Three-Dimensional Modeling of Plant 4-Hydroxyphenylpyruvate Dioxygenase, a Molar Target of Triketone-Type Herbicides

Hitoshi KAKIDANI* and Kenji HIRAI
Sagami Chemical Research Center, 2743-1 Hayakawa, Ayase, Kanagawa 252-1193, Japan
Toho Corporation, 2743-1 Hayakawa, Ayase, Kanagawa 252-1123, Japan

(Received May 23, 2003; Accepted July 22, 2003)

Homology modeling of plant 4-hydroxyphenylpyruvate dioxygenase (4-HPPD) was carried out using structural information on the Pseudomonas fluorescens enzyme as a template. A three-dimensional structural model of the barley enzyme was investigated in detail to compare its catalytic pocket with that of the Pseudomonas enzyme. Striking conservation of the catalytic pocket provided a rational basis for designing potent 4-HPPD inhibitors by the in silico docking of compounds and model/template enzymes.

Key words: 4-HPPD, enzyme inhibitor, homology modeling, SWISS-MODEL.

INTRODUCTION

4-Hydroxyphenylpyruvate dioxygenase (4-HPPD, EC 1.13.11.27) catalyzes the oxidation of 4-hydroxyphenylpyruvate, giving homogentisate (HGA) and CO₂, through a complex mechanism involving decarboxylation, dioxygenation and rearrangement.10 This enzyme is a molecular target of a group of bleaching herbicides that includes pyrazolate, benzobicyclon and isoaflutole. The mode of action of these compounds is postulated to be competitive inhibition of 4-HPPD, either as an active compound or as a prodrug. In most organisms including higher plants, 4-HPPD plays a major role in tyrosine metabolism. The role of the reaction product HGA is, however, divergent. In higher plants and cyanobacteria, HGA is a common precursor of plastoquinone and tocopherols, both indispensable for photosynthesis and normal physiology. In some gram-negative bacteria such as Shewanella,1 Legionella2 and Vibrio,3 HGA is converted to different types of compounds such as melanin, fluorescent substances and hemolysin. In mammals, HGA is further metabolized in the TCA cycle. In patients with a rare hereditary disease, tyrosinemia type I (deficient in fumarylacetoacetase activity), 4-HPPD activity leads to an accumulation of toxic compounds.7 4-HPPD inhibitors, therefore, are being studied as therapeutic agents as well as herbicides.

From a structural viewpoint only one enzyme from Pseudomonas fluorescens has been studied in detail by X-ray crystallography.9 Overall protein folding was quite characteristic (derived from four βαββαβ modules), and Fe(II) in the active site was found to be coordinated by two histidine residues (His161 and His240) and one glutamate residue (Glu322). These structural features are also observed in 2,3-dihydroxybiphenyl 1,2-dioxygenase (DHBD; EC 1.13.11.39)10 and catechol 2,3-dioxygenase (C 23 O; EC 1.13.11.2).10 Based on these findings, the complex of NTBC·(2-(2-nitro-4-trifluoromethylbenzoyl)cyclohexane-1,3-dione) and Pseudomonas 4-HPPD was modeled recently.10 In this article, we describe homology modeling of plant 4-HPPD in an effort to better understand how the enzyme-inhibitor complex forms. This approach should facilitate the design of potent 4-HPPD inhibitors.

METHODS

1. Multiple Alignment and Associated Phylogenetic Representation of Amino Acid Sequences

Amino acid sequences of 4-HPPD were taken from the database at the NCBI web site (http://www.ncbi.nlm.nih.gov/). CLUSTAL W11 presented on the DNA Data Bank of Japan web site (http://www.ddbj.nig.ac.jp/) was employed for the multiple alignment. A phylogenetic tree based on the alignment was drawn using TreeView 1.6.6 (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html).

