
Resolution (Å)	50.00 – 2.00 (2.03 – 2.00)	50.00 – 1.70 (1.73 – 1.70)	50.00 - 2.00 (2.03 - 2.00)
Beamline	19-ID	21-ID-G	21-ID-G
Wavelength (Å)	0.979	0.979	0.979
Space group	$P3_221$	$P3_221$	$I4_1$
Unit cell: a, b, c (Å)	108.2, 108.2, 80.7	108.1, 108.1, 80.2	128.6, 128.6, 122.9
Protein chains in the asymmetric unit	1	1	2
Completeness (%)	100 (100)	99.9 (100)	100.0 (100.0)
Number of unique reflections	37017	60027	67353
Redundancy	12.2 (12.2)	6.9 (6.7)	6.9 (6.6)
$\langle I \rangle / \langle \sigma(I) \rangle$	41.5 (3.2)	35.4 (2.4)	18.5 (1.7)
CC ½ - highest resolution shell	(0.92)	(0.73)	(0.62)
R_{merge} [#]	0.069 (0.846)	0.074 (0.852)	0.108 (0.902)
Wilson B factor (Å^2)	38.5	25.7	28.5
Structure refinement			
R_{work} / R_{free} [‡]	0.154/0.196	0.134/0.150	0.147/0.174
Mean ADP value (Å²)		31	44
Number of protein atoms	2426	2442	3437
Mean ADP for protein (Å²)	49	28	43
Number of water molecules	250	429	659
Mean ADP for water (Å²)	56	45	50
Median ADP (Occupancy) of the ligand in the active site, by chain	48 (1)	24 (1)	-
Bond lengths rmsd (Å)	0.015	0.010	0.011
Bond angles rmsd (°)	1.8	1.5	1.4
Rotamer outliers, (%)	0	0	0
Ramachandran outliers (%)[§]	0	0	0
Ramachandran favored (%)	99.0	99.0	98.4
Clashscore/Clashscore percentile (%)	2.22/100	1.00/99	0.29/100
MolProbity score	1.00	0.80	0.61

[b] Values in parentheses are for the highest resolution shell.

[#] $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where ‹I(hkl)› is the mean of I observations I_i(hkl) of reflection hkl.

‡$R = \sum ||F_o| - |F_c|| / \sum |F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively, calculated for working set (R_{work}) and for 5% of the reflections omitted from refinement (R_{free}).

§Ramachandran plot statistics are calculated by MolProbity.

Table 3. Summary of crystallization conditions and data collection, processing, and refinement statistics for deposited structures of *Sm*GhrB[c]

