Structure of the Pyruvate Dehydrogenase Multienzyme Complex E1 Component from *Escherichia coli* at 1.85 Å Resolution[†,‡]

Palaniappa Arjunan,[$,§,||] Natalia Nemeria,[||,⊥] Andrew Brunskill,[$,§] Krishnamoorthy Chandrasekhar,[$,§] Martin Sax,[$] Yan Yan,[⊥] Frank Jordan,[⊥] John R. Guest,[#] and William Furey*[,$,§]

Biocrystallography Laboratory, Veterans Affairs Medical Center, P. O. Box 12055, University Drive C, Pittsburgh, Pennsylvania 15240, USA, Department of Pharmacology, University of Pittsburgh School of Medicine, 1340 BSTWR, Pittsburgh, Pennsylvania 15261, USA, Department of Chemistry, Rutgers University, Newark, New Jersey 07102, USA, and Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, S100 2TN, United Kingdom

Received September 27, 2001; Revised Manuscript Received January 22, 2002

ABSTRACT: The crystal structure of the recombinant thiamin diphosphate-dependent E1 component from the *Escherichia coli* pyruvate dehydrogenase multienzyme complex (PDHc) has been determined at a resolution of 1.85 Å. The *E. coli* PDHc E1 component E1p is a homodimeric enzyme and crystallizes with an intact dimer in an asymmetric unit. Each E1p subunit consists of three domains: N-terminal, middle, and C-terminal, with all having α/β folds. The functional dimer contains two catalytic centers located at the interface between subunits. The ThDP cofactors are bound in the "V" conformation in clefts between the two subunits (binding involves the N-terminal and middle domains), and there is a common ThDP binding fold. The cofactors are completely buried, as only the C2 atoms are accessible from solution through the active site clefts. Significant structural differences are observed between individual domains of E1p relative to heterotetrameric multienzyme complex E1 components operating on branched chain substrates. These differences may be responsible for reported alternative E1p binding modes to E2 components within the respective complexes. This paper represents the first structural example of a functional pyruvate dehydrogenase E1p component from any species. It also provides the first representative example for the entire family of homodimeric ($\alpha 2$) E1 multienzyme complex components, and should serve as a model for this class of enzymes.

The pyruvate dehydrogenase multienzyme complex (PDHc)[1] converts the 3-carbon product of glycolysis, pyruvic acid, to acetyl-coenzyme A, one of two compounds needed for condensation to citrate and required for the tricarboxylic acid (Krebs, or citric acid) metabolic cycle:

$$\text{pyruvate} + \text{CoA} + \text{NAD}^+ \rightarrow \text{acetyl-CoA} + \text{CO}_2 + \text{NADH} + \text{H}^+ \quad (1)$$

It is a nearly ubiquitous complex and is arguably the most important member of a larger class of related multienzyme complexes, the 2-oxoacid dehydrogenase multienzyme complexes. Because of its location at such a key junction in sugar metabolism, PDHc is subject to considerable regulation, both homotropic [by substrate and the thiamin diphosphate (ThDP, the vitamin B1 diphosphate) and Mg(II) cofactors], and heterotropic. The fundamental reactions of this complex are carried out by three enzymatic components, named E1, E2, and E3. While there is much structural information available for E3 (a flavoenzyme), for E2 there is only partial structural information in the literature. So far as the E1 component is concerned, there have been recognized three types of such 2-oxoacid dehydrogenase enzymes, an $\alpha 2$ homodimer, an $\alpha 2\beta 2$ heterotetramer, both in lower organisms, and in mammals an $\alpha 2\beta 2$ heterotetramer subject to regulation by a kinase and phosphatase system. Hol and collaborators just in the recent past have reported structures of E1 from the second and third type, both specific for a branched chain 2-oxo acid (*1*, *2*). The best characterized 2-oxoacid dehydrogenase multienzyme complex with an $\alpha 2$ homodimer E1 quaternary structure is from *Escherichia coli* (*3*), with the three enzymatic components: pyruvate dehydrogenase [EC 1.2.4.1; E1, molecular weight 99 474 utilizing thiamin diphosphate (ThDP) as a cofactor (*4*)]; dihydrolipoamide

[†] Supported at Pittsburgh by NIH-GM-61791 (to W.F.), at Rutgers by NIH-GM-50380, NIH-GM-62330, and the Rutgers University Busch Fund (to F.J.), and at Sheffield by the Biotechnology and Biological Sciences Research Council (UK, to J.G.)

[‡] Coordinates have been deposited in the Protein Data Bank with the accession code 1L8A.

* To whom correspondence should be addressed. Tel. 412-683-9718; fax: 412-688-6945; e-mail: fureyw@pitt.edu.

[$] Veterans Affairs Medical Center.

[§] University of Pittsburgh School of Medicine.

[⊥] Rutgers University.

[#] University of Sheffield.

[||] These authors contributed equally to the success of this study.

[1] Abbreviations: PDHc, pyruvate dehydrogenase multienzyme complex; ThDP, thiamin diphosphate, vitamin B1 diphosphate; E1, E2, E3, first, second and third enzymatic components, respectively, of multienzyme complexes related to and including PDHc; E1p, first enzymatic component of the dehydrogenase multienzyme complex with specificity for pyruvate; BDHc, branched chain dehydrogenase multienzyme complex; E1b, first enzymatic component of dehydrogenase multienzyme complexes with specificity for branched chains; PEG, polyethylene glycol; MAD, multiple anomalous diffraction; NCS, noncrystallographic symmetry; rms, root-mean square.

