Acc. Chem. Res. **2006,** *39,* 550−557

Domain Relationships in Thiamine Diphosphate-Dependent Enzymes

RONALD G. DUGGLEBY*
School of Molecular and Microbial Sciences, University of Queensland, Brisbane, Queensland 4072, Australia

Received February 9, 2006

ABSTRACT
Three-dimensional structures have been determined for 13 different enzymes that use thiamine diphosphate (ThDP) as a cofactor. These enzymes fall into five families, where members within a family have similar structures. In different families, there are similarities between some domains that clearly point to a common ancestor for all of these enzymes. Where the enzyme structures differ, evolutionary relationships between families can be discerned. Here, I present an analysis of these families and propose an evolutionary pathway to explain the diversity of structures that are now known.

Introduction

An understanding of protein and enzyme evolution is useful for several reasons. First, the phylogenetic analysis of equivalent proteins from different organisms is an aid to taxonomy, allowing deductions about how the organisms synthesizing those proteins are related to one another. Second, related enzymes have related enzymology; therefore, if the mechanism of one enzyme is well-understood, we can often extend this understanding to similar enzymes. Moreover, defining the conserved structural features imposed by evolutionary constraints lead to hypotheses on the roles of individual residues in the functioning of the enzyme. This, in turn, might allow biotechnologists to modify proteins and develop novel, technologically useful enzymes. Third, because protein evolution depends upon the underlying mechanisms of molecular genetics, understanding how proteins change can reveal information about these genetic mechanisms. Finally, in the burgeoning field of genome science, the assignment of function to genes relies almost completely upon being able to recognize similarities and relationships between proteins.

Thiamine diphosphate (ThDP, Figure 1) is a cofactor for a variety of different enzymes[1] that are found in all forms of life. After the initial three-dimensional structure determination of transketolase[2] (TK), structures of pyruvate oxidase[3] (POX) and pyruvate decarboxylase[4] (PDC) soon followed. Since then, structures of most of the major types of ThDP-dependent enzymes have been determined. These studies, together with sequence analysis, have revealed that the enzymes can be classified into five broad families.

The largest of these has been termed[5] the POX family, which also includes PDC, indolepyruvate decarboxylase, benzoylformate decarboxylase (BFDC), oxalyl-CoA decarboxylase (OCDC), acetohydroxyacid synthase (AHAS), acetolactate synthase (ALS), benzaldehyde aldolase (BAL), and N^2-(2-carboxyethyl)arginine synthase. In addition to these nine, for which the three-dimensional structure has been determined, sequencing studies have shown that there are several other enzymes that fall into the same family. These include glyoxylate carboligase (GCL), CDP-4-aceto-3,6-dideoxygalactose synthase (the product of the *YerE* gene), sulfoacetaldehyde acetyltransferase, phenylpyruvate decarboxylase, 2-ketoisovalerate decarboxylase, 2-hydroxyphytanoyl-CoA lyase, and (1*R*,6*R*)-2-succinyl-6-hydroxy-2,4-cyclohexadiene-carboxylate synthase. POX is atypical of this family, by virtue of being the only member catalyzing a redox reaction. For most enzymes in this family (including POX), the reaction catalyzed (or at least the first stage) is the decarboxylation of a 2-ketoacid. Therefore, in the ensuing analysis, this family will be described as the "DC family".

The reactions catalyzed by most members of the DC family are summarized in Figure 1. The 2-ketoacid decarboxylases act on their substrate via route *1*, releasing carbon dioxide to form the resonance-stabilized enamine/α-carbanion ThDP derivative. This then reacts with a proton (route *a*) to release the aldehyde product. POX, AHAS, ALS, and GCL also follow route *1*, but the fate of the intermediate differs. In POX, there is a reaction with phosphate and oxygen (route *b*) forming acetyl phosphate and hydrogen peroxide, while a reaction with a 2-ketoacid (route *c*) occurs for the remaining three enzymes. Finally, BAL reaches the intermediate via route *2* and exits by route *a*.

The TK family is smaller, comprising phosphoketolase, dihydroxyacetone synthase, and 1-deoxy-D-xylulose-5-phosphate synthase, but only the structure of TK itself has been determined (although from four different species). The reaction catalyzed by TK is defined in Figure 1 by route *2f*. Enzymes falling into the same family as pyruvate:ferredoxin oxidoreductase (PFOR; route *1d* in Figure 1) are less easy to define. Only one structure (from *Desulfovibrio africanus*) has been determined,[6] but there are several variants of this enzyme with different numbers of subunits and domains. These variations are, at least in part, related to the identity of the final electron acceptor (ferredoxin, flavodoxin, or $NADP^+$). For example, the pyruvate:$NADP^+$ oxidoreductase of *Euglena gracilis*[7] has been shown to be an enlarged variant of PFOR (1803 residues) compared to the 1232 residues for *D. africanus* PFOR that uses ferredoxin. It is, to some extent, a semantic

Ron Duggleby was born in England (1945) and obtained his Ph.D. in Canada (Queen's University, 1972) before moving to Australia in 1975. He is now Reader in Biochemistry and Molecular Biology at the University of Queensland. In his first employment (1966), he was assigned to a project studying an enzyme. Showing a lack of imagination and originality that beggars belief, 40 years later, he is still studying enzymes. His current interests are focused on enzyme structure and mechanism, particularly those that use thiamine diphosphate as a cofactor.

* To whom correspondence should be addressed: School of Molecular and Microbial Sciences, University of Queensland, Brisbane, Queensland 4072, Australia. Telephone: +617-3365-4615. Fax: +617-3365-4699. E-mail: ronald.duggleby@uq.edu.au.