*Sm*GhrB PDB ID	5v7n	5v7g	5v6q	5j23	5uog	5v72
Complex	$NADP^+$ and 2-Keto-D-gluconic acid	NADPH and Oxalate	$NADP^+$ and Malonate	2-phospho-ADPR	*Apo* Unliganded	*Apo* Citrate
Diffraction images DOI	10.18430/M35V7N	10.18430/M35V7G	10.18430/M35V6Q	10.18430/M35J23	10.18430/M35UOG	10.18430/M35V72
Crystallization						
Protein stock conc. (mg/ml)	13.2	13.2	13.2	13.2	13.2	13.2
Ligands added	5 mM $NADP^+$ 100 mM 2-Keto-D-gluconic acid hemicalcium salt, pH 7.0	5 mM NADPH 50 mM Oxalic acid, pH 7.0	5 mM $NADP^+$	5 mM NADPH 50 mM Glycolic acid	5 mM NADPH	none
Protease added	rTEV	rTEV	rTEV	rTEV	none	chymotrypsin
Precipitant mix	0.2 M Magnesium formate pH=5.9, 20% w/v PEG 3350 (TOP96 #24)	0.1 M Sodium cacodylate pH=6.5, 18% w/v PEG 8000, 0.2 M Sodium acetate (TOP96 #41)	1.1 M Malonic acid, 0.072 M Succinic acid, 0.15 M Ammonium citrate tribasic, 0.18 M DL-Malic acid, 0.096 M Ammonium tartrate dibasic, 0.24 M Sodium acetate, 0.3 M Sodium formate, pH=7.0 (MCSG Suite 2 #20)		0.1 M Citric acid pH=3.5, 2 M Ammonium sulfate, (MCSG Suite 3 #61)	0.1 M Sodium citrate pH=5.0, 20% w/v PEG 6000 (MCSG Suite 2 #9)
Cryoprotection	drying over 1 M Sodium chloride	drying over 1 M Sodium chloride	Paratone-N	drying over 1M Sodium chloride	20% Ethylene glycol	air-drying
Data processing						
Resolution (Å)	50.00 - 1.75 (1.78 - 1.75)	50.00 - 1.75 (1.78 - 1.75)	50.00 - 1.95 (1.98 - 1.95)	50.00 - 2.30 (2.34 - 2.30)	50.00 - 2.40 (2.44 - 2.40)	50.00 - 2.10 (2.14 - 2.10)
Beamline	21-ID-F	21-ID-G	21-ID-G	21-ID-G	19-ID	21-ID-G
Wavelength (Å)	0.979	0.979	0.979	0.979	0.979	0.979
Space group	*H3*	*H3*	*H3*	*H3*	*H3*	$P2_1$
Unit cell: a, b, c (Å)	175.9, 175.9, 135.3	178.2, 178.2, 133.8	176.5, 176.5, 135.5	175.7, 175.7, 136.7	176.6, 176.6, 135.0	63.2, 157.9, 64.8
Protein chains in the asymmetric unit	4	4	4	4	4	4
Completeness (%)	99.6 (100.0)	98.8 (99.7)	99.6 (98.6)	100.0 (100.0)	100.0 (100.0)	98.4 (99.1)
Number of unique reflections	156897	157652	114355	70083	61899	68211
Redundancy	4.9 (4.9)	4.5 (4.3)	4.6 (4.5)	3.9 (3.9)	3.2 (3.2)	3.2 (3.2)
$\langle I \rangle / \langle \sigma(I) \rangle$	19.8 (1.8)	17.9 (2.1)	22.8 (2.1)	18.2 (1.8)	22.7 (2.0)	12.5 (2.3)
CC ½ - highest resolution shell	(0.61)	(0.78)	(0.69)	(0.68)	(0.69)	(0.79)
R_{merge}	0.113 (0.878)	0.081 (0.667)	0.105 (0.691)	0.077 (0.662)	0.061 (0.630)	0.096 (0.431)
Wilson B factor ($Å^2$)	28.1	26.4	30.2	45.8	56.9	28.0
Structure refinement						
R_{work} / R_{free}	0.144/0.167	0.150/0.180	0.144/0.157	0.146/0.169	0.136/0.190	0.164/0.204
Mean ADP value ($Å^2$)	33	31	34	47	58	35
Number of protein atoms	9511	9448	9446	9454	9325	9327
Mean ADP for protein ($Å^2$)	32	30	34	47	58	35
Number of water molecules	1350	1809	841	603	972	1151
Mean ADP for water ($Å^2$)	41	41	39	46	54	40
Median ADP (Occupancy) of the ligand in the active site, by chain	26 (1), 28 (1), 35 (1), 38 (1)	28 (0.8), 23 (0.7), 18 (0.8), 25 (1.0)	33 (1), 28 (1), 30 (1), 26 (1)	-	-	-, -, 33 (1), -
Bond lengths rmsd (Å)	0.011	0.011	0.009	0.013	0.009	0.015
Bond angles rmsd (°)	1.5	1.5	1.3	1.5	1.3	1.5
Rotamer outliers (%)	0.80	0.5	0.51	0.60	1.04	1.24
Ramachandran outliers[§] (%)	0.32	0.32	0.08	0.32	0.00	0.00
Ramachandran favored (%)	97.31	97.4	97.4	97.9	97.5	98.2
Clashscore/Clashscore percentile (%)	0.51/100	0.31/100	0.41/100	0.87/1000	0.69/100	0.69/100
MolProbity score	0.81	0.73	0.77	0.80	0.83	0.80

[c] Data prepared as for Table 2.

