Structure Based Protein Engineering of Aldehyde Dehydrogenase from *Azospirillum brasilense* to Enhance Enzyme Activity against Unnatural 3-Hydroxypropionaldehyde

Hyeoncheol Francis Son¹ and Kyung-Jin Kim¹,²*

¹KNU Institute for Microorganisms, Kyungpook National University, Daegu 41566, Republic of Korea
²School of Life Sciences, BK21 FOUR KNU Creative BioResearch Group, Kyungpook National University, Daegu 41566, Republic of Korea

3-Hydroxypropionic acid (3HP) is a platform chemical and can be converted into other valuable C3-based chemicals. Because a large amount of glycerol is produced as a by-product in the biodiesel industry, glycerol is an attractive carbon source in the biological production of 3HP. Although 8 HP-producing aldehyde dehydrogenases (ALDHs) have been reported so far, the low conversion rate from 3-hydroxypropionaldehyde (3HPA) to 3HP using these enzymes is still a bottleneck for the production of 3HP. In this study, we elucidated the substrate binding modes of the eight 3HP-producing ALDHs through bioinformatic and structural analysis of these enzymes and selected protein engineering targets for developing enzymes with enhanced enzymatic activity against 3HPA. Among ten *Ab*KGSADH variants we tested, three variants with replacement at the Arg281 site of *Ab*KGSADH showed enhanced enzymatic activities. In particular, the *Ab*KGSADH⁷⁵‡ variant exhibited improved catalytic efficiency by 2.5-fold compared with the wild type.

**Keywords:** 3-Hydroxypropionate, aldehyde dehydrogenase, *Azospirillum brasilense*

### Introduction

3HP is one of the top building block chemicals issued in 2004 and 2010 by the United States Department of Energy. As a platform chemical, 3HP can be converted into other valuable C3-based chemicals, such as acrylic acid, acrylamide, malonic acid, and other 3HP- or acryl-based polymers [1-4]. Several chemical approaches to produce 3HP industrially using oxidation from 1,3-propanediol or 3HPA, and hydration from acrylic acid have been reported [5, 6]. However, several advantages associated with free of petroleum-based raw materials, low cost, and environmental issues are driving the production of 3HP through biological methods [7].

For the biological production of 3HP, two biosynthetic pathways have been developed using glucose or glycerol as carbon sources. In the glucose pathway, 3HP can be produced by a carbon fixation pathway such as the 3HP/4HB cycle [8-11]. In the glycerol pathway, glycerol dehydratase converts glycerol to 3HPA, and then ALDH converts 3HPA to 3HP [12, 13]. In recent years, a large amount of glycerol has been produced as a byproduct in the biodiesel industry, therefore glycerol is attracting attention as a suitable carbon source for 3HP production [14, 15].

Eight 3HP-producing ALDH enzymes have been reported to date, including *Bd*DhaS, *Cn*GapD⁴, *Ec*AldH, *Kp*PuuC, *Kp*Yd c W, *Kp*YneI, *Sc*Ald⁴, and *Ab*KGSADH (Table S1) [16-22]. Among these, only one crystal structure of *Ab*KGSADH has been reported, and this report explained how *Ab*KGSADH stabilize to catalyze various chemicals such as α-ketoglutaric semialdehyde, succinic semialdehyde, and 3-HPA [23]. The structural information was also utilized to enhance its reactivity [24]. However, enzyme reactivity of ALDH for the conversion of 3HPA to 3HP is still low, and the activity imbalance in between DhaB and ALDH causes toxic 3HPA to accumulate in the cell. Moreover, the high expression of ALDH enzyme can be a burden on the host strain, therefore ALDH increasing activity by enhancing 3HPA specificity has been required [25].

Here, we elucidate the unique substrate specificity of ALDH enzymes by bioinformatic analysis. Crystal structure of *Ab*KGSADH, structure homology modeling, and molecular docking simulation revealed unnatural 3HPA binding modes of eight 3HP-producing ALDHs, and several variants with enhanced activity were obtained successfully. The results from this study could be utilized for the improved production of 3HP.
Materials and Method

Protein Homology Modeling

Protein homology modeling was performed using SWISS-MODEL [26, 27]. Model structures of seven 3HP-producing ALDHs were built. NCBI accession codes of target sequences are: AFQ57871.1 (BdDhaS), WP_011617618.1 (CnGapD4), OSL13388.1 (EcAldH), WP_020323708.1 (KpPuuC), AB727355.1 (KpYdcW), WP_015958390.1 (KpYneI), and NP_015019.1 (ScAld4). PDB codes of template structures, amino acid sequence identity between query and template sequences, and QMEAN scores are shown in Table S2.