10.1021/bi0118557 CCC: $22.00
Published on Web 03/28/2002

acetyltransferase [EC 2.3.1.12; E2, molecular weight 65 959, which contains covalently bound lipoyl groups (*5*)]; and lipoamide dehydrogenase [EC 1.8.1.4; E3, molecular weight 50,554, containing tightly bound FAD (*6*)]. This multienzyme complex has the stoichiometry 24 E1, 24 E2, and 12 E3 (corresponding to a total calculated molecular weight of 4.57×10^6) and performs the following series of reactions:

$$\text{pyruvate} + \text{E1-ThDP-Mg}^{2+} \rightarrow \text{E1-hydroxyethylidene-ThDP-Mg}^{2+} + \text{CO}_2 \quad (2)$$

$$\text{E1-hydroxyethylidene-ThDP-Mg}^{2+} + \text{E2-lipoamide} \rightarrow \text{E1-ThDP-Mg}^{2+} + \text{E2-acetyldihydrolipoamide} \quad (3)$$

$$\text{E2-acetyldihydrolipoamide} + \text{CoA} \rightarrow \text{E2-dihydrolipoamide} + \text{acetyl-CoA} \quad (4)$$

$$\text{E2-dihydrolipoamide} + \text{E3-FAD} \rightarrow \text{E2-lipoamide} + \text{E3-FADH}_2 \quad (5)$$

$$\text{E3-FADH}_2 + \text{NAD}^+ \rightarrow \text{E3-FAD} + \text{NADH} \quad (6)$$

Although much is known about ThDP catalysis in general, very little structural information on the important PDHc E1 component is available yet. Very recently, Eisenberg and collaborators reported (*7*) the crystal structure of the PDHc E1β subunit from *Pyrobaculum aerophilum*, a member of the $\alpha2\beta2$ class. This subunit however, is nonfunctional in the absence of its α subunit counterpart as it cannot bind cofactors or assemble into the appropriate quaternary structure. Thus, this structure provided no information about the active site or catalytic mechanism, although it provided some information about helix−helix packing at β subunit interfaces. We here report the high-resolution X-ray structure determination results for recombinant E1p from *E. coli* PDHc. This structure is the first example of an $\alpha2$ homodimeric E1 quaternary structure and is also the first example of a functional E1 with specificity for pyruvate. A BLAST protein search of recently sequenced genomes revealed remarkably high degrees of identity between the *E. coli* E1p subunits and E1p subunits from gram negative bacteria (such as *Vibrio cholerae*, *Haemophilus influenzae*, *Azotobacter vinelandii*, *Neisseria meningitidis*), as well as from actinomycetes (such as *Mycobacterium leprae* and *Mycobacterium tuberculosis*). Therefore, the structure here reported takes on special significance as the first representative from a family of related enzymes crucial for metabolism in many pathogenic bacteria.

EXPERIMENTAL PROCEDURES

The homodimeric ($\alpha2$) E1p structure from *E. coli* was determined by MAD (multiple anomalous diffraction) phasing methods supplemented with solvent flattening and NCS (noncrystallographic symmetry) averaging. Recombinant E1p enzyme from *E. coli* was purified and assayed as previously reported (*8*, *9*). The E1p protein was crystallized by the sitting drop vapor diffusion method by mixing equal amounts of protein and precipitant solutions. The best crystals were obtained with a reservoir solution containing 15−20% PEG2000 monomethyl ether, 10% propanol, 0.2% NaN_3, and 60 mM Hepes buffer (pH 7.05) at 22 °C. Drops were 6−10 μL consisting of equal parts of reservoir and protein solution. The crystal growth took between 2 and 5 weeks. The crystals typically have dimensions of $0.15 \times 0.30 \times 0.50$ mm.

An initial low-temperature (−180 °C) native X-ray diffraction data set was collected from a single crystal on a Rigaku RU200 rotating anode with a Bruker multiwire area detector. The crystals are monoclinic with space group $P2_1$ and cell constants $a = 82.18$, $b = 142.66$, $c = 82.88$ Å, and $\beta = 102.05°$. For structure determination, a data set from a native E1p crystal and data at three wavelengths from a selenomethionine analogue crystal were collected at beam line X12C of the National Synchrotron Light Source at Brookhaven National Laboratory. The E1p crystal diffracted to 1.85 Å resolution and the selenomethionyl substituted crystal to 2.30 Å resolution. The synchrotron data sets were processed with the DENZO (*10*) software program and yielded essentially the same cell parameters as from the in-house data. The Matthews coefficient V_m is 2.37 Å^3 Da^{-1}, based on 4 subunits/cell. The V_m value is in the standard range of 1.7−3.5 and is derived from the assumption that one dimer with a mass of 199 kDa (including cofactors) is in the asymmetric unit. The presence and orientation of a noncrystallographic 2-fold axis was identified by self-rotation function analysis.

The direct methods program SnB (*11*) was used with peak anomalous wavelength data to 3.50 Å resolution to easily locate the selenium atoms in an asymmetric unit. The top 40 peaks from the SnB solution were found to be consistent with the known NCS operator direction and were accepted as selenium sites. All 40 sites were subsequently found to be correct. Although there are 42 selenium atoms in the asymmetric unit, two are disordered and were not observed, even in the final structure. After refinement of the 40 selenium atomic parameters, a set of MAD phases was obtained with an overall figure of merit 0.78 in the 20 to 2.60 Å resolution range. Solvent flattening and NCS averaging then improved the phase set, and extended the resolution to 2.30 Å. All selenium atom refinement, phasing, averaging, solvent flattening, and map calculations were performed with the PHASES (*12*) package on an SGI workstation. The resulting 2.30 Å electron density map was of very good quality, and nearly all of the polypeptide chain (Cα-trace) was readily obtained using the graphics program O (*13*). The map used for chain tracing was computed from phases *after phase combination with the original MAD phases* at the end of the averaging iterations, and not the averaged map itself. Thus, the two subunits were not forced to be strictly identical (although they were very similar). Accordingly, both subunits were traced independently. The E1p amino acid sequence from *E. coli* was used to fit the side chains into the electron density map. The current model is approximately 91% complete, with three missing regions as yet unidentified due to weak and ambiguous electron density. The missing regions and corresponding chain breaks are identical in both subunits.

The model was refined by simulated annealing without imposing any NCS using the program XPLOR (*14*) with the synchrotron native data set to 1.85 Å resolution. The refinement procedure included periodic examinations of omit and difference maps, as well as the introduction of water molecules and individual *B* values. A final examination of difference electron density maps still did not produce interpretable electron density for the three missing regions