Table 4. Overview of GHRA and GHRB crystal structures in the closed conformation with a ligand bound in the active site[d]

2HADH subfamily	Organism	UniProt Accession ID	Cofactor bound	Ligand in the substrate-binding site	Year deposited	Resolution, Å	PDB ID	Reference	Comment about the ligand in the substrate-binding site
GHRA	*Rhizobium etli*	Q2KDT2	NADPH	oxalic acid	2016	1.90	5tsd	-	CSA+
			$NADP^+$	(2R,3R)-tartaric acid	2015	1.45	5bqf	-	CSA (bound differently) (http://molstack.bioreproducibility.org/c/FdBO/)
	Rhodobacter sphaeroides	Q3IWN8	$NADP^+$	sulfate ion	2015	1.85	4zqb	-	RSA
	Xanthobacter autotrophicus	A7IIH0	NADPH	2-(N-morpholino)-ethanesulfonic acid (MES)	2017	2.20	5vg6	-	Buffer molecule
	Rhizobium meliloti	Q92T34	$NADP^+$	sulfate ion	2014	2.00	4weq	This paper	RSA Figure 5B
			NADPH	oxalic acid	2015	1.70	4z0p	This paper	CSA+ Figure 5A
GHRB	*Homo sapiens*	Q9UBQ7	NADPH	(2R)-2,3-dihydroxypropanoic acid (hydroxypyruvate)	2006	2.20	2gcg	[18]	Substrate Figure 5F
	Pseudomonas stutzeri	O69054	NAD^+	sulfite ion	2012	1.95	4e5k	[2]	RSA
			NAD^+	sulfate ion	2013	2.65	4nu6	[47]	RSA
	Pyrococcus furiosus	Q8U3Y2	$NADP^+$	glyoxylic acid?	2015	1.40	5aov	[13]	Substrate? (http://molstack.bioreproducibility.org/c/FdBO/)
	Pyrococcus horikoshii	O58320	$NADP^+$	sulfate ion	2005	1.70	2dbq	[19]	RSA
			$NADP^+$	sulfate ion	2005	2.61	2dbr	[19]	RSA
			$NADP^+$	sulfate ion	2005	2.45	2dbz	[19]	RSA
	Pyrococcus yayanosii	F8AEA4	$NADP^+$	(2R)-2,3-dihydroxypropanoic acid Incorrectly modelled	2015	2.00	5aow	[13]	Incorrectly modelled (http://molstack.bioreproducibility.org/c/FdBO/)
				Re-modelled as malonate	2017	2.00	6bii	This paper	RSA (malonate) (http://molstack.bioreproducibility.org/c/FdBO/)
	Rhodotorula graminis	Q7LLW9	NAD^+	sulfate ion	2008	2.50	2w2l	-	RSA
	Rhizobium meliloti	Q92LZ4	$NADP^+$	2-keto-D-gluconic acid	2017	1.75	5v7n	This paper	Substrate Figure 5E
			NADPH	oxalic acid	2017	1.75	5v7g	This paper	CSA+ Figure 5C
			$NADP^+$	malonic acid	2017	1.95	5v6q	This paper	RSA Figure 5D

[d] Ligands are named "close substrate analogs" (CSA) if they have a carboxyl group and an oxygen atom on the second carbon, which warrants a close similarity to an actual substrate for 2HADH (2-keto/2-hydroxy acid). CSA+ denotes a CSA bound in the same way as a substrate. Other substrate analogs are named "remote substrate analogs" (RSA).