10.1021/ar068022z CCC: $33.50
Published on Web 05/28/2006

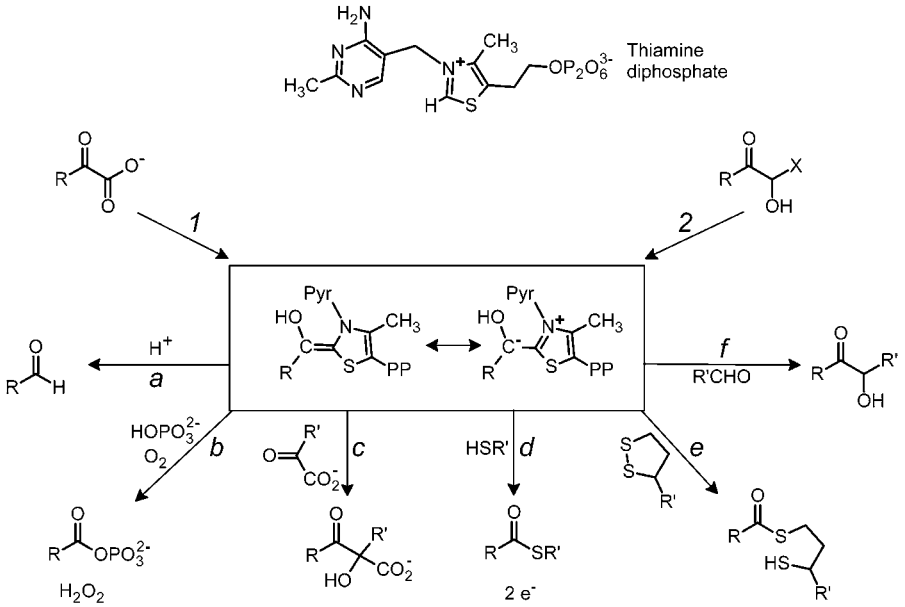

FIGURE 1. ThDP and reactions catalyzed by ThDP-dependent enzymes. The upper part of the figure shows the structure of ThDP, while the lower part summarizes most of the reactions catalyzed by ThDP-dependent enzymes. In these reactions, a 2-ketoacid (upper left) or an acyloin (upper right) substrate is cleaved adjacent to the carbonyl group with the transfer of the R−C=O moiety to the cofactor and release of the remainder of the substrate. The box encloses the resonating enamine/α-carbanion intermediate; "Pyr" represents 2,5-dimethyl-4-amino-pyrimidine and "PP" is the ethyl diphosphate tail. The intermediate reacts by routes *a*−*f*, completing the catalytic cycle. This representation of the reactions is adapted from Pohl et al.[1]

question whether such variations define different enzymes in a broader family or if they should all be considered as variants of a single PFOR. The only other enzymes that clearly belong in this family without being simply variant PFORs are 2-ketoglutarate:ferredoxin oxidoreductase and 2-ketoisovalerate:ferredoxin oxidoreductase. Because the 2-ketoacid substrate and the electron acceptor vary across this group of enzymes, it will be designated as the OR family.

Finally, the 2-ketoacid dehydrogenases (route *1e* in Figure 1) are multienzyme complexes containing three major catalytic components termed E1, E2, and E3, with ThDP being bound to the E1 component. These have been divided traditionally into two types depending upon whether E1 is made up of a single chain or has a two-chain (α plus *β*) structure. Here, these will be identified as the K1 and K2 families, respectively. The former is exemplified by *Escherichia coli* pyruvate dehydrogenase E1 (PDH), while the *Homo sapiens* branched-chain 2-ketoacid dehydrogenase E1 (BCDH) is an example of the latter. The K2 family includes the E1 component of 2-ketoglutarate dehydrogenase and the E1 of PDH from eukaryotes and Gram-positive bacteria. Acetoin dehydrogenase, although not acting on a 2-ketoacid, also belongs in the K2 family. On the other hand, PDH from Gram-negative bacteria and actinomycetes appears to stand alone in the K1 family.

A comparison between POX, PDC, and TK[8] revealed that they have common features, with each composed of three domains. One of these domains ("Pyr") interacts with the pyrimidine ring of the cofactor, while a second ("PP") binds the diphosphate tail. The Pyr domain contains a conserved glutamate residue[9] that is essential for activation of the cofactor, while the PP domain contains the conserved $GDGX_{25-30}NN$ motif,[10] which anchors ThDP to these enzymes via the diphosphate tail and a coordinating divalent metal ion. TK is a circular permutation of POX and PDC,[8] with the Pyr domain located in the N-terminal domains of POX and PDC but in the middle domain of TK. Similarly, the PP domain is C-terminal in POX and PDC but N-terminal in TK. It was also pointed out[8] that the Pyr and PP domains have similar topology to one another, suggesting that there may be a deeper evolutionary relationship underlying the structures of ThDP-dependent enzymes. Other workers have described the structural similarities between pairs of enzymes; for example, BCDH versus TK,[11] PFOR versus TK,[6] POX versus AHAS,[12] and PDH versus BCDH.[13] However, structural relationships between domains across this entire group of proteins have not been explored systematically. Here, I examine these similarities within and between ThDP-dependent enzymes. Some of the conclusions have been noted previously by other authors, but here, these are brought together and, combined with some new observations, suggest a possible scheme for the origins of this diverse group of proteins.

Materials and Methods

For this analysis, I selected one member of each family. In each case, I chose the structure that had been determined to the highest resolution [as available from the Protein Data Bank (PDB) in January 2006]. These are listed in Table 1. Where the asymmetric unit contains several copies of the protein, the one listed first in the PDB file was used. Structural relatives of various domains were

Table 1. Representative Members of the Families of ThDP-Dependent Enzymes

family	example	species	PDB ID	resolution (Å)	reference
DC	BFDC	*P. putida*[a]	1Q6Z	1.00	Polovnikova et al.[14]
TK	TK	*E. coli*	1QGD	1.90	Isupov et al.[b]
OR	PFOR	*D. africanus*	1KEK	1.90	Chabrière et al.[15]
K1	PDH	*E. coli*	1L8A	1.85	Arjunan et al.[13]
K2	BCDH	*H. sapiens*	1X7Y	1.57	Wynn et al.[16]

[a] *P. putida* = *Pseudomonas putida*. [b] Unpublished results.

identified using the Protein Structure Comparison Service SSM[17] at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm). Structural overlays were made with the Swiss-PdbViewer[18] obtained from the Expert Protein Analysis System Proteomics Server (http://www.expasy.org/spdbv). In some cases, it was necessary to obtain a preliminary overlay using SSM. Figures were prepared using ChemSketch (http://www.acdlabs.com), TopDraw,[19] and PyMol (http://www.pymol.org).