Table 1: Crystallographic Data and Refinement Statistics

sp. gp. and cell constants, $P2_1$	$a = 82.18$ Å	$b = 142.66$ Å	$c = 82.88$ Å	$\beta = 102.05°$
data collection	native	Se-Met l1	Se-Met $\lambda 2$	Se-Met $\lambda 3$
resolution (Å)	1.85	2.30	2.30	2.30
completeness (%) all, (last shell)	92.1 (70.2)	99.9 (99.9)	99.9 (99.9)	99.9 (99.9)
total reflections	528 145	615 083	619 947	632 674
unique reflections	146 183	82 749	82 749	82 749
R_{merge}	0.054	0.132	0.128	0.117
$<I/\sigma>$ all, (last shell)	17.9 (3.9)	10.9 (3.9)	11.2 (4.1)	11.8 (4.7)
X-ray source	X12C	X12C	X12C	X12C
wavelength (Å)	1.07200	0.97932	0.97904	0.93000
phasing power				
iso		1.89	1.59	
ano		3.47	3.45	2.09
Refinement				
resolution range (Å)	8.00−1.85			
no. of reflections ($I > 1\sigma$)	125 670			
R-factor all, (last shell)	0.189 (0.236)			
R_{free} all, (last shell)	0.236 (0.261)			
no. of residues	1602			
no. of protein atoms	12 740			
no. of water molecules	682			
average B factor (Å^2)	17.5			
rms deviations				
bond lengths (Å)	0.006			
bond angles (°)	1.273			

(residues 1−55, 401−413, and 541−557), and hence they remain absent in the model. Mass spectrometry analyses of washed and dissolved crystals then confirmed these regions are in fact present in the crystals and must be disordered, as opposed to having been cleaved away by proteolysis prior to crystallization. For the data between 8.00 and 1.85 Å resolution with $I/\sigma > 1$, the *R*-factor is 0.189 (0.236 in the last shell) and the R_{free} (based on 6% of the data) is 0.236 (0.261 in the last shell). The rms deviations from ideal geometry are 0.006 Å for bond lengths and 1.273° for bond angles. Data collection and refinement statistics are given in Table 1. The final model contains two independent subunits, each with 801 amino acids out of a possible 886, and two ThDP−Mg cofactor pairs. This model also includes 682 water molecules and two cis-prolines (Pro463) per asymmetric unit. The model was analyzed with PROCHECK (*15*), and 89% of the residues are in the most favored regions of the Ramachandran plot, with only a single residue per subunit (Asp521) in an unfavorable region. Coordinates have been deposited in the Protein Data Bank with the accession code 1L8A.

RESULTS

(A) Overall Description of the Structure. The crystallographic asymmetric unit contains one E1p dimer, and the structure determination gives, therefore, independent results for two subunits. The two independent E1p monomers are, nevertheless, almost identical stereochemically and are related by a noncrystallographic 2-fold symmetry axis. After least-squares alignment the rms deviation between corresponding Cα atoms in the dimer is 0.37 Å. The results described are therefore valid for both subunits. Accordingly, Ramachandran plots are also quite similar for both subunits, and 89% of the residues are in the most favored regions. Only one residue, Asp521, falls in an unfavorable region. This residue is part of a loop penetrating into the active site approaching the ThDP. Comparison of the two subunits shows that only seven residues (Glu70, Asn271, Glu283, Glu486, Glu675, Glu743, and Glu781) have different side chain orientations, presumably due to packing. Electron density over the entire molecule is generally very well defined; however, several surface residues have weak or missing side-chain densities. In addition, there are breaks in the polypeptide backbone (no interpretable electron density) for residues 1 to 55, 401 to 413, and 541 to 557. The average *B* value for main chain atoms is 16.0 Å^2 and for side chain atoms it is 18.6 Å^2. A ribbon drawing illustrating a single E1p subunit is shown in Figure 1.

(B) Structure of the E1p Subunit. Each PDHc E1p subunit consists of a single polypeptide chain of 886 amino acids, which folds into three distinct domains termed the N-terminal, middle, and C-terminal according to their consecutive locations along the chain. A topology diagram illustrating the individual domain folds is shown in Figure 2. All three domains have α/β folds, although they differ from each other. The N-terminal domain (residues 56 to 470) is larger than the other two domains and is centered on a five strand parallel β sheet. There are several connecting helices on both sides of the sheet and some additional helices on top of it, as well as several small helices that form the outermost ridge close to the N-terminus. The β/α connection after the third strand contains a loop involved in cofactor binding. Two of the three disordered regions (1−55 and 401−413) for which there is no interpretable electron density are in the N-terminal domain. The middle domain (residues 471−705) has six helices surrounding a six-strand parallel β-sheet and is connected to the N-terminal domain by an extended loop region (residues 458−493). The final disordered region (541−557) is in the middle domain and is located at the surface of the molecule. The C-terminal domain (residues 706−886) contains a mixed β-sheet with four strands parallel and one antiparallel with the sheet approximately perpendicular to those in the other two domains. The β-sheet is flanked by six helices with the last helix extending toward the C-terminus. The amino acid sequence along with the secondary structure is shown in Figure 3. Two subunits within a crystallographic asymmetric unit are tightly packed to form a functional dimer.

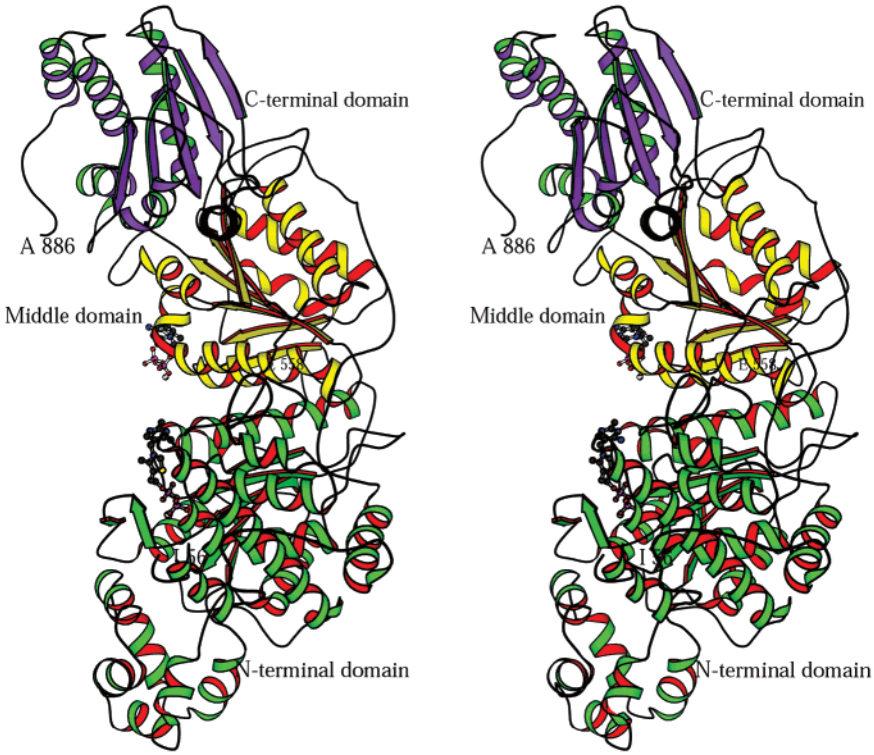

FIGURE 1: Stereo ribbon diagram for a single PDHc E1p subunit. ThDP cofactors are shown in a ball-and-stick representation. The figure was created with the program MOLSCRIPT (*16*).