2. Homology Modeling

Homology modeling of plant 4-HPPD was performed with the ExPASy web server presented at the SWISS-
MODEL\textsuperscript{18, 19} web site (http://www.expasy.org/swissmod/SWISS-MODEL.html). Swiss-PdbViewer (spdbv) 3.7b2 was downloaded from the same site. Following the on-line manual provided, manipulation of the viewer and alignments were carried out. Structural data on \textit{Pseudomonas} 4-HPPD presented at The Protein Data Bank web site (http://www.rcsb.org/pdb/) was referred to throughout. The actual template used for the modeling was a \textit{Pseudomonas} 4-HPPD A subunit structure file (1 CJXA.pdb), pre-tailored from the entire structure file (1 CJX.pdb) with SWISS-MODEL. Plant 4-HPPD sequences, the barley sequence in particular, were aligned to the template sequence on spdbv, taking into consideration the multiple alignment output and secondary structure of the template. Among several alternate alignments sent to the ExPASy web server, the most favorable, which gave the lowest energy and longest coverage, was adopted. The output structural model was then energy minimized using the GROMOS96 force field calculation on spdbv. Finally Fe(II) was inserted into the model at the position identical to that in the template structure by maximally fitting these two structures.

3. Insertion of 4-HPPD Inhibitors into the Model

The three-dimensional structure of NTBC was drawn with HyperChem 5 (Hypercube, Inc., Gainesville, Florida). The Monte Carlo calculation in the software was used to minimize energy (total energy = 50.2 kJ/mol). This structure was inserted into the template structure and the model structure. In the former case, a complex structure similar to that of Wu \textit{et al.}\textsuperscript{10} was obtained. For drawing the three-dimensional structure of pyrazolate, benzobicyclob and isoauxifolute metabolites, the geometry optimization tool in HyperChem 5 was applied instead of the Monte Carlo calculation.

RESULTS

1. Homology Modeling of Barley 4-HPPD

In the initial attempt at the homology modeling of plant 4-HPPD, a very limited region was automatically modeled by SWISS-MODEL ("First Approach mode") due to low overall sequence identity with the template. It was then necessary to manually align the target and template sequences on spdbv ("Optimise mode"). To this end, we combined the multiple sequence alignment and investigations of the structural and functional features of the template enzyme.

Eighteen amino acid sequences of 4-HPPD were picked from the NCBI database. These included 4 plant enzymes (Barley, Carrot, Arabidopsis and Coleus), 4 mammalian enzymes (Human, Rat, Mouse and Pig) and 4 bacterial enzymes (\textit{Pseudomonas}, Legionella, Vibrio and Shewanella). The multiple alignment (data not shown) and associated phylogenetic representation (Fig. 1) revealed a divergent relationship among these sequences. This phylogenetic tree was consistent with the modern taxonomical classification of all organisms (http://www.ncbi.nlm.nih.gov/Taxonomy/) and other trees based on specific\textsuperscript{15} or combined protein data.\textsuperscript{16, 17} For unknown reasons, the bacterial sequences were very divergent, considering that all of these bacteria belong to gammaproteobacteria. As plant (barley in particular) and \textit{Pseudomonas} enzymes were distantly related to each other (i.e. below the twilight zone\textsuperscript{16}), homology modeling was not

![Phylogenetic Tree](http://example.com/phylogenetic_tree.png)

\textit{Fig. 1.} Unrooted representation of the results of the phylogenetic analysis of 4-HPPD amino acid sequences. Horizontal bar indicates evolutionary distance measured by the number of amino acid replacements per site.
feasible. A closer look at the alignment, however, provided a clue to manually align the barley sequence to the *Pseudomonas* sequence. The C-terminal domain, which is presumably responsible for the catalytic function, was much more conserved than the N-terminal domain. Within the C-terminal domain, several sequence clusters with a high degree of conservation were observed. These clusters included His161, His240 and Glu322 (numbering in *Pseudomonas* enzyme) which were bound to Fe(II). These findings suggested a high degree of structural conservation of the catalytic pocket despite apparent sequence divergence.

The alignment adopted for the modeling (Fig. 2) covered most of the barley enzyme sequence (Arg36-Gln427, 392/434 amino acids). This alignment exhibited low amino acid identity (26%, 89/342 amino acids) and a high degree of insertions/deletions (15 sites, 60 amino acids in total). The barley model based on this alignment nevertheless exhibited profound stability (total energy = $-10,300 \text{ kJ/mol}$) as determined from the GROMOS96 force field calculation after the addition of hydrogen atoms and energy minimization.

Comparison of the outlook of model and template structures revealed gross similarity with some differences in the molecular surface area (Fig. 3). All of the 15 insertion/deletion sites were located in the surface area. Among them, an

---

**Fig. 2.** Sequence alignment adopted for the modeling of barley 4-HPPD. Three amino acid residues involved in coordination to Fe(II) are boxed.