Results and Discussion

DC Domains. Proteins of the DC family consist of three domains of similar size. Each domain contains a central, five- or six-stranded, parallel β sheet linked by α helices and connecting loops. The connectivity of the N-terminal (Pyr) and C-terminal (PP) domains is identical but distinctly different from that of the middle domain (Figure 2). The similarity between the Pyr and PP domains is suggestive of a common evolutionary history, but the limited number of possible topologies allows more mundane explanations. Therefore, I compared the three-dimensional structures of the two domains. As shown in Figure 3A, the two domains overlay quite well with an overall root-mean-square deviation (rmsd) of 1.58 Å for 95 Cα atoms. The similarity is more clearly evident when the major elements of secondary structure are examined. The central β sheets of the two domains are almost identical (Figure 3B). The α-helical core also displays a good correspondence (Figure 3C) apart from one small helix (Pyr, 25−33; PP, 379−387) that shows a lateral shift of about 8 Å. This structural similarity strongly implies a common origin for the two domains, despite the fact that a sequence comparison (not shown) displays no convincing trace of their relatedness.

Omitted from Figure 3C are two helices that differ markedly: helix 100−105 from the Pyr domain and the corresponding helix 458−469 from the PP domain (see parts A and C of Figure 2). Both of these helices are at the interface between adjacent monomers in the biologically active tetramer of BFDC (Figure 3D). ThDP is also bound at the subunit interface, interacting with the Pyr domain of one monomer and the PP domain of its neighbor. Thus, assembly to a dimer (or a multiple thereof) is required

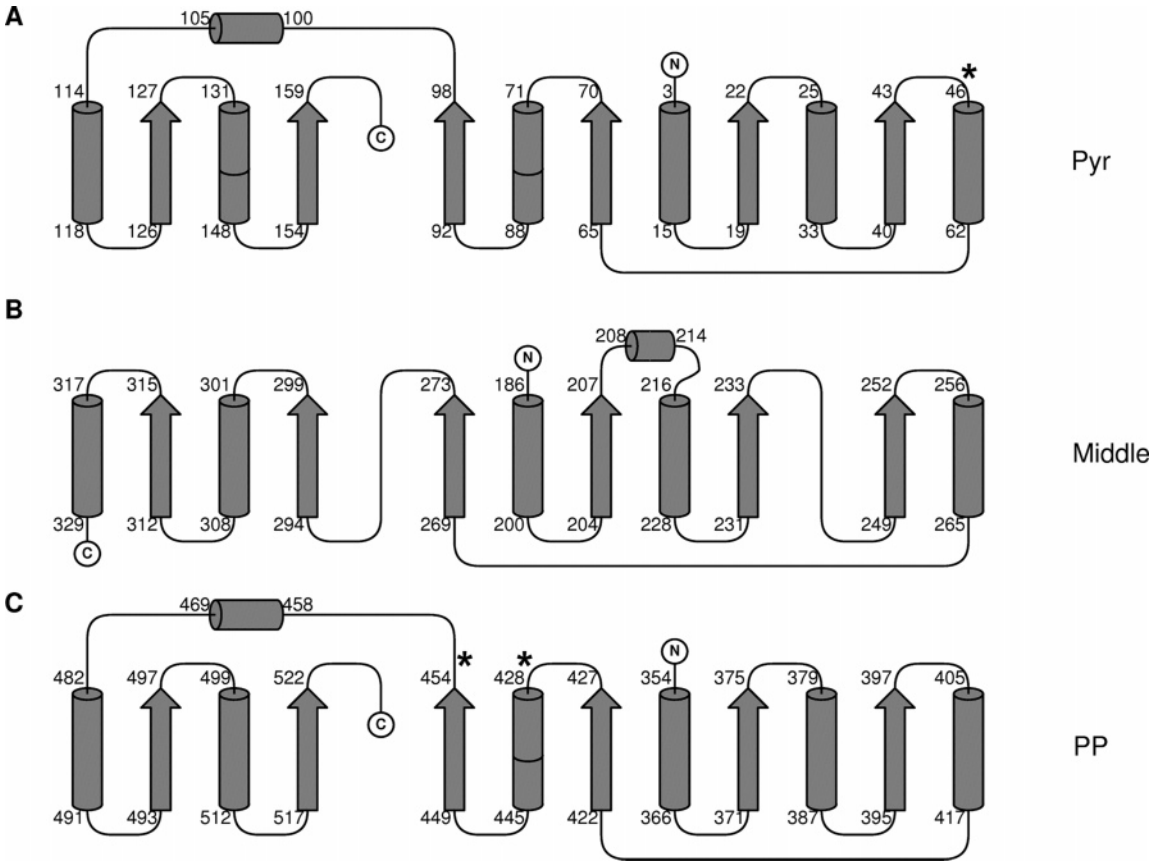

FIGURE 2. Topology cartoons for the three domains of BFDC. "N" and "C" represent the termini of each domain, while cylinders and arrows represent α helices and β strands, respectively. In the Pyr domain (A), helix 71−88 has a kink, as indicated by the central curved line. The corresponding helix 428−445 in the PP domain (C) has a similar kink. Asterisks indicate the location of the catalytic glutamate (A, residue 47) and the metal ion-binding residues D428 and N455 (C).

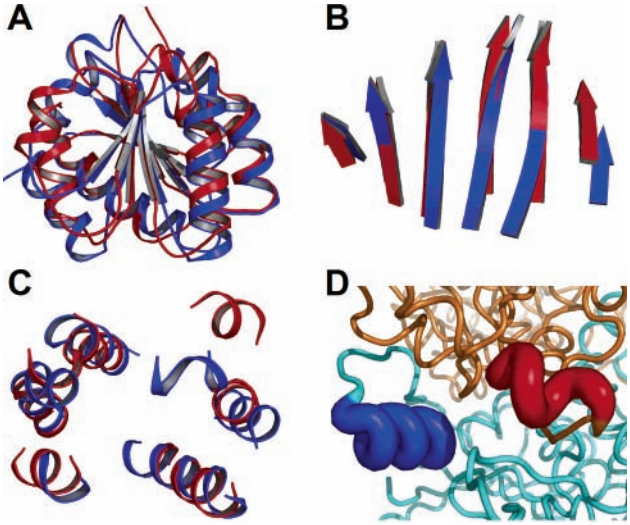

FIGURE 3. Comparison of the Pyr and PP domains of BFDC. A shows an overlay of the two domains. The central β sheets of the two domains are shown in B, while C shows their α helices. The Pyr domain is colored blue, and the PP domain is colored red. D shows the interface between a pair of subunits with thick segments emphasizing the regions that differ most between domains. These are helix 100−105 (shown in red) from the Pyr domain of the gold subunit and helix 458−469 (in blue) from the PP domain of the cyan subunit.

for enzyme activity. It seems that, after the duplication event that allowed the two domains to follow separate evolutionary paths, the basic structure has been preserved intact, apart from two regions needed to promote assembly of the active dimer.