(C) The E1p-Dimer. E. coli PDHc E1p functions as a dimer of MW 200 000 containing two identical polypeptide chains of 886 amino acids each and requires ThDP and Mg^{2+} as cofactors. The two subunits selected to form the crystallographic asymmetric unit are the two most tightly associated and form a functional dimer encompassing a pair of symmetrical active sites at the dimer interface. The dimer assignment is unambiguous as the meaningful dimer is that which forms the active sites at subunit-subunit interfaces, in a manner previously observed in all other ThDP dependent enzymes of known structure. A view of the dimer is shown in Figure 4. The dimer is roughly 103 × 95 Å wide in two dimensions and 71 Å thick along the central noncrystallographic 2-fold axis relating the two subunits. Within the dimer, subunit−subunit contacts involve residues from all three domains. There are tight packing interactions between helices of residues 237−249 and 273−285 on one side of the subunit in the N-terminal domain with 2-fold related helices in the other subunit. Subunit−subunit interactions involving the C-terminal domain mainly are contacts between the two C-terminal domains themselves. However, the C-terminal helix residues 831−841 also make contacts with the N-terminal domain residues 165−170. The dimer arrangement positions loops at the carboxyl end of the N-terminal domain sheet in one subunit in contact with loops at the carboxyl end of the middle domain sheet in the second subunit. Most of these interactions are in or close to the active sites.

(D) Cofactor Binding. In PDHc E1p, the ThDP cofactors are situated at the dimer interface with residues from both subunits contributing to cofactor binding. The diphosphate end of the cofactor binds only to N-terminal residues from one subunit, whereas the aminopyrimidine end binds to both the same N-terminal domain and to middle domain residues of the second subunit. The diphosphate group interacts with the protein in two ways, directly through hydrogen bonding and indirectly through coordination to a protein-bound Mg^{2+} ion. The octahedrally coordinated Mg^{2+} binding site contains three protein ligands: the side chains of the highly conserved ThDP-binding residues Asp230 and Asn260 and the main chain oxygen atom of Gln262. Additional binding to two oxygen atoms from the diphosphate and one water molecule completes the octahedral coordination. Other interactions of the diphosphate with protein involve hydrogen bonds to the residues Ser109, Gln140, and His142. The cofactor binding site is shown in Figure 5.

The aminopyrimidine and thiazolium rings in the cofactor are bound in a cleft between the two subunits. The cofactor is in the "V" conformation with the torsion angles (*18*) ϕ_T and ϕ_P being 108 and −68.6°, respectively. The thiazolium ring mainly interacts with residues Ile569 and Asp521, which pack against it. Stacking between the cofactor's aminopyrimidine ring and the side chain ring of Phe602 is evident. A number of hydrogen bonds also participate in binding the aminopyrimidine ring and are at least partly responsible for the proper orientation of this end of the cofactor in the active site. The 4′-amino group forms a conserved hydrogen bond to the main chain oxygen of Val192 in the first subunit and is also close to the His640 side chain in the second subunit. An important hydrogen bonding interaction (conserved in all ThDP dependent enzymes of known structure) is made by N1′ of the pyrimidine ring with the side chain of Glu571 (Figure 5). This interaction can induce tautomerization of the aminopyrimidine ring, converting the 4′-amino group to an imino group. The result of tautomerization is to place a relatively basic imino group in close proximity to and ideally oriented to abstract the C2 hydrogen, thus activating the cofactor (*19*). During turnover, the C2 atom of ThDP reacts

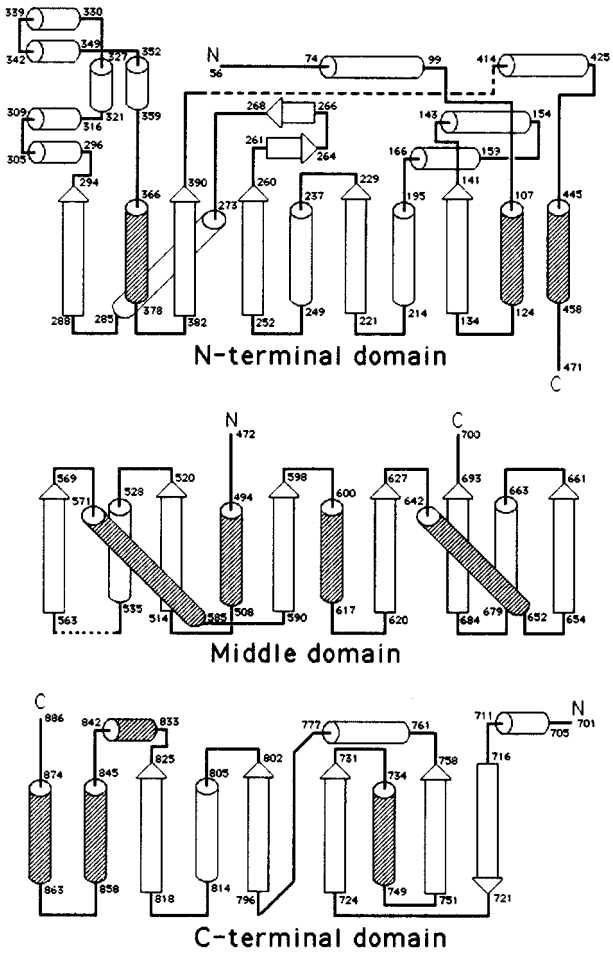

FIGURE 2: Topology diagrams for the N-terminal, middle, and C-terminal domains. Cylinders represent α-helices; arrows represent β strands. Open cylinders are behind the β sheets; shaded ones in front. Dotted lines indicate missing regions in the N-terminal and middle domains.

