

Decrypting the Biochemical Function of an Essential Gene from *Streptococcus pneumoniae* Using ThermoFluor® Technology*

Received for publication, November 24, 2004, and in revised form, December 22, 2004
Published, JBC Papers in Press, January 5, 2005, DOI 10.1074/jbc.M413278200

Theodore E. Carver,^{a,b,c} Brian Bordeau,^{a,b} Maxwell D. Cummings,^{