Pyr and PP Domains. The core structure of both the Pyr and PP domains is quite well-conserved across the five families. An example of each is shown in Figure 4, comparing Pyr (A; rmsd = 1.63 Å for 95 Cα atoms) and PP (B; rmsd = 1.23 Å for 118 Cα atoms) of BFDC and PDH. Overlaying all five structures for each domain (not shown) clearly supports the idea of a common ancestry for all ThDP-dependent enzymes. Nevertheless, the Pyr and PP domains have undergone some divergence, with the most elaborate changes in the K1 family exemplified by PDH. Thus, the overlay in Figure 4B shows a contiguous segment of BFDC (residues 350−525), but for PDH, three discontinuous segments are depicted (residues 105−157, 193−295, and 361−393). When the topology of the entire domain (Figure 4C) is compared with the PP domain of BFDC (Figure 2C), the underlying differences are revealed (colored red in Figure 4C). First, there are extensions to both ends of the domain in PDH, composed of two and three α helices at the N and C termini, respectively. Second, a helix−loop (residues 458−475; Figure 2C) in BFDC is missing and is replaced in PDH by a short hairpin segment that contains a very small, two-stranded, antiparallel β sheet (residues 262−267). Note that the BFDC helix 458−469 is one of the segments mentioned earlier as mediating subunit interactions (Figure 3D). In PDH, the small β sheet is also at the subunit interface. Third, there is a four-helix insert prior to the final strand of the main β sheet, and this forms a discrete lobe in the three-

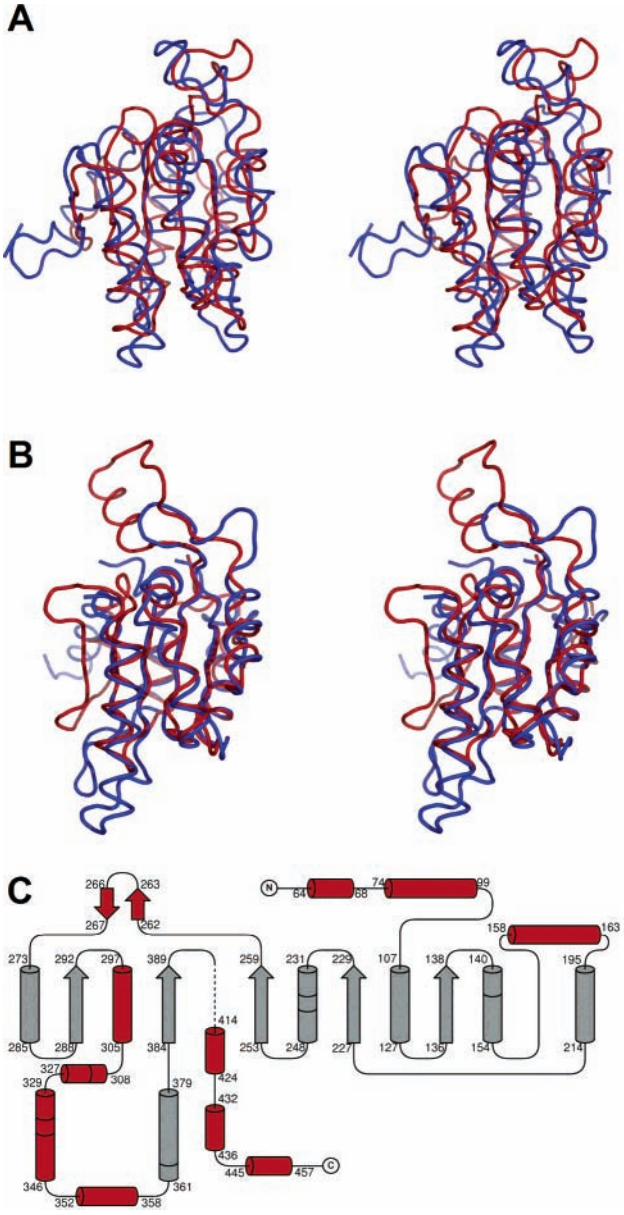

FIGURE 4. Pyr and PP domains of BFDC and PDH. A shows in stereo an overlay of the Pyr domain of BFDC (red; residues 2−176) and the Pyr core of PDH (blue; residues 492−540 and 558−700). B shows in stereo an overlay of the PP domain of BFDC (red; residues 350−525) and the PP core of PDH (blue; residues 105−157, 193−295, and 361−393). C shows a topology cartoon for the complete PP domain of PDH (residues 56−471, apart from residues 401−413 that are undefined in the structure and shown by a broken line). The representations are the same as described in Figure 2. Regions shown in red differ from the core PP domain.

dimensional structure.[13] Finally, the β strand corresponding to residues 395−397 in BFDC is missing, replaced by α helix 158−163 in PDH. This helix cannot be regarded as the structural equivalent of the missing BFDC β strand, because it occupies a very different position in space (not shown).

Third Domain of DC and TK. The idea that the structure of TK is a circular permutation of the DC family raises the question of the possible structural relationship between the third domains of two families of proteins.