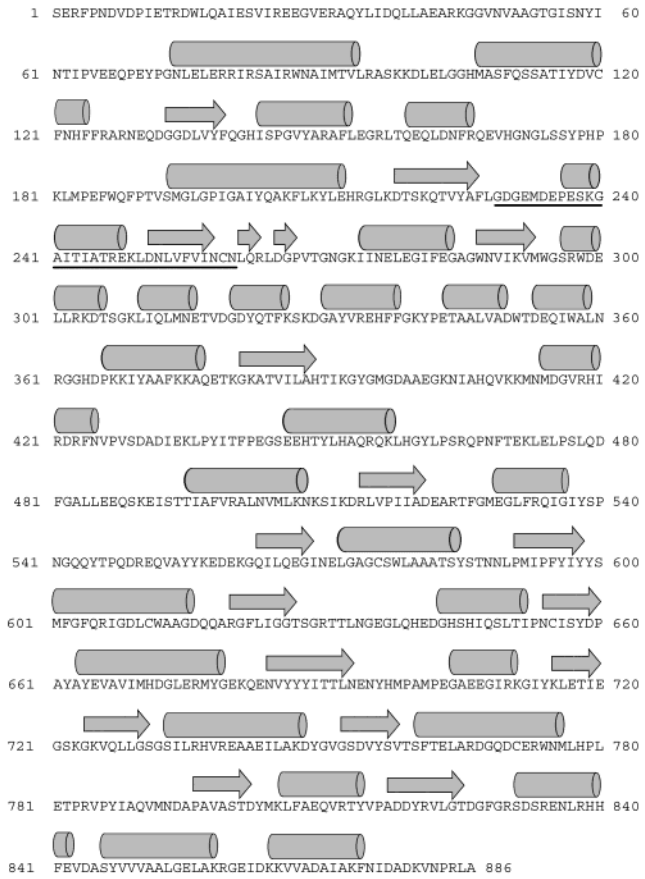

FIGURE 3: *E. coli* PDHc E1p sequence with secondary structure indicated. Helices are represented by cylinders and beta strands by arrows. The conserved ThDP fingerprint binding region is underlined.

with pyruvate to form 2-α-lactylThDP, which in turn undergoes decarboxylation to the 2-α-hydroxyethylidene-ThDP, an intermediate whose reactive enamine/2-α-carbanion moiety is common to all ThDP enzymes (see reaction 2 in introduction). The cofactor is almost completely buried, as only the C2 atom is accessible from solution through the active site cleft. Packing contacts with Met194, which is wedged between the cofactor rings may also be important in stabilizing the cofactor conformation as was shown for the corresponding Ile415 in PDC (*20*).

(E) Active Site Interactions. The active site cavity of E1p has dimensions of 18 × 8 × 21 Å^3 with a wide mouth and very deep cleft. The ThDP binding site is accessible from solvent only through the cleft between the two subunits. This cleft contains residues His106, Ser109, Gln140, His142, Tyr177, Met194, Asp230, Glu235, Asn260, Leu264, and Lys392 from one subunit and residues Asp521, Glu522, Ile569, Glu571, Tyr599, Phe602, Glu636, and His640 from the other subunit. The key residues involved in cofactor binding and likely to interact with substrate and/or reaction intermediates are shown in Figure 5. On one side of the thiazolium ring in the vicinity of the C2 atom is a cluster of histidine residues, His106, His142, and His640, that likely help attract and orient the negatively charged substrate pyruvate. The Glu522 side chain is positioned above, but close to the positively charged thiazolium ring and may help neutralize it as well as aid in orienting the substrate. Whereas His142 is involved in binding the cofactor's diphosphate group, His640, His106, and Glu522 might function in proton transfer during catalysis and in stabilization of the covalent intermediate formed between substrate and ThDP. It is possible that Tyr177 might also be able to interact with reaction intermediates (*21*). A water molecule O4 present in the active site is involved in hydrogen bonding with His640 as well as with the cofactor's 4′-amino group. This water molecule is close to the C2 carbon of the thiazolium ring and to protein residues His106 and Tyr177. It almost certainly will be displaced upon substrate binding. Asp521 points into the active site cleft and might also be indirectly involved in the reaction mechanism through interactions with Glu522. An electrostatic potential drawing of the active site is shown in Figure 6.

(F) Solvent Structure. A total of 682 ordered water molecules were located in the asymmetric unit of the E1p crystal. The average B-value of these water molecules is 18.1 Å^2. Every fixed water molecule is hydrogen bonded to a protein donor or acceptor atom, and the entire solvent structure is organized around protein polar groups. There are six water molecules that are tightly bound (average *B* value 9.5 Å^2) and make several hydrogen bonds to cofactor atoms (O13, O21, N4′, and Mg^{2+}) as well as to protein residues Ser109, Gln140, Asp258, Glu571, and His640. Water molecule O1 is a ligand for Mg^{2+} linking it to the protein,

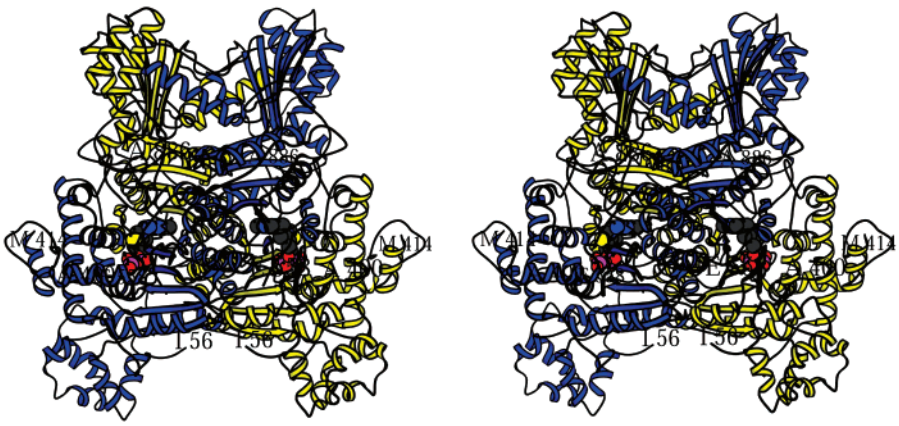

FIGURE 4: Stereo ribbon diagram for the E1p dimer generated by MOLSCRIPT (*16*). The ThDP cofactors are shown by a space-filling representation.