---

**Fig. 3.** The template and model structures.

(a) *Pseudomonas fluorescens* 4-HPPD (1 CIX.pdb). Ethyl mercury is included in the N-terminal domain (left part, space-filled, carbon atoms in gray and mercury atom in magenta). Fe(II) and acetate are included in the C-terminal domain (right part, space-filled, Fe atom in gold, carbon atoms in gray and oxygen atoms in red). (b) Barley 4-HPPD model. Fe(II) is included in the C-terminal domain (space-filled, Fe atom in gold). Arrow indicates inserted 15-amino acid sequences.
insertion of 15 consecutive amino acids in the C-terminal region (Gly385-G399, Fig. 2) was noteworthy. As this insertion was common to plant enzymes and was located very close to Glu379, care was taken to evaluate the effect of this insertion. Fortunately, this region in the model was properly folded and was stabilized by many hydrogen bonds. This region apparently did not interfere with the catalytic pocket though it might affect the entry of substrate and/or inhibitors.

As SWISS-MODEL does not involve Fe(II) in the calculation, this atom was finally inserted into the model structure by superposing the model and the template structures. Due to the apparent high degree of structural conservation of the catalytic pocket, the positioning of Fe(II) was unequivocal (Fig. 4). It is most likely that the substrate and/or inhibitors are bound to the large pocket surrounded by eight $\beta$ sheets and one $\alpha$ helix. In the template structure, acetate and water molecules fill this space (water molecules, not shown in the figures).

2. Possible Interactions between Inhibitors and the Enzyme

Insertion of the NTBC molecule in the model structure was then carried out manually on the graphics interface (Fig. 5a). In this process, a collision between the enzyme and the inhibitor was avoided, and the geometry around Fe(II) was taken into consideration. As a result, reasonable positioning of the inhibitor was achieved while no energy optimization was employed. A schematic drawing of the coordination around Fe(II) for the enzyme-inhibitor complex was made (Fig. 5b), along with that of the presumptive enzyme-substrate complex (Fig. 5c). We postulate that the 6-membered ring formed by the inhibitor and Fe(II) is more stable than the 5-membered ring formed by the substrate and Fe(II). Furthermore, the enzyme-inhibitor complex would not undergo decarboxylation, which is the first step in the catalytic reaction. Taken together, triketone/1,3-diketone-motif is a reasonable common structure of potent inhibitors.

As most of the NTBC atoms in the model were located within 10 Å from Fe(II), amino acid residues within 10 Å from Fe(II) were examined. When these residues (34 in the barley model) were compared for 16 organisms (multiple alignment derived from Fig. 1, Legionella and Shewanella were omitted), striking conservation was observed (Fig. 6). This result indicates that the catalytic pocket of 4-HPPD is highly conserved while the overall amino acid sequence is rather divergent. This conservation may reflect the complex catalytic mechanism of 4-HPPD. Figure 6 also revealed group-specific amino acid residues (e.g. a leucine residue

---

![Fig. 4](image-url)

Fig. 4. Magnified C-terminal domain of the template and model structures.

(a) *Pseudomonas fluorescens* 4-HPPD (1 C2X.pdb), Fe(II) and acetate are included (ball and stick). His161, His240 and Glu322 are highlighted. (b) *Barley* 4-HPPD model. Fe(II) is included (ball and stick). Color representations are the same as in Fig. 3. His211, His295 and Glu379 are highlighted. Arrow indicates inserted 15-amino acid sequences.
Fig. 5. Model of barley 4-HPPD-NTBC complex. (a) All atoms (except for hydrogen) within 12 Å of the C(8) atom of NTBC\textsuperscript{150} are shown. (b) Schematic representation of Fe(II) coordination for (a). (c) Supposed coordination formed in the catalytic reaction.

![Diagram of the barley 4-HPPD-NTBC complex](image)

Fig. 6. Comparison of amino acids located within 10 Å from Fe(II). Arrows indicate conserved histidine and glutamate residues involved in coordination to Fe(II). Some group-specific amino acids are boxed.

![Comparison of amino acids](image)

common in plants). It is worth investigating whether these amino acid residues are involved in the specificity toward inhibitors.