ASSOCIATED CONTENT

Supporting Information

The following files are available free of charge.

Steady-state kinetic plots for substrates and NADPH (Figures S1 and S2). Comparison of the protein domains and their structural changes upon cofactor binding (Figures S3 and S4). Multiple sequence alignment of representative sequences of the GHRA and GHRB subfamilies (Figure S5). Genomic context of *Sm*GhrA and *Sm*GhrB (Figure S6). Results of the activity screening for various substrates with NADPH as a cofactor (Tables S1 and S2). Results of the protein thermal shift assay (Table S3). Table of key resources used in this study (Table S4). (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: kwozniak@chem.uw.edu.pl. *E-mail: kginal@cent.uw.edu.pl. *E-mail: wladek@iwonka.med.virginia.edu. Phone: (434) 243-6865. Fax: (434) 243-2981.

Author Contributions

Conceptualization: IGS, KG, KW, and WM; Methodology: IGS, JK, DM, KBH, and OG; Experiments: JK, IGS, DM, KBH, OG, and PS; Data analysis and interpretation: IGS, JK, DM and WM; Writing the original draft: JK, IGS and DM; Review & Editing: IGS, DM, MWG, KG, KW and WM; Visualization: IGS, JK, and DM; Supervision: MWG, KG, KW, and WM; Funding Acquisition, MWG, KG, KW, and WM. All authors read and approved the final manuscript.

†These authors contributed equally.

Funding

The presented work was supported by NIGMS grants U54-GM094662, R01-GM117080, and NIH Big Data to Knowledge (BD2K) grant HG008424. The project was also supported by NCN (Maestro, No. UMO-2012/04/A/ST5/00609; Opus, No. UMO-2014/15/B/NZ1/03357) and FNP (TEAM) grants. Results shown in this report are derived from work performed at Argonne National Laboratory, Structural Biology Center at the Advanced Photon Source. Argonne is operated by UChicago Argonne, LLC, for the U.S. Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817).

Notes

The authors declare that they have no conflicts of interest with the contents of this article.

ACKNOWLEDGEMENTS

We thank David Cooper, Matthew Zimmerman, and Barat Venkataramany for critically reading the manuscript.

ABBREVIATIONS

2HADH, D-2-hydroxyacid dehydrogenase; GHPR, previously delineated glyoxylate/hydroxypyruvate reductase subfamily of 2HADH; GHRA and GHRB, glyoxylate/hydroxypyruvate reductase subfamilies A and B (defined in this work); *Sm*GhrA and *Sm*GhrB, GHRA (SMc02828) and GHRB (SMc04462) from *Sinorhizobium meliloti*; *Hs*GhrB (*Hs*GRHPR), human GHRB; *Ec*GhrA and *Ec*GhrB, GHRA and GHRB from *Escherichia coli*; *Re*GxrA, GHRA from *Rhizobium etli*; IPTG, isopropyl β-D-thiogalactopyranoside; PDB, Protein Data Bank; rTEV, recombinant Tobacco Etch Virus protease.

REFERENCES

(1) Stoll, V. S., Manohar, A. V., Gillon, W., MacFarlane, E. L., Hynes, R. C., and Pai, E. F. (1998) A thioredoxin fusion protein of VanH, a D-lactate dehydrogenase from Enterococcus faecium: cloning, expression, purification, kinetic analysis, and crystallization, *Protein Sci. 7*, 1147-1155.

(2) Zou, Y., Zhang, H., Brunzelle, J. S., Johannes, T. W., Woodyer, R., Hung, J. E., Nair, N., van der Donk, W. A., Zhao, H., and Nair, S. K. (2012) Crystal structures of phosphite dehydrogenase provide insights into nicotinamide cofactor regeneration, *Biochemistry 51*, 4263-4270.

(3) Tarmy, E. M., and Kaplan, N. O. (1968) Kinetics of Escherichia coli B D-lactate dehydrogenase and evidence for pyruvate-controlled change in conformation, *J. Biol. Chem. 243*, 2587-2596.