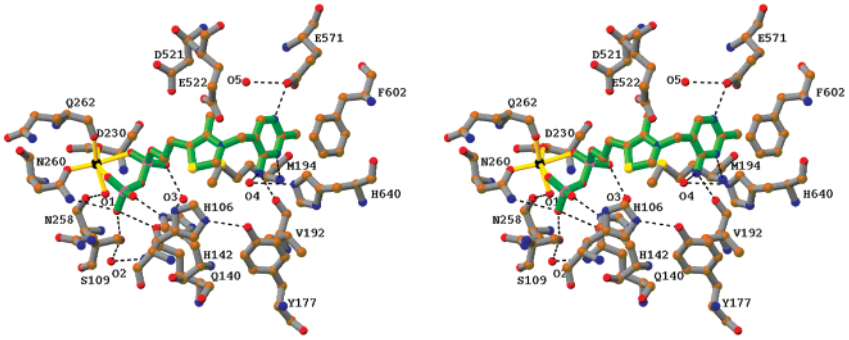

FIGURE 5: Stereo drawing of the cofactor binding and active site environment. The ThDP cofactor is situated at the interface between the N-terminal and Middle domains from different subunits. Residues with numbers less than 471 are from the N-terminal domain of one subunit whereas those numbered greater than 470 are from the middle domain of the "other" subunit. Several water molecules are involved in hydrogen bonding with the cofactor and are included in the drawing. The figure was created with the program RIBBONS (*17*).

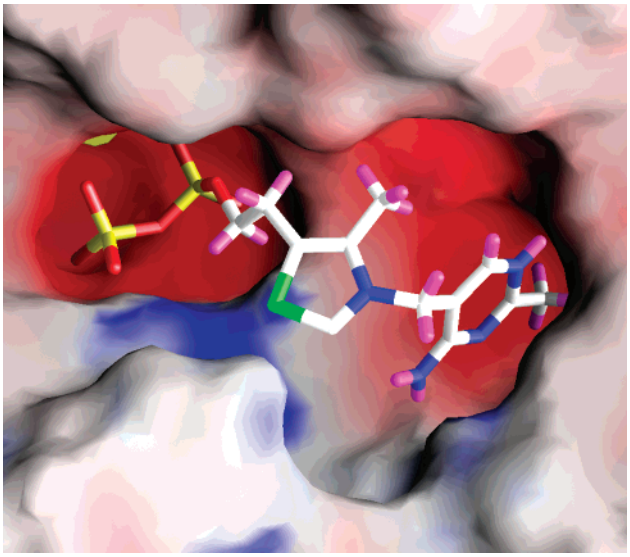

FIGURE 6: Electrostatic potential surface drawing showing the active site region. Note that some of the residues above the ThDP have been removed for visibility purposes (including E522), and the corresponding negative charge at the top of the drawing is therefore not visible. The figure was created with the program GRASP (*22*).

and water molecule O4 is hydrogen bonded to the 4′-amino nitrogen atom of the cofactor as well as to His640.

(G) Comparison of Homodimer E1p with Heterotetramer E1b. While an extensive comparison between homodimeric and heterotetrameric E1's is beyond the scope of this text, some large-scale differences are nevertheless apparent, and a simple comparison is warranted. When individual domains of *E. coli* PDHc E1p and branched chain *P. putida* E1b are compared, they appear generally similar but with substantial differences in certain regions. The N-terminal domain of E1p corresponds to the α-subunit of the E1b-heterotetramer while the E1p Middle domain corresponds to the E1b *β*-subunit N-terminal domain. However, the E1p N-terminal residues 56−80 are oriented differently than the corresponding E1b α-subunit residues. Also, several small helices (residues 295−365) found in the outermost ridge on the E1p N-terminal domain are not present in E1b, indicating that they may have a function that is absent in E1 heterotetramers. The E1p C-terminal domain corresponds to the E1b *β*-subunit C-terminal domain. Rms deviations between the corresponding three domains of E1p and E1b are 1.67, 2.00, and 1.87 Å, respectively, for the Cα atoms of matched residues as calculated with the program O (*13*); however, only roughly 50% of the Cα atoms match in each domain. When the entire molecules are considered as a whole, again only about half (404) of the residues appear to be structurally equivalent (rms deviation 2.12 Å), indicating structural differences between the two enzymes. There are also large differences in the nature and lengths of loop regions connecting strands in the *β*-sheets within domains. Because of the large structural differences in some areas, it is not possible to identify corresponding residues, and hence the overall rms deviation of 2.12 Å is misleadingly low since these regions were not included in the analysis. A superposition of the

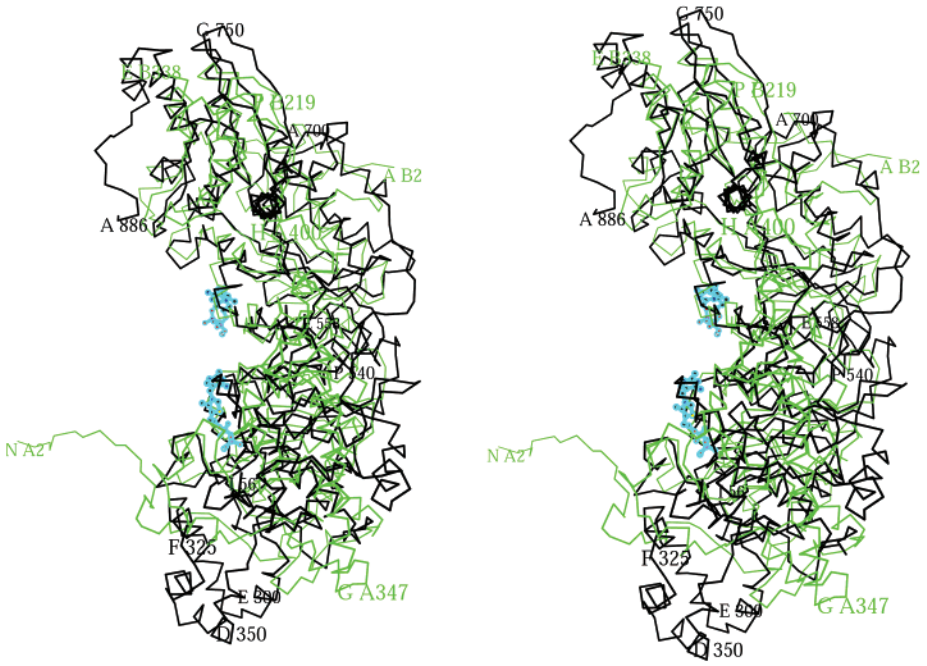

FIGURE 7: Stereo drawing of superposition of the *E. coli* PDHc E1p subunit and its *P. putida* BDHc E1b counterparts after least squares alignment. Colors are black and green for the E1p and E1b structures, respectively. The figure was created with the program MOLSCRIPT (*16*).