Some potent bleaching herbicides whose target is speculated to be 4-HPPD are obviously not triketone-type compounds, but most likely converted to triketone-type compounds in the environment or by plant metabolism. Typical examples are pyrazolate (1), benzobicyclon (2) and isoxaflutole (3) (Fig. 7). Recent studies revealed that active compounds derived from 1, 2 and 3 are 1', 2' and 3', respectively. We calculated possible three-dimensional structures of 1', 2' and 3' to insert them in the
Herbicides (prodrugs)  

Active metabolites

Fig. 7. Three structurally related herbicides and their active metabolites. Actual herbicides, pyrazolate (1), benzobicyclon (2) and isoxaflutole (3), are shown on the left and corresponding active metabolites are on the right.

Fig. 8. Superposed compound structures and surrounding amino acid residues that are presumably important for the inhibitor-enzyme interaction.

barley 4-HPPD model structure as was made for NTBC. The superposition of these compounds along with NTBC (Fig. 5a) indicated several amino acid residues that are possibly important for the interaction with the inhibitors (Fig. 8). His211, His293 and Glu379 are obviously not in direct contact with inhibitors, and solely involved in the coordination around Fe(II). Asn267, Gln278 and Gln292 are possible hydrogen-bond acceptors and/or receptors. Gln278 and Gln292 (Gln225 and Gln229 in Psuedomonas, respectively) were speculated to be H-bonded with a hydroxyl group in the substrate ρ-hydroxyphenylpyruvate." It is, therefore, possible that they are involved in the
interaction with polar groups on the aromatic ring of the inhibitors. Likewise, Asn267 may be H-bonded with another polar group. Leu362, Phe404 and Phe409 may provide important hydrophobic interactions. Phe409 could be involved in edge-to-face stacking interactions\textsuperscript{328} with the aromatic ring of the inhibitors. Leu362 and Phe404 are located close to the second ring of the inhibitors, though not in all cases. Phe404 restricts the size of the inhibitor at one end (right in the figure), and the adjacent Gly403 allows a bulky group to be positioned. The other end (upper left in the figure) has more space to be explored. It is implied that in the upper left direction, a bulkier group is allowed for an inhibitor to bind to the enzyme. Lastly, it has to be noted that all of the amino acids shown in Fig. 8 are entirely conserved among species tested.

**DISCUSSION**

While structure-based drug design is considered a useful research tool in the pharmaceutical industry, it has been much less developed in the agrochemical industry. This difference is mostly due to a higher degree of discrepancy between in vitro (receptor-agent interaction in the laboratory) and in vivo (efficacy in the field) data for agrochemicals, herbicides in particular. Another reason is the capital-intensive nature of this approach. Finding a target molecule (receptor) and its structural determination require expensive instruments and long-term R&D expenditure in general. This situation, however, has been changing as genome projects and bioinformatics proceed. The accumulation of available protein structural and functional data is also an important factor. We in this article presented an easy-to-access procedure for designing potent enzyme inhibitors based on a known target receptor structure. Fortunately, 4-HPPD was found to be a good model for this approach as a bacterial enzyme structure was determined by X-ray crystallography. While the overall amino acid sequence was obviously divergent, particularly in plant enzymes compared with bacterial enzymes (Fig. 1), homology modeling was possible as catalytic pockets were highly conserved. Among 4 plant 4-HPPDs whose amino acid sequence is available, barley was arbitrarily chosen and extensively investigated. Modeling of other plant enzymes is in progress, among which the carrot model exhibited remarkable stability (data not shown). Modeling of the human enzyme was even easier as no long insertion in the C-terminal domain was involved.

In this article we emphasized the striking conservation of the catalytic pocket. This implies that herbicide design is possible using the *Pseudomonas* structure without considering plant models. We, however, assume that detailed analyses will reveal substantial differences, reflecting substrate/inhibitor specificity. Thr163, for example (boxed as bacteria-specific, Fig. 6) is replaced by valine in other groups (Val213 for barley). As this residue faces the catalytic pocket, it is considered a possible site affecting the specificity.

The quantitative analysis of enzyme-inhibitor complexes is now a demanding task. To this end, we are introducing an automatic docking software. This procedure along with the detailed characterization of the catalytic pocket surface should help in the design of potent 4-HPPD inhibitors in the near future.

**REFERENCES**