Molecular Docking Simulation

Molecular docking simulations of 3HPA to 3HP-producing ALDH structures were performed by AutoDock Vina software [28]. Crystal structures of AbKGSADH (PDB code 5X5U) [23] and seven model structures of 3HP-producing ALDHs, including BsDhaS, CnGapD4, EcAldH, KpPuuC, KpYdcW, KpYneI, and ScAld4, were superimposed and then docking simulations were performed. The 3HPA ligand structure was prepared using the JLigand software [29]. The pbdqt files were generated using AutoDock Tools, and all steps were performed by the AutoDock Vina manual [30]. The grid size for AbKGSADH was x = 44, y = 40, z = 34, and the grid center was designated at x = -31.496, y = -11.696, z = -2.0. The grid size for BsDhaS was x = 44, y = 40, z = 32, and the grid center was designated at x = -31.725, y = -11.576, z = -2.897. The grid size for CnGapD4 was x = 44, y = 40, z = 32, and the grid center was designated at x = -31.227, y = -11.339, z = -2.604. The grid size for EcAldH was x = 42, y = 40, z = 32, and the grid center was designated at x = -31.382, y = -11.676, z = -3.182. The grid size for KpPuuC was x = 44, y = 40, z = 40, and the grid center was designated at x = -31.571, y = -11.640, z = -2.65. The grid size for KpYdcW was x = 44, y = 40, z = 32, and the grid center was designated at x = -30.644, y = -10.669, z = -3.018. The grid size for KpYneI was x = 42, y = 40, z = 42, and the grid center was designated at x = -30.794, y = -11.876, z = -2.618. The grid size for ScAld4 was x = 44, y = 40, z = 32, and the grid center was designated at x = -31.586, y = -11.833, z = -2.735. 90 docking poses in each ALDH were generated, and one pose with the most appropriate direction and distance in between aldehyde group of 3HPA and catalytic residue was selected.

Site-Directed Mutagenesis and Enzyme Preparation

Site-directed mutations were performed using a QuikChange kit (Agilent, USA), and mutated nucleotide sequences were confirmed. AbKGSADH enzyme was prepared with same method as described in our previous study [23]. AbKGSADH wild type and variants sub-cloned into pProEX-HTa vector (Thermo Fischer Scientific, USA) were transformed into an E. coli BL21(DE3)-T1R strain, and each strain was cultured to an OD600 of 0.6 in fresh LB medium with 100 mg 1\(^{-1}\) ampicillin at 37°C. AbKGSADH protein expression was induced by 0.5 mM IPTG. After 20 h at 18°C, the cells were harvested by 4,000 g for 15 min at 4°C. The cells were resuspended in ice-cold 40 mM Tris-HCl, pH 8.0, and disrupted by ultrasonication. Cell debris was removed by centrifugation at 13,000 g for 30 min, and the lysate was applied onto a Ni-NTA agarose column (Qiagen, Germany). After washing with 40 mM Tris and 25 mM imidazole, pH 8.0, the AbKGSADH protein was eluted with 40 mM Tris and 300 mM imidazole, pH 8.0.

Enzyme Activity Assay

The activity of AbKGSADH wild type and variants was determined by measuring the increase of absorbance at 340 nm. Enzyme reaction was performed with a reaction mixture of 0.5 mM total volume at 25°C. For kinetic analysis, reaction mixtures contained 100 mM Tris-HCl, pH 8.0, 500 μM NAD, 50 nM AbKGSADH enzyme, and various concentrations of 3HPA (0.5-10 mM). The reactions were initiated by the addition of enzyme. All reactions were performed in duplicate. The initial velocity of each measurement was calculated with extinction coefficient of NADH (6.22 M\(^{-1}\) cm\(^{-1}\)). Statistical analysis for k\(_{\text{cat}}\) and k\(_{\text{m}}\) values was performed by Michaelis Menten models using OriginPro software (OriginLab, USA).