(4) Kleczkowski, L. A., and Randall, D. D. (1988) Purification and characterization of a novel NADPH(NADH)-dependent hydroxypyruvate reductase from spinach leaves. Comparison of immunological properties of leaf hydroxypyruvate reductases, *Biochem. J. 250*, 145-152.

(5) Cramer, S. D., Ferree, P. M., Lin, K., Milliner, D. S., and Holmes, R. P. (1999) The gene encoding hydroxypyruvate reductase (GRHPR) is mutated in patients with primary hyperoxaluria type II, *Hum. Mol. Genet. 8*, 2063-2069.

(6) Xue, Y. P., Zeng, H., Jin, X. L., Liu, Z. Q., and Zheng, Y. G. (2016) Enantioselective cascade biocatalysis for deracemization of 2-hydroxy acids using a three-enzyme system, *Microb. Cell Fact. 15*, 162.

(7) Patel, R. N. (2013) Biocatalytic synthesis of chiral alcohols and amino acids for development of pharmaceuticals, *Biomolecules 3*, 741-777.

(8) Hummel, W. (1999) Large-scale applications of NAD(P)-dependent oxidoreductases: recent developments, *Trends Biotechnol. 17*, 487-492.

(9) Grant, G. A. (1989) A new family of 2-hydroxyacid dehydrogenases, *Biochem. Biophys. Res. Commun. 165*, 1371-1374.

(10) Fauvart, M., Braeken, K., Daniels, R., Vos, K., Ndayizeye, M., Noben, J. P., Robben, J., Vanderleyden, J., and Michiels, J. (2007) Identification of a novel glyoxylate reductase supports phylogeny-based enzymatic substrate specificity prediction, *Biochim. Biophys. Acta 1774*, 1092-1098.

(11) Fujii, T., Shimizu, M., Doi, Y., Fujita, T., Ito, T., Miura, D., Wariishi, H., and Takaya, N. (2011) Novel fungal phenylpyruvate reductase belongs to d-isomer-specific 2-hydroxyacid dehydrogenase family, *Biochim. Biophys. Acta 1814*, 1669-1676.

(12) Baker, D. P., Kleanthous, C., Keen, J. N., Weinhold, E., and Fewson, C. A. (1992) Mechanistic and active-site studies on D(--)-mandelate dehydrogenase from Rhodotorula graminis, *Biochem. J. 281*, 211-218.

(13) Lassalle, L., Engilberge, S., Madern, D., Vauclare, P., Franzetti, B., and Girard, E. (2016) New insights into the mechanism of substrates trafficking in Glyoxylate/Hydroxypyruvate reductases, *Sci. Rep. 6*, 20629.

(14) Costas, A. M., White, A. K., and Metcalf, W. W. (2001) Purification and characterization of a novel phosphorus-oxidizing enzyme from Pseudomonas stutzeri WM88, *J. Biol. Chem. 276*, 17429-17436.

(15) Kim, K. H., Janiak, V., and Petersen, M. (2004) Purification, cloning and functional expression of hydroxyphenylpyruvate reductase involved in rosmarinic acid biosynthesis in cell cultures of Coleus blumei, *Plant Mol. Biol. 54*, 311-323.

(16) Nunez, M. F., Pellicer, M. T., Badia, J., Aguilar, J., and Baldoma, L. (2001) Biochemical characterization of the 2-ketoacid reductases encoded by ycdW and yiaE genes in Escherichia coli, *Biochem. J. 354*, 707-715.

(17) Duan, X., Hu, S., Zhou, P., Zhou, Y., Liu, Y., and Jiang, Z. (2014) Characterization and crystal structure of a first fungal glyoxylate reductase from Paecilomyes thermophila, *Enzyme Microb. Technol. 60*, 72-79.