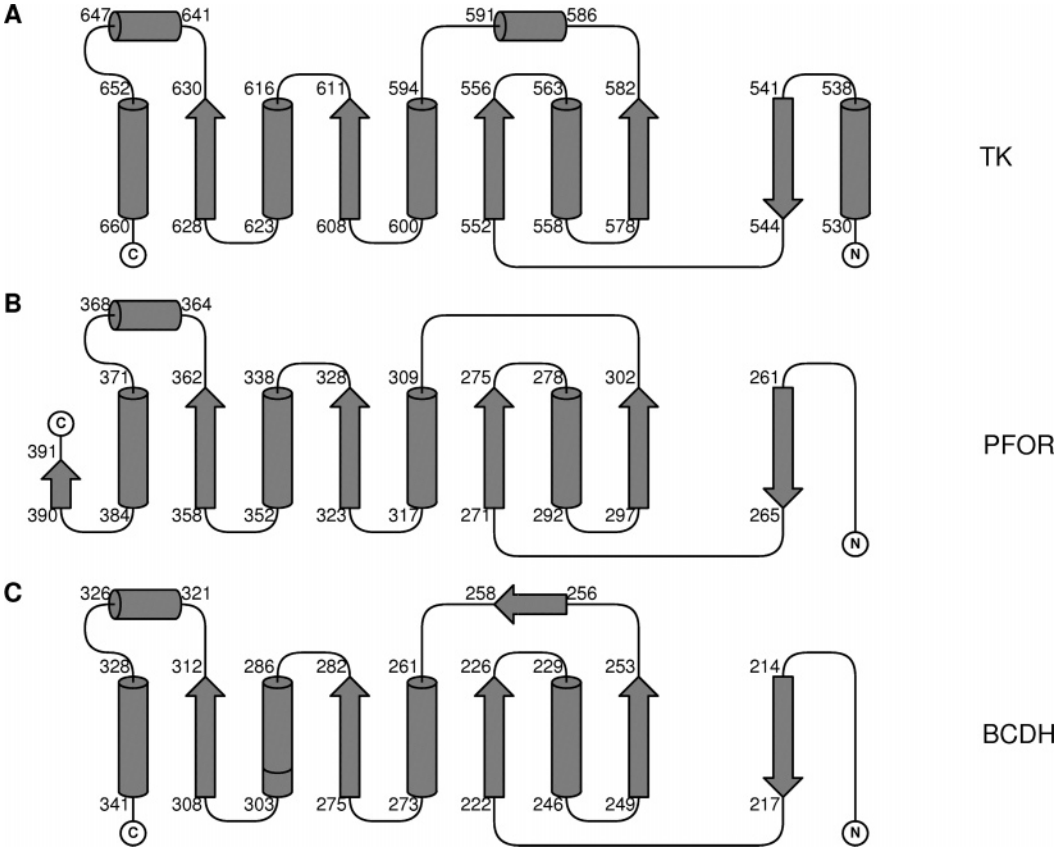

FIGURE 5. Topology cartoons for the third domain of TK (A) and its structural relatives in PFOR (domain II, B) and BCDH (β subunit, C). The representations are the same as described in Figure 2.

Despite the different order of domains in the protein sequence, this third domain occupies a similar position in space. To investigate whether the basic structure of the C-terminal domain of TK resembles that of the middle domain of DC, their topologies were examined. As shown in Figure 5A, this domain in TK contains a five-stranded β sheet with the first strand antiparallel to the remaining four. Its connectivity is clearly different from that of the middle domain of DC (Figure 2B), making a common origin highly unlikely. The possible origins of these domains were investigated.

A close structural match (Figure 6A; rmsd = 1.43 Å for 88 Cα atoms) was found between this DC domain and $NADP^+$/NADPH-binding component dIII of the proton-translocating transhydrogenase (1U2D from *Rhodospirillum rubrum*).[20] This is a significant observation because this domain in some DC proteins also binds nucleotides: FAD in POX[3] and AHAS[12] or ADP in OCDC.[21] If this apparent similarity between dIII and POX, AHAS, and OCDC really reflects an evolutionary relationship, then the nucleotide should be bound in an equivalent position in all four proteins. It has been shown previously that this is true in comparisons between AHAS and POX[12] or OCDC and AHAS.[21] Therefore, I overlaid the protein component of dIII upon the middle domain of AHAS[22] and compared the location of the nucleotides (parts B and C of Figure 6). There is very good agreement between the position of $NADP^+$ in dIII and FAD in AHAS, confirming the hypothesis that the middle domain of the AHAS (and by implication of all of the DC proteins) is derived from dIII or its ancestor. I suggest that, in the evolution of the DC proteins, the middle domain was first recruited for its

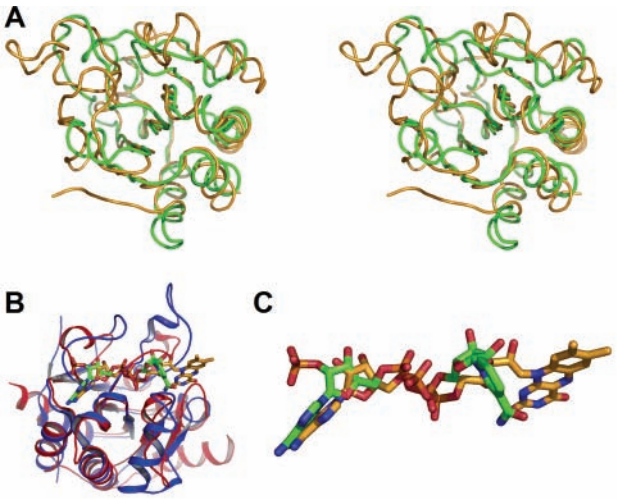

FIGURE 6. Structural similarities between dIII and the middle domain of DC proteins. A shows in stereo an overlay of the middle domain of BFDC (green; residues 183−330) and the component dIII of the proton-translocating transhydrogenase from *R. rubrum* (gold; residues 30−203). B shows an overlay of the middle domain of AHAS (red; residues 281−458) and this dIII protein (blue). FAD is shown in a stick representation with carbon colored gold, nitrogen colored blue, oxygen colored red, and phosphorus colored orange. $NADP^+$ is shown with the same color coding, except that carbon is green. C shows a close-up of the nucleotides illustrated in B.