Results and Discussion

Unnatural Substrate Availability of ALDHs

According to the enzyme nomenclature databases operated by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology, ALDHs belonging to the EC 1.2.1.- group (oxidoreductases acting on the aldehyde- or oxo-group of donors with NAD or NADP as acceptor) are divided into 107 classes (1.2.1.1-107) based on their specific activity [31]. Among these 107 classes, 2 classes have been deleted, 12 classes have been transferred to another entry, and 13 classes have no annotated nucleotide/amino acid sequence data in the database. Another 34 classes have amino acid sequences that refer to their entries, but no crystal structure has been reported to verify their structural fold and elucidate detailed molecular mechanisms and catalyses. Finally, the selected crystal structures of the remaining 46 different ALDH classes were compared (Tables S3, S4). Interestingly, when the selected crystal structures were superimposed on each other, approximately 70% of the ALDH classes displayed a similar overall structure (ALDH fold) and the remaining ALDHs have the GapDH fold (15%) or their own unique shape (15%) (Tables S3, S4). Moreover, many enzymes with the ALDH fold are multi-classified in the EC 1.2.1.- category. These observations indicate that most ALDH enzymes have similar structural conformations, and differences in only several amino acids involved in the constitution of the substrate binding pocket cause ALDH enzymes to accommodate different substrates. Due to these structural properties, ALDH enzymes have relatively broad substrate specificities, which in turn allow some ALDH enzymes to utilize unnatural aldehyde chemicals as substrates.
Unnatural 3HPA Binding Mode Prediction

The oxidation reaction from 3HPA to 3HP is a key step in the efficient production of 3HP from glycerol. However, because enzymes utilizing the unnatural 3HPA as a substrate have not yet been reported in natural organisms, use of the unique enzyme property of the ALDH enzymes with a broad substrate specificity is crucial for the enzymatic conversion of 3HPA to 3HP. Although eight 3HP-producing ALDHs have been reported to date, their low activities against 3HPA still remain a bottleneck for highly efficient 3HP production [16-22]. Here, an understanding of the substrate binding sites of the eight ALDHs and detailed structural comparisons of these ALDHs can be considered essential for the development of the ALDH enzymes with high activity against 3HPA. Because the crystal structure was reported only in AbKGSADH among the eight 3HP-producing ALDHs [23], we first attempted to obtain the three-dimensional structure of the seven other ALDHs through homology modeling. We used the known structure with the highest amino acid homology to each ALDH as a template model for more accurate structural prediction, and all seven modeled structures had reasonable model quality estimation scores (Table S2).

We also performed the molecular docking simulations of the 3HPA molecule into the eight 3HP-producing ALDHs to identify the 3HPA binding mode of these enzymes (Fig. 1). Although the 3HPA molecule was positioned in a slightly different mode in each enzyme, the aldehyde group where the catalysis occurs was directed...
toward the catalytic residues in all eight enzymes. When we compared the structures of the eight 3HP-producing ALDHs, the two catalytic residues, Glu253 and Cys287 in AbKGSADH, were completely conserved in all eight enzymes, indicating that these enzymes catalyze the reaction in an identical manner (Fig. 1). We then selected ten residues located in the vicinity of the 3HPA molecule and compared these residues in eight 3HP-producing ALDHs. Among eight enzymes, the EcAldH and KpPuuC enzymes, with 83% sequence identity, contained all ten residues, and the other six enzymes contained variable residues (Fig. 2). Of the ten selected residues, three residues, Phe156, Val286, and Phe450 in AbKGSADH, were highly conserved in the 3HP-producing ALDHs, with the exception of KpYnel, which contained Tyr156, Asp285, and His450 residues, respectively. Seven other residues, including Ser109, Asn159, Gln160, Arg163, Arg281, Ile288, and Pro444 in AbKGSADH, were variable in the 3HP-producing ALDHs (Fig. 2). In particular, three residues, Ser109, Ile288, and Pro444 in AbKGSADH, were highly variable, so that a dominant residue was not found (Fig. 2).

### Protein Engineering of AbKGSADH to Improve 3HPA Utilization

As we described above, the residues involved in stabilization of the 3HPA substrate are quite diverse among the known 3HP-producing ALDHs, and these variable residues can be potential targets for enzyme engineering for more efficient 3HP production. Of ten residues involved in substrate binding, we selected four relatively more variable residues, Ser109, Arg281, Ile288, and Pro444 in AbKGSADH, for target engineering sites. We then replaced these four residues of AbKGSADH with the various corresponding residues located in seven other 3PH-producing ALDHs, and generated the following ten AbKGSADH variants: AbKGSADHS109G, AbKGSADHS109L, AbKGSADHR281Y, AbKGSADHR281F, AbKGSADHR281Q, AbKGSADHI288T, AbKGSADHP444F, AbKGSADHP444G, AbKGSADHP444L, and AbKGSADHP444S.