(18) Booth, M. P., Conners, R., Rumsby, G., and Brady, R. L. (2006) Structural basis of substrate specificity in human glyoxylate reductase/hydroxypyruvate reductase, *J. Mol. Biol. 360*, 178-189.

(19) Yoshikawa, S., Arai, R., Kinoshita, Y., Uchikubo-Kamo, T., Wakamatsu, T., Akasaka, R., Masui, R., Terada, T., Kuramitsu, S., Shirouzu, M., and Yokoyama, S. (2007) Structure of archaeal glyoxylate reductase from Pyrococcus horikoshii OT3 complexed with nicotinamide adenine dinucleotide phosphate, *Acta Crystallogr. D Biol. Crystallogr. 63*, 357-365.

(20) Jones, K. M., Kobayashi, H., Davies, B. W., Taga, M. E., and Walker, G. C. (2007) How rhizobial symbionts invade plants: the Sinorhizobium-Medicago model, *Nat. Rev. Microbiol. 5*, 619-633.

(21) Galibert, F., Finan, T. M., Long, S. R., Puhler, A., Abola, P., Ampe, F., Barloy-Hubler, F., Barnett, M. J., Becker, A., Boistard, P., Bothe, G., Boutry, M., Bowser, L., Buhrmester, J., Cadieu, E., Capela, D., Chain, P., Cowie, A., Davis, R. W., Dreano, S., Federspiel, N. A., Fisher, R. F., Gloux, S., Godrie, T., Goffeau, A., Golding, B., Gouzy, J., Gurjal, M., Hernandez-Lucas, I., Hong, A., Huizar, L., Hyman, R. W., Jones, T., Kahn, D., Kahn, M. L., Kalman, S., Keating, D. H., Kiss, E., Komp, C., Lelaure, V., Masuy, D., Palm, C., Peck, M. C., Pohl, T. M., Portetelle, D., Purnelle, B., Ramsperger, U., Surzycki, R., Thebault, P., Vandenbol, M., Vorholter, F. J., Weidner, S., Wells, D. H., Wong, K., Yeh, K. C., and Batut, J. (2001) The composite genome of the legume symbiont Sinorhizobium meliloti, *Science 293*, 668-672.

(22) UniProt, C. (2015) UniProt: a hub for protein information, *Nucleic Acids Res. 43*, D204-212.

(23) Katoh, K., and Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability, *Mol. Biol. Evol. 30*, 772-780.

(24) Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses, *Bioinformatics 25*, 1972-1973.

(25) Price, M. N., Dehal, P. S., and Arkin, A. P. (2010) FastTree 2--approximately maximum-likelihood trees for large alignments, *PloS One 5*, e9490.

(26) Han, M. V., and Zmasek, C. M. (2009) phyloXML: XML for evolutionary biology and comparative genomics, *BMC Bioinformatics 10*, 356.

(27) Tong, L., Lee, S., and Denu, J. M. (2009) Hydrolase regulates NAD+ metabolites and modulates cellular redox, *J. Biol. Chem. 284*, 11256-11266.

(28) Zimmerman, M. D., Grabowski, M., Domagalski, M. J., Maclean, E. M., Chruszcz, M., and Minor, W. (2014) Data management in the modern structural biology and biomedical research environment, *Methods Mol. Biol. 1140*, 1-25.

(29) Berman, H. M., Gabanyi, M. J., Groom, C. R., Johnson, J. E., Murshudov, G. N., Nicholls, R. A., Reddy, V., Schwede, T., Zimmerman, M. D., Westbrook, J., and Minor, W. (2015) Data to knowledge: how to get meaning from your result, *IUCrJ 2*, 45-58.

(30) Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006) HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes, *Acta Crystallogr. D Biol. Crystallogr. 62*, 859-866.

(31) Otwinowski, Z., and Minor, W. (1997) [20] Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307-326.

(32) Vagin, A., and Teplyakov, A. (2010) Molecular replacement with MOLREP, *Acta Crystallogr. D Biol. Crystallogr. 66*, 22-25.