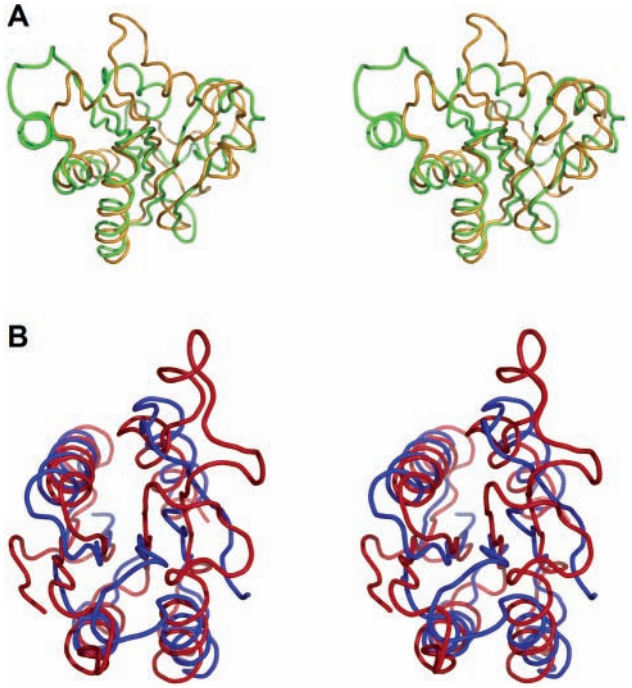

FIGURE 7. Structural comparisons of domain X. A shows in stereo an overlay of residues 520−663 of TK (green) with residues 259−415 of PFOR (gold). B shows in stereo an overlay of residues β199−β342 of BCDH (red) with residues 2−121 of 1MB0 (blue).

nucleotide-binding properties. During further development, this property has been lost by BFDC and most other proteins in the family but retained by the three enzymes mentioned above.

Unlike the third domain of the DC family, the third domain in TK has structural equivalents in each of the remaining families. Figure 5 compares its topology in TK, PFOR (domain II), and BCDH (β subunit, K2 family), while Figure 7A compares the three-dimensional structure of this domain of TK with that of PFOR (rmsd = 1.51 Å for 54 Cα atoms). A similar domain is observed in the K1 family. The structures are clearly related, although the ultimate origin of this domain is not clear. Searching the PDB (using the domain from BCDH as the probe) gave the best match with 1MB0 (Figure 7B; rmsd = 1.45 Å for 62 Cα atoms), which is a cell-division response regulator from *Caulobacter crescentus*.[23] This protein is part of a family of bacterial regulatory proteins with a flavodoxin-like fold. However, there is no obvious biological rationale for a flavodoxin becoming a domain of a ThDP-dependent enzyme. Therefore, until the origin of this domain is more clearly established, it will be designated as domain X.

Evolution of ThDP-Dependent Enzymes. The structural relationships described here allow a possible evolutionary tree to be proposed (Figure 8). The most ancient common ancestor of all ThDP-dependent enzymes is a protein with a core of approximately 170 residues consisting of a six-stranded, parallel β sheet with linking loops and α helices. This protein would have possessed binding sites for both the pyrimidine and diphosphate ends of the cofactor. Because these two binding sites would have been located in different parts of this ancestral protein (see asterisks in parts A and C of Figure 2), dimerization would create two interfaces that would each bind ThDP (inset A in Figure 8). This binding could enhance the intrinsic catalytic activity of the cofactor,[24] creating a primitive 2-ketoacid decarboxylase. Duplication of the gene would allow independent evolution of the two resulting proteins, creating a more efficient enzyme by formation of an $\alpha_2\beta_2$-heterotetramer. A remnant of this structure still exists, as deduced from gene sequencing; sulfopyruvate decarboxylase (SPDC) contains two types of subunits corresponding to the Pyr and PP domains.[25] It should be pointed out that in Figure 8 only half of this heterotetramer is shown. The remainder of this figure, apart from the two boxed insets, is simplified in a similar way.

After the appearance of this $\alpha_2\beta_2$-heterotetramer, there is a divergence, with one branch leading to the DC family and the other being the starting point for the remaining four families. For the DC group, fusion of the Pyr and PP domain genes would be the next step on the pathway, and again, there is an extant remnant of this intermediate, phosphopyruvate decarboxylase (PPDC).[25] Although the three-dimensional structure has not been determined, the gene sequence shows clearly that it contains the Pyr and PP domains only. At this point, the transhydrogenase dIII subunit (TH3) was recruited and the gene interposed between the Pyr and PP coding sequence creating the structure that is characteristic of the main members of the DC family. Most likely, there was an intermediate structure in which TH3 was a separate subunit and gene fusion was a later event. In further evolution, most of the DC family (e.g., BFDC) lost the $NADP^+$/NADPH binding properties of TH3, while in a few, the specificity was altered to FAD (POX and AHAS) or ADP (OCDC).

The branch leading to the remaining four families starts by fusion of the gene for protein X, creating a new heterotetramer that would resemble the TK family, except that the PP and Pyr domains were on separate subunits. Fusion of the genes for the two subunits led to the TK family. This precursor would also have developed into the K2 family; then, fusion of the genes encoding the α and β subunits finally led to the K1 proteins. Clearly, this is a simplification of the true situation because both the K1 and K2 proteins have acquired additional segments, in both the Pyr and PP (e.g., Figure 4C) domains, some of which are undoubtedly needed for interaction with other components of this large multienzyme complex.[26] Note that, despite the identical order of domains in TK and K1, they probably evolved separately from a common ancestor. It is proposed that K2 is on the direct pathway to K1.

As shown in Figure 8, the C terminus of the Pyr domain in the DC family is connected through the TH3 domain to the N terminus of the PP domain. In contrast, the K1, K2, and TK families all have the C terminus of the PP domain connected directly to the N terminus of the Pyr domain. However, this different connectivity does not imply that the interface between these two domains is different between the families. In all cases, the domain interface developed at a stage when the Pyr and PP domains were separate subunits, as they have remained