E1p and E1b structures after least squares alignment is shown in Figure 7.

DISCUSSION

ThDP serves as a cofactor in many biochemical reactions involving cleavage of a carbon−carbon bond adjacent to a keto-group. Crystal structure comparisons of three ThDP-dependent enzymes, yeast pyruvate decarboxylase (PDC), yeast transketolase (TK), and *Lactobacillus plantarum* pyruvate oxidase (POX) revealed that ThDP is always bound in a cleft formed by hydrophobic amino acids at the interface between two subunits within a dimer, and there is a common ThDP binding fold (*23*). The E1p structure is consistent with this dimeric binding mode, as are all ThDP-dependent enzymes of known structure. The ThDP binding fold involves a sequence approximately 30 amino acid residues long. For *E. coli* PDHc E1p, the sequence commences with the highly conserved G(229)DG and terminates with N(258)CN. The GDG..X_{26}..N(C)N fingerprint binding motif (*24*) is conserved in nearly all of the hundreds of sequences of ThDP dependent enzymes in sequence data banks. Site-directed mutagenesis studies on PDHc E1p from *E. coli* have since confirmed the importance of the ThDP fold in maintaining enzyme activity (*25*). In the related enzyme PDC (*19*, *26*), the ThDP binding sequence is located near the C terminal end, but in PDHc E1p it is located near the N terminal end, indicative of a different binding domain assembly. This different domain ordering relative to PDC has also been reported in transketolase (*23*) and BDHc E1b (*1*).

In the active site the cofactor conformation, although similar to that found in other ThDP dependent enzymes, differs significantly. In particular, the value of ϕ_T is roughly 12° larger than that found in other ThDP dependent enzymes, including its cousin, PDC (*19*, *26*) and the related BDHc E1b (*1*). This represents a small but significant shift away from the standard "V" conformation and effectively reduces

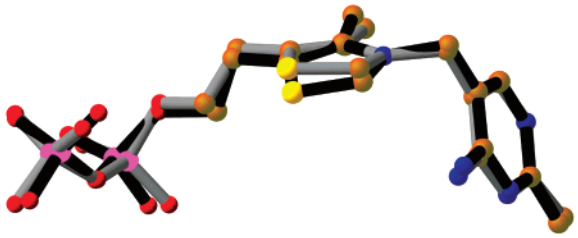

FIGURE 8: Superposition of the ThDP cofactors in *E. coli* PDHc E1p and *P. putida* BDHc E1b after least squares alignment based on the cofactors alone. The E1p and E1b cofactors are shown with gray and black bonds, respectively. The figure was created with the program RIBBONS (*17*).

steric clashes between the thiazolium C2 hydrogen (or reaction intermediates covalently bound to C2) and an N4′ bound hydrogen from the aminopyrimidine ring. This may be an indication of flexibility in cofactor conformation previously unobserved in enzyme bound ThDP. In this regard, it is worth noting that while the diphosphate and aminopyrimidine ends of the cofactor have positions and orientations tightly constrained by directed hydrogen bonds and/or ring stacking interactions, the thiazolium ring forms no such interactions and thus may be expected to have greater mobility. A superposition of the E1p and E1b ThDP cofactors after least squares alignment is shown in Figure 8.

Within the *E. coli* PDHc multienzyme complex the ThDP dependent E1p component catalyzes the decarboxylation of pyruvate and subsequently the reductive acetylation of a lipoamide group bound to the E2 component. The lipoyl domain of E2 has to approach E1p closely for the attached lipoamide group to penetrate the active site channel and reach the acyl-ThDP intermediate. Recognition of the lipoyl domain by the E1p component within the complex thus makes for an elegant, effective, and necessary mechanism of substrate channeling (*27*). The surface regions potentially involved in molecular interactions with the lipoyl domain

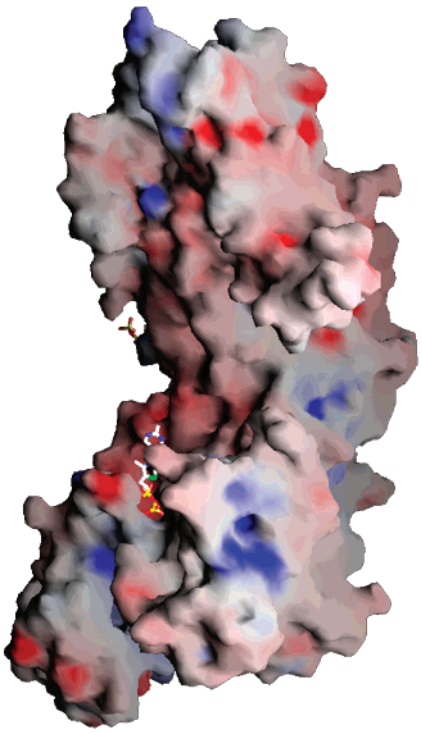

FIGURE 9: Electrostatic potential surface drawing for a single subunit. The orientation is similar to that in Figure 1. The figure was created with the program GRASP (*22*).

are found in the E1p N-terminal and middle domains; mainly residues His106, Ser109, Tyr177, Glu521, His640, and Phe602, forming the rim of the channel opening. The crystal structure of E1p from *E. coli* shows that the catalytic ThDP is buried at the bottom of a 21 Å deep funnel-shaped hole at the interface between subunits. For the lipoamide group to reach the active site, assuming that no major conformational changes occur, the lipoyl domain has to come into close contact with the E1p surface, and the lipoyl-lysine side-chain must become fully extended. In the active site, the water molecule O4 (Figure 5) is assumed to mimic the position of cofactor bound substrate. This water molecule is very close to residues His106, Tyr177, and Glu522 and hydrogen bonded to residue His640. These residues therefore are likely to play important roles in substrate binding with C2 of the ThDP thiazolium ring, and may help to stabilize the thioester bond linking the acyl group to the E2-bound dihydrolipoamide swinging arm as part of the catalytic mechanism.