(33) Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments, *Acta Crystallogr. D Biol. Crystallogr. 67*, 235-242.

(34) Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains, *Acta Crystallogr. D Biol. Crystallogr. 62*, 1002-1011.

(35) Porebski, P. J., Cymborowski, M., Pasenkiewicz-Gierula, M., and Minor, W. (2016) Fitmunk: improving protein structures by accurate, automatic modeling of side-chain conformations, *Acta Crystallogr. D Biol. Crystallogr. 72*, 266-280.

(36) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr. D Biol. Crystallogr. 53*, 240-255.

(37) Painter, J., and Merritt, E. A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion, *Acta Crystallogr. D Biol. Crystallogr. 62*, 439-450.

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

(39) Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr. D Biol. Crystallogr. 66*, 12-21.

(40) Read, R. J., Adams, P. D., Arendall, W. B., 3rd, Brunger, A. T., Emsley, P., Joosten, R. P., Kleywegt, G. J., Krissinel, E. B., Lutteke, T., Otwinowski, Z., Perrakis, A., Richardson, J. S., Sheffler, W. H., Smith, J. L., Tickle, I. J., Vriend, G., and Zwart, P. H. (2011) A new generation of crystallographic validation tools for the Protein Data Bank, *Structure 19*, 1395-1412.

(41) Grabowski, M., Langner, K. M., Cymborowski, M., Porebski, P. J., Sroka, P., Zheng, H., Cooper, D. R., Zimmerman, M. D., Elsliger, M. A., Burley, S. K., and Minor, W. (2016) A public database of macromolecular diffraction experiments, *Acta Crystallogr. D Biol. Crystallogr. 72*, 1181-1193.

(42) Porebski, P. J., Sroka, P., Zheng, H., Cooper, D. R., and Minor, W. (2017) Molstack - interactive visualization tool for presentation, interpretation, and validation of macromolecules and electron density maps, *Protein Sci.*

(43) Mdluli, K., Booth, M. P., Brady, R. L., and Rumsby, G. (2005) A preliminary account of the properties of recombinant human Glyoxylate reductase (GRHPR), LDHA and LDHB with glyoxylate, and their potential roles in its metabolism, *Biochim. Biophys. Acta 1753*, 209-216.

(44) Timm, S., Nunes-Nesi, A., Parnik, T., Morgenthal, K., Wienkoop, S., Keerberg, O., Weckwerth, W., Kleczkowski, L. A., Fernie, A. R., and Bauwe, H. (2008) A cytosolic pathway for the conversion of hydroxypyruvate to glycerate during photorespiration in Arabidopsis, *Plant Cell 20*, 2848-2859.

(45) Rauch, B., Pahlke, J., Schweiger, P., and Deppenmeier, U. (2010) Characterization of enzymes involved in the central metabolism of Gluconobacter oxydans, *Appl. Microbiol. Biotechnol. 88*, 711-718.

(46) Janiak, V., Petersen, M., Zentgraf, M., Klebe, G., and Heine, A. (2010) Structure and substrate docking of a hydroxy(phenyl)pyruvate reductase from the higher plant Coleus blumei Benth, *Acta Crystallogr D Biol Crystallogr 66*, 593-603.

(47) Hung, J. E., Fogle, E. J., Garg, N., Chekan, J. R., Nair, S. K., and van der Donk, W. A. (2014) Chemical rescue and inhibition studies to determine the role of Arg301 in phosphite dehydrogenase, *PLoS One 9*, e87134.

(48) Domenech, J., and Ferrer, J. (2006) A new D-2-hydroxyacid dehydrogenase with dual coenzyme-specificity from Haloferax mediterranei, sequence analysis and heterologous overexpression, *Biochim. Biophys. Acta 1760*, 1667-1674.

(49) Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J. Mol. Biol. 372*, 774-797.