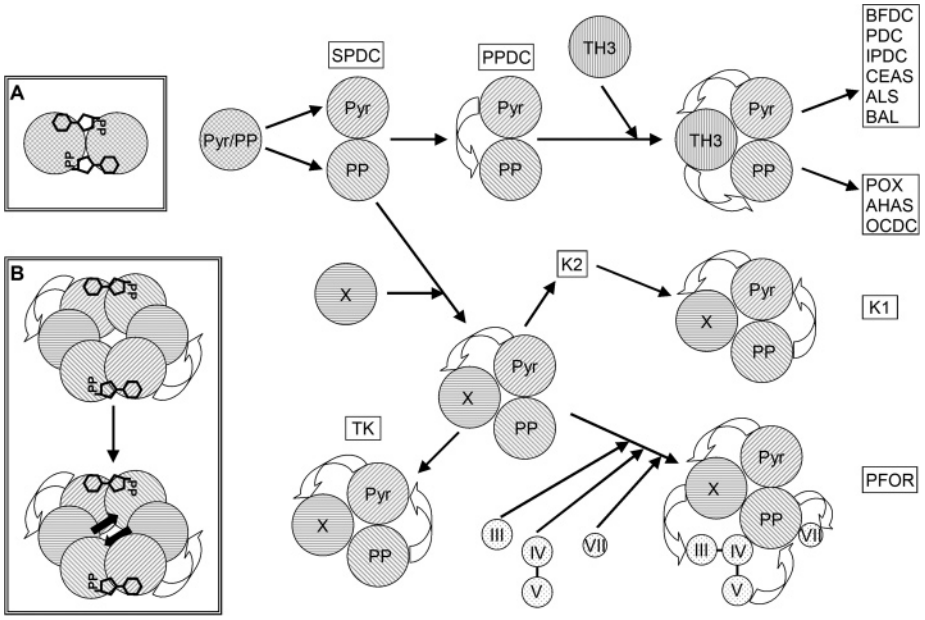

FIGURE 8. Proposed evolutionary pathway of ThDP-dependent enzyme families. The main figure shows how a primordial Pyr/PP protein underwent duplication and subsequent fusions (indicated by curved white arrows) and the addition of extra domains to create the extant enzymes. In all cases, the active enzyme would contain two copies (or a multiple thereof) of the unit shown. Domains are not drawn to scale; different shadings distinguish various types of the domain. Abbreviations not defined elsewhere are indolepyruvate decarboxylase (IPDC) and N^2-(2-carboxyethyl)arginine synthase (CEAS). Inset A shows how a dimer of the primordial Pyr/PP protein would bind ThDP to form a primitive enzyme. Inset B shows how domain fusion (indicated by the black arrows) would allow PFOR to possess a ThDP-binding site within a single subunit.

in SPDC. The way in which the genes subsequently fused has not perturbed this interface.

The evolution of the OR family followed a somewhat more complex route.[27] Only the structure of the *D. africanus* enzyme has been solved, in which there are seven domains numbered I−VII from the N terminus. These correspond to Pyr (domain I), PP (domain VI), and X (domain II), plus four other domains as described elsewhere.[6] Domain V has a structural similarity to ferredoxin. Domains III−V and VII were evidently recruited in stages as illustrated in Figure 8, because there are examples of ORs from other organisms that appear to be intermediates and/or side branches of the pathway that led to the *D. africanus* enzyme. Thus, the enzyme from *Halobacterium halobium* contains one subunit consisting of domains III, I, and II and a second subunit equivalent to domain VI. PFOR from *Thermotoga maritima* contains four subunits (α, β, γ, and δ) corresponding to domains I + II, VI, III, and IV + V, respectively. Fusion of these genes, plus the recruitment of the small domain VII, completed the development of *D. africanus* PFOR. According to the scheme shown in Figure 8, the common ancestor would have contained two subunits corresponding to domains I + II and VI and the present day ORs developed by adding extra subunits and domains. Kletzin and Adams[27] postulate a somewhat different sequence of events, with an ancestral type containing four subunits (similar to *T. maritima* PFOR) that underwent both gene fusions (*D. africanus* type) and gene loss (*H. halobium* type).

There is one final aspect of the structure of PFOR (and, presumably, the entire OR family) that merits a comment. In contrast to the other four families, ThDP is bound across the Pyr/PP interface within the same PFOR subunit. This difference may be explained by the way in which the X−PP fusion event occurred. As illustrated for a simplified PFOR (omitting domains III−V and VII) in Figure 8 (inset B), fusion across the neighboring subunit (thick black arrows) links the Pyr and PP domains to create a dimeric PFOR, with its ThDP-binding site totally contained within one subunit.

Preservation of the Chemical Mechanism. The structural changes that have occurred during the evolution of the various families of ThDP-dependent enzymes have been accompanied by alterations in the type of reaction catalyzed and the substrate specificity. However, as is illustrated in Figure 1, these changes have been achieved without major modification of the underlying chemical mechanism. The active site may have changed to recognize and correctly orient the various donor substrates, but in all cases, the reaction passes through a very similar enamine/α-carbanion intermediate. This then reacts with various acceptor substrates whose identity is again determined with relatively minor changes in the constellation of amino acid side chains that make up the active site. The most radical changes occur where completion of the catalytic cycle involves an oxidation (POX and the K1, K2, and OR families). In these cases, additional domains or subunits provide conduits for electron flow. However, the conformation and orientation of ThDP is preserved with minimal change[5] because of the similarity of the Pyr/PP interface across the protein families. This interface first arose at an early evolutionary stage, when the Pyr and PP domains were separate subunits (Figure 8).

Conclusions

Gene sequencing has identified many actual and potential ThDP-dependent enzymes across all living organisms. In almost every case, the sequence allows the protein to be allotted to one of the major families that have been identified from three-dimensional structure determination. The only ones that do not fit within these families appear to be modified versions that have gained, lost, or not acquired one or more domains or subunits (e.g., PPDC and *H. halobium* PFOR). In light of the overall structural similarity between the K1 and K2 enzymes, they should probably be merged into a single 2-ketoacid dehydrogenase (KD) family. Therefore, four families (DC, TK, OR, and KD) seem to cover all structural possibilities for this group of enzymes. It appears that there was a small ancestral homodimeric decarboxylase of about 170 residues and a duplication event created the basic Pyr and PP domains found in all of these enzymes. This basic structure has diversified through recruitment of other domains and insertion of extra sequences. The origin of some of these has been established, but for others, such as domain X found in all but the DC family, a flavodoxin ancestor is possible; however, this remains to be proven.

Drs. Luke Guddat and Jennifer McCourt made some useful comments on this Account. The author also thanks the Australian Research Council for many years of grant support.