We measured the specific activities and kinetic parameters of these ten variants and compared them with the same parameters for AbKGSADHWT. The AbKGSADHWT and AbKGSADH3109G variants showed somewhat decreased activities compared with AbKGSADHWT (Fig. 3A). The kinetic parameters of these two variants indicate that the decreased activities are due to decreased \( k_{cat} \) values rather than decreased substrate affinity (Table 1). The phenomenon was more dramatic in the AbKGSADH3109G variant, and we suspect that replacement...

---

**Fig. 2. Sequence alignment of core residues contributing 3HPA binding pocket formation.** Twelve amino acid residues contributing unnatural 3HPA substrate binding pocket are aligned. Graphical position weight matrix at the top of the alignment table shows the dominance of each site amino acid.

**Fig. 3. Kinetic parameters and relative catalytic efficiency.** (A) Relative catalytic efficiency represented by \( k_{cat}/K_m \) value. The catalytic efficiency of ten variants were compared with that of wild type AbKGSADH. (B) Enzyme kinetic graphs of AbKGSADHWT (left) and AbKGSADH3109G (right).
of Ser109 by a small hydrophobic group increased the binding affinity for 3HPA, however, the altered binding conformation to the molecule by the replacement severely hampered the conversion ability of the variant. The AbKGSADH^{R281Y} variant showed only 28% activity compared to the wild type, and kinetic analysis of the variant revealed that decreased activity was due mainly to decreased $k_{cat}$ value (Fig. 3A and Table 1). In the case of the variants mutated at the Pro444 site, all four variants showed decreased activities compared with the wild type (Fig. 3A). Kinetic analysis of the four variants indicated that the levels of decreased activities were almost proportional to the levels of decreased $k_{cat}$ values, with the $K_M$ values similar to each other (Table 1). Based on these observations, we conclude that three residues, Ser109, Ile288, and Pro444 of AbKGSADH, provide a more suitable conformation for 3HP binding and enzyme catalysis compared with the residues possessed by other 3HP-producing enzymes.

Interestingly, the three variants mutated at the Arg281 position of AbKGSADH exhibited substantially increased catalytic efficiencies. The AbKGSADH^{R281F} variant displayed a $k_{cat}$ value almost identical to AbKGSADH WT, but displayed a dramatically decreased $K_M$ value, resulting in 23% enhanced catalytic efficiency for this variant. The AbKGSADH^{R281Q} variant displayed higher values in both $k_{cat}$ and $K_M$, resulting in a 16% improvement in catalytic efficiency for this variant. The most remarkable improvement was observed in the AbKGSADH^{R281Y} variant. This variant showed 2.5-fold increased catalytic efficiency with approximately half the $K_M$ value and 24% higher $k_{cat}$ value compared with AbKGSADH WT, indicating that the enhanced catalytic efficiency was due to both the higher substrate affinity and increased conversion rate to product (Table 1 and Figs. 3A and 3B). Based on these results, we believe that the location of an amino acid with a bulky or ring side-chain, such as Arg, Gln, Phe, and Tyr, at the position of Arg281 of AbKGSADH is crucial for stabilization of the 3HPA substrate. In particular, the tyrosine residue appears to be most optimal for the unnatural 3HPA substrate, In particular, the tyrosine residue appears to be most optimal for the unnatural 3HPA substrate, in fact, half of the 3HP-producing ALDHs have the tyrosine residue at this site.

In summary, bioinformatic analysis of ADLHs belonging to the EC 1.2.1.- category shows that most ALDHs have the same overall fold, and their substrate specificity depends on amino acid residues involved in the formation of the substrate binding pocket. ALDH enzymes with a conventional ALDH fold can utilize a wide range of aldehyde chemicals as substrates, indicating that unnatural aldehydes can also be catalyzed by these ALDHs. Structure homology modeling and molecular docking simulation allowed us to identify detailed substrate binding modes of the eight 3HP-producing ALDHs reported to date and determine amino acids that can be protein engineering targets to improve utilization of the unnatural 3HPA substrate. Of the ten AbKGSADH variants, three variants in which the Arg281 site of KGSADH was mutated showed enhanced 3HPA utilization ability. In particular, AbKGSADH^{R281Q} exhibited improved catalytic efficiency by 2.5-fold. This study offers valuable information in the pursuit of highly efficient 3HP production.

Acknowledgments
This work was supported the National Research Foundation of Korea (NRF) grant funded by the Korea government (MIST) (No. 2020R1A4A1018332 and 2021R1C1C2004411).

Conflict of interest
The authors have no financial conflicts of interest to declare.