In multienzyme complexes, the E2 binding domain is in many cases responsible for binding E1 and/or E3 to the E2 core. However, in the *A. vinelandii* PDHc complex site-directed mutagenesis of E2 has shown that E1p interacts with *both* the E2 binding and catalytic core domains. Limited proteolysis experiments of the *A. vinelandii* E1p component have further suggested that the N-terminal region of E1p is required for binding to the E2 core domain as the enzyme produced after deletion of 48 amino acids at the N-terminus could no longer interact with E2 (*28*). The specific binding interactions are thought to be mainly of an electrostatic nature involving negatively charged residues on the N-terminal domains of E1p and positively charged residues on the binding and catalytic domains of E2. Removal of the N-terminal residues has no influence on the interactions of E1p with free lipoyl domains or on catalysis in the acyl-transfer. The high sequence identity (60%) between the *E. coli* and *A. vinelandii* E1p's suggests a similar binding and catalytic mechanism in the *E. coli* multienzyme complex. A comparison of the first 80 N-terminal amino acids from these enzymes indicates 38% sequence identity and 58% sequence similarity, supporting the notion of a similar role in the *E. coli* complex. Surprisingly, sequence comparisons of E1 homodimers with E1 heterotetramers indicate no sequence similarity between these components in the first 82 residues of the corresponding N-terminal domains. This suggests that the binding mode and function of this region may be different in homodimeric and heterotetrameric E1 enzymes. Unfortunately, because of disorder the first 55 N-terminal residues are not observed in the E1p crystal structure, and no specific details are available although the general location of this region relative to the overall structure is apparent. It is possible that this region will be ordered in a different crystal form, or in complexes involving E2 components or fragments thereof. An electrostatic potential drawing is shown in Figure 9; however, due to the absence of the first 55 N-terminal residues in the crystal structure, the figure is incomplete.

The PDHc E1p structure from *E. coli* represents the first structural example for any intact and functional pyruvate dehydrogenase multienzyme complex E1 component. It is also the first structural example for any member of the large family of homodimeric E1's, and because of the high degree of sequence similarity, is well suited to serve as an initial model for all members of this class. It clearly shows that the conformations of some regions differ significantly from that of E1 heterotetramers, and these structural differences may help explain binding differences during assembly within multienzyme complexes. Additional biochemical and crystallographic studies of E1p in complex with E2/E3 subcomplex are indicated to further understanding of binding, assembly, and catalytic functions within multienzyme complexes.

ACKNOWLEDGMENT

The authors are grateful to S. Swaminathan, D. Kumaran, and S. Eswaramoorthy for their assistance and guidance during data collection at the NSLS in Brookhaven National Laboratory.

REFERENCES

1. Aevarsson, A., Seger, K., Turley, S., Sokatch, J., and Hol, W. (1999) *Nature Struct. Biol. 6*, 785−792.
2. Aevarsson, A., Chuang, J. L., Wynn, R. M., Turley, S., Chuang, D. T., and Hol, W. (2000) *Structure 8*, No 3, 277−291.
3. Reed, L. J. (1974) *Acc. Chem. Res. 7*, 40−46.
4. Stephens, P. E., Darlison, M. G., Lewis, H. M., and Guest, J. R. (1983) *Eur. J. Biochem. 133*, 155−162.
5. Stephens, P. E., Darlison, M. G., Lewis, H. M., and Guest, J. R. (1983) *Eur. J. Biochem. 133*, 481−489.
6. Stephens, P. E., Darlison, M. G., Lewis, H. M., and Guest, J. R. (1983) *Eur. J. Biochem. 135*, 519−527.
7. Kleiger, G., Perry, J., and Eisenberg, D. (2001) *Biochemistry 40*, 14484−14492.
8. Nemeria, N., Volkov, A., Brown, A., Yi, J., Zipper, L., Guest, J., and Jordan, F. (1998) *Biochemistry 37*, 911−922.
9. McNally, A. J., Motter, K., and Jordan, F. (1995) *J. Biol. Chem. 270*, 19736−19743.
10. Otwinowski, Z., and Minor, W. (1997) *Methods Enz. 276*, 307−326.

11. Weeks, C. M., and Miller, R. (1999) *Acta Crystallogr. D 55*, 492−500.
12. Furey, W., and Swaminathan, S. (1997) *Methods Enz. 276*, 590−620.
13. Jones, T. A., Zou, J., Cowtan, S., and Kjeldgaard, M. (1991) *Acta Crystallogr. A 47*, 110−119.
14. Brünger, A. T., Krukowski, A., and Erickson. J. W. (1990) *Acta Crystallogr. A 46*, 585−593.
15. Leskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
16. Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946−950.
17. Carson, M. (1991) *J. Appl. Crystallog. 24*, 958−961.
18. Pletcher, J., Sax, M., Blank, G., and Wood, M. (1977) *J. Am. Chem. Soc. 99*, 1396−1403.
19. Arjunan, P., Umland, T., Dyda, F., Swaminathan, S., Furey, W., Sax, M., Farrenkopf, B., Gao, Y., Zang, D., and Jordan, F. (1996) *J. Mol. Biol. 256*, 590−600.
20. Guo, F., Zhang, D., Kahyaoglou, A. Farid, R. S., and Jordan, F. (1998) *Biochemistry 37*, 13379−13391.
21. Nemeria, N., Yan, Y., Zhang, Z., Brown, A. M., Arjunan, P., Furey, W., Guest, J. R., and Jordan, F. (2001) *J. Biol. Chem. 276*, 45969−45978.
22. Nicholls, A., Sharp, K., and Honig, B. (1991) *Proteins* 11, 281−296.
23. Muller, Y. A., Lindqvist, Y., Furey, W., Schulz, G. E., Jordan, F., and Schneider, G. (1993) *Structure 1*, 95−103.
24. Hawkins, C. F., Borges, A., and Perham, R. N. (1989) *FEBS Lett. 255*, 77−82.
25. Yi, J., Nemeria, N., McNally, A., Jordan, F., Machado, R., and Guest, J. R. (1996) *J. Biol. Chem. 271*, 33192−33200.
26. Dyda, F., Furey, W., Swaminathan, S., Sax, M., Farrenkopf, B., and Jordan, F. (1993) *Biochemistry 32*, 6165−6170.
27. Jones, D. D., Stott, K. M., Reche, P. A., and Perham, R. N. (2001) *J. Mol. Biol. 305*, 49−60.
28. Hengeveld, A. F., Schoustra, S. E., Westphal, A. H., and De Kok, A. (1999) *Eur. J. Biochem. 265*, 1098−1107.

BI0118557