References

(1) Pohl, M.; Sprenger, G. A.; Müller, M. A new perspective on thiamine catalysis. *Curr. Opin. Biotechnol.* **2004**, *15*, 335−342.
(2) Lindqvist, Y.; Schneider, G.; Ermler, U.; Sundström, M. Three-dimensional structure of transketolase, a thiamine diphosphate dependent enzyme, at 2.5 Å resolution. *EMBO J.* **1992**, *11*, 2373−2379.
(3) Muller, Y. A.; Schulz, G. E. Structure of the thiamine- and flavin-dependent enzyme pyruvate oxidase. *Science* **1993**, *259*, 965−967.
(4) Dyda, F.; Furey, W.; Swaminathan, S.; Sax, M.; Farrenkopf, B.; Jordan, F. Catalytic centers in the thiamin diphosphate dependent enzyme pyruvate decarboxylase at 2.4-Å resolution. *Biochemistry* **1993**, *32*, 6165−6170.
(5) Mosbacher, T. G.; Mueller, M.; Schulz, G. E. Structure and mechanism of the ThDP-dependent benzaldehyde lyase from *Pseudomonas fluorescens*. *FEBS J.* **2005**, *272*, 6067−6076.
(6) Chabrière, E.; Charon, M.-H.; Volbeda, A.; Pieulle, L.; Hatchikian, E. C.; Fontecilla-Camps, J.-C. Crystal structures of the key anaerobic enzyme pyruvate:ferredoxin oxidoreductase, free and in complex with pyruvate. *Nat. Struct. Biol.* **1999**, *6*, 182−190.
(7) Nakazawa, M.; Inuia, H.; Yamaji, R.; Yamamoto, T.; Takenaka, S.; Ueda, M.; Nakano, Y.; Miyatake, K. The origin of pyruvate:NADP$^+$ oxidoreductase in mitochondria of *Euglena gracilis*. *FEBS Lett.* **2000**, *479*, 155−156.
(8) Muller, Y. A.; Lindqvist, Y.; Furey, W.; Schulz, G. E.; Jordan, F.; Schneider, G. A thiamin diphosphate binding fold revealed by comparison of the crystal structures of transketolase, pyruvate oxidase and pyruvate decarboxylase. *Structure* **1993**, *1*, 95−103.
(9) Candy, J. M.; Duggleby, R. G. Structure and properties of pyruvate decarboxylase and site-directed mutagenesis of the *Zymomonas mobilis* enzyme. *Biochim. Biophys. Acta* **1998**, *1385*, 323−338.
(10) Hawkins, C. F.; Borges, A.; Perham, R. N. A common structural motif in thiamin pyrophosphate-binding enzymes. *FEBS Lett.* **1989**, *255*, 77−82.
(11) Ævarsson, A.; Seger, K.; Turley, S.; Sokatch, J. R.; Hol, W. G. Crystal structure of 2-oxoisovalerate and dehydrogenase and the architecture of 2-oxo acid dehydrogenase multienzyme complexes. *Nat. Struct. Biol.* **1999**, *6*, 785−792.
(12) Pang, S. S.; Duggleby, R. G.; Guddat, L. W. Crystal structure of yeast acetohydroxyacid synthase: A target for herbicidal inhibitors. *J. Mol. Biol.* **2002**, *317*, 249−262.
(13) Arjunan, P.; Nemeria, N.; Brunskill, A.; Chandrasekhar, K.; Sax, M.; Yan, Y.; Jordan, F.; Guest, J. R.; Furey, W. Structure of the pyruvate dehydrogenase multienzyme complex E1 component from *Escherichia coli* at 1.85 Å resolution. *Biochemistry* **2002**, *41*, 5213−5221.
(14) Polovnikova, E. S.; McLeish, M. J.; Sergienko, E. A.; Burgner, J. T.; Anderson, N. L.; Bera, A. K.; Jordan, F.; Kenyon, G. L.; Hasson, M. S. Structural and kinetic analysis of catalysis by a thiamin diphosphate-dependent enzyme, benzoylformate decarboxylase. *Biochemistry* **2003**, *42*, 1820−1830.
(15) Chabrière, E.; Vernède, X.; Guigliarelli, B.; Charon, M. H.; Hatchikian, E. C.; Fontecilla-Camps, J. C. Crystal structure of the free radical intermediate of pyruvate:ferredoxin oxidoreductase. *Science* **2001**, *294*, 2489−2490.
(16) Wynn, R. M.; Machius, M.; Chuang, J. L.; Li, J.; Tomchick, D. R.; Chuang, D. T. Roles of His^{291}-α and His^{146}-β' in the reductive acylation reaction catalyzed by human branched-chain α-ketoacid dehydrogenase: Refined phosphorylation loop structure in the active site. *J. Biol. Chem.* **2003**, *278*, 43402−43410.
(17) Krissinel, E.; Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2256−2268.
(18) Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714−2723.
(19) Bond, C. S. TopDraw: A sketchpad for protein structure topology cartoons. *Bioinformatics* **2003**, *19*, 311−312.
(20) Mather, O. C.; Singh, A.; van Boxel, G. I.; White, S. A.; Jackson, J. B. Active-site conformational changes associated with hydride transfer in proton-translocating transhydrogenase. *Biochemistry* **2004**, *43*, 10952−10964.
(21) Berthold, C. L.; Moussatche, P.; Richards, N. G. J.; Lindqvist, Y. Structural basis for activation of the thiamin diphosphate-dependent enzyme oxalyl-CoA decarboxylase by adenosine diphosphate. *J. Biol. Chem.* **2005**, *280*, 41645−41654.
(22) McCourt, J. A.; Pang, S. S.; Duggleby, R. G.; Guddat, L. W. Elucidating the specificity of binding of sulfonylurea herbicides to acetohydroxyacid synthase. *Biochemistry* **2005**, *44*, 2330−2338.
(23) Guillet, V.; Ohta, N.; Cabantous, S.; Newton, A.; Samama, J. P. Crystallographic and biochemical studies of DivK reveal novel features of an essential response regulator in *Caulobacter crescentus*. *J. Biol. Chem.* **2002**, *277*, 42003−42010.
(24) Krampitz, L. O. Catalytic functions of thiamin diphosphate. *Annu. Rev. Biochem.* **1969**, *38*, 213−240.
(25) Graupner, M.; Xu, H.; White, R. H. Identification of the gene encoding sulfopyruvate decarboxylase, an enzyme involved in biosynthesis of coenzyme M. *J. Bacteriol.* **2000**, *182*, 4862−4867.
(26) Frank, R. A. W.; Pratap, J. V.; Pei, X. Y.; Perham, R. N.; Luisi, B. F. The molecular origins of specificity in the assembly of a multienzyme complex. *Structure* **2005**, *13*, 1119−1130.
(27) Kletzin, A.; Adams, M. W. Molecular and phylogenetic characterization of pyruvate and 2-ketoisovalerate ferredoxin oxidoreductases from *Pyrococcus furiosus* and pyruvate ferredoxin oxidoreductase from *Thermotoga maritima*. *J. Bacteriol.* **1996**, *178*, 248−257.

AR068022Z