References
1. Borodina I, Kildegaard KR, Jensen NB, Blicher TH, Maury J, Shersky S, et al. 2015. Establishing a synthetic pathway for high-level production of 3-hydroxypropionic acid in Saccharomyces cerevisiae via beta-alanine. Metab. Eng. 27:57-64.
2. Karp EM, Eaton TR, Nogue VSI, Vorotnikov V, Biddy MJ, Tan ECD, et al. 2017. Renewable acrylonitrile production. Science 358:1307-1310.
3. Kumar V, Ashok S, Park S. 2013. Recent advances in biological production of 3-hydroxypropionic acid. Biotechnol. Adv. 31:945-961.
4. Valdehuesa KNG, Liu HW, Nisola GM, Chung WJ, Lee SH, Park SJ. 2013. Recent advances in the metabolic engineering of microorganisms for the production of 3-hydroxypropionic acid as C3 platform chemical. Appl. Microbiol. Biotechnol. 97:3309-3321.
5. Haas T, Brossmer C, Meier M, Arntz D, Freund A. 2000. Process for preparing 3-hydroxypropionic acid or its salt. Patent Application No. EP0819670.

Table 1. Kinetic parameters of AbKGSADH wild type and ten variants.

| AbKGSADH     | $k_{cat}$ [min⁻¹] | $K_M$ [mM] | $k_{cat}/K_M$ [(mM min)⁻¹] |
|--------------|-------------------|------------|----------------------------|
| Wild type    | 1291.4 ± 119.2    | 4.581 ± 0.921 | 281.9                      |
| S109G        | 359.8 ± 52.6      | 1.646 ± 0.726 | 218.5                      |
| S109L        | 1021.2 ± 149.8    | 4.156 ± 1.377 | 245.9                      |
| R281Y        | 1595.0 ± 93.6     | 2.338 ± 0.372 | 662.3                      |
| R281E        | 1317.0 ± 108.2    | 3.808 ± 0.728 | 345.9                      |
| R281Q        | 1891.6 ± 236.6    | 5.764 ± 1.462 | 328.2                      |
| I288T        | 367.0 ± 101.2     | 4.593 ± 2.814 | 79.9                       |
| P444F        | 1171.0 ± 62.6     | 4.714 ± 0.547 | 248.4                      |
| P444G        | 395.2 ± 95.0      | 3.512 ± 2.070 | 112.5                      |
| P444L        | 1130.4 ± 141.4    | 4.204 ± 1.184 | 268.9                      |
| P444S        | 1039.8 ± 305.8    | 4.690 ± 2.996 | 221.7                      |
6. Behr A, Botulinski A, Carduck FJ SM. 1996. process for preparing 3-hydroxypropionic acid. Patent Application No. EP0579617.
7. Jiang X, Meng X, Xian M. 2009. Biosynthetic pathways for 3-hydroxypropionic acid production. Appl. Microbiol. Biotechnol. 82: 995-1003.
8. Rathnasighe C, Raj SM, Lee Y, Catherine C, Ashoka S, Park S. 2012. Production of 3-hydroxypropionic acid via malonyl-CoA pathway using recombinant Escherichia coli strains. J. Biotechnol. 157: 633-640.
9. Suyama A, Hisuchi Y, Urushihara M, Maeda Y, Takagawa K. 2017. Production of 3-hydroxypropionic acid via the malonyl-CoA pathway using recombinant fission yeast strains. J. Biosci. Bioeng. 124: 392-399.
10. Hugler M, Huber H, Stetter KO, Fuchs G. 2003. Autotrophic CO₂ fixation pathways in archaea (Crenarchaeota). Arch. Microbiol. 179: 160-173.
11. Hugler M, Menendez C, Schagger H, Fuchs G. 2002. Malonyl-coenzyme A reductase from Chloroflexus aurantiacus, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO₂ fixation. J. Bacteriol. 184: 2404-2410.
12. Forage RG, Foster MA. 1982. Glycerol fermentation in Bacillus subtilis. Appl. Environ. Microbiol. 43: 160-167.
13. Lebedev AA, Young P, Isupov MN, Moroz OV, Vagin AA, Murshudov GN. 2012. JLigand: a graphical tool for the CCP4 template-refinement library. Acta Crystallogr. D, Biol. Crystallogr. 68: 431-440.
14. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. 2009. AutoDock4 and autodockTools4: Automated docking with selective receptor flexibility. J. Comput. Chem. 30: 2785-2791.
15. Moss GP. Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB), IUBMB, Accessed 17 Aug. 2021. <https://www.qmul.ac.uk/sbcs/iubmb/>.