(50) Shabalin, I. G., Filippova, E. V., Polyakov, K. M., Sadykhov, E. G., Safonova, T. N., Tikhonova, T. V., Tishkov, V. I., and Popov, V. O. (2009) Structures of the apo and holo forms of formate dehydrogenase from the bacterium Moraxella sp. C-1: towards understanding the mechanism of the closure of the interdomain cleft, *Acta Crystallogr. D Biol. Crystallogr. 65*, 1315-1325.

(51) Popov, V. O., and Lamzin, V. S. (1994) NAD(+)-dependent formate dehydrogenase, *Biochem. J. 301*, 625-643.

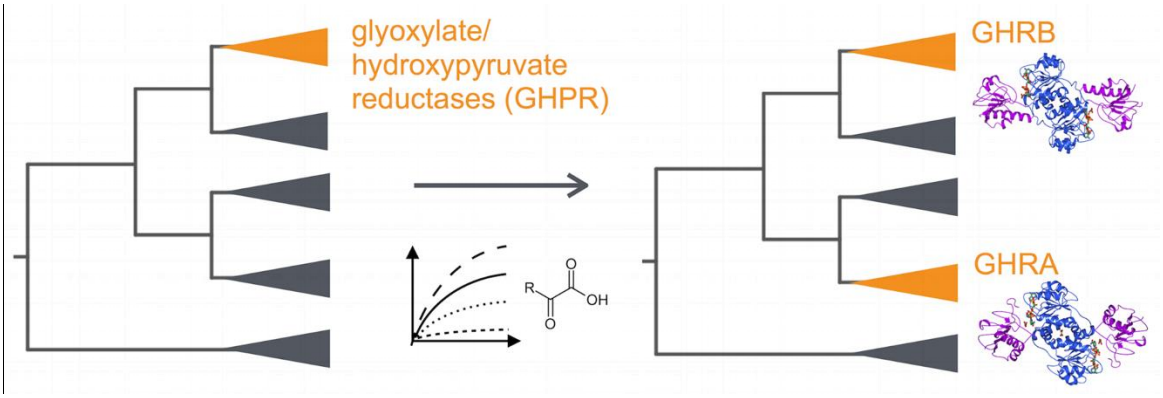
(52) Cahn, J. K., Werlang, C. A., Baumschlager, A., Brinkmann-Chen, S., Mayo, S. L., and Arnold, F. H. (2017) A general tool for engineering the NAD/NADP cofactor preference of oxidoreductases, *ACS Synth. Biol. 6*, 326-333.

(53) Novichkov, P. S., Kazakov, A. E., Ravcheev, D. A., Leyn, S. A., Kovaleva, G. Y., Sutormin, R. A., Kazanov, M. D., Riehl, W., Arkin, A. P., Dubchak, I., and Rodionov, D. A. (2013) RegPrecise 3.0--a resource for genome-scale exploration of transcriptional regulation in bacteria, *BMC Genomics 14*, 745.

(54) Ravcheev, D. A., Khoroshkin, M. S., Laikova, O. N., Tsoy, O. V., Sernova, N. V., Petrova, S. A., Rakhmaninova, A. B., Novichkov, P. S., Gelfand, M. S., and Rodionov, D. A. (2014)

Comparative genomics and evolution of regulons of the LacI-family transcription factors, *Front. Microbiol. 5*, 294.

FOR TABLE OF CONTENTS USE ONLY



Structural, biochemical, and evolutionary characterization of glyoxylate/hydroxypyruvate reductases shows their division into two distinct subfamilies

Jan Kutner[‡,§,†], Ivan G. Shabalin[‡,†], Dorota Matelska[‡,¶,†], Katarzyna B. Handing[‡], Olga Gasiorowska[‡], Piotr Sroka[‡], Maria W. Gorna[§], Krzysztof Ginalski[¶,], Krzysztof Wozniak[§,*], and Wladek Minor[‡,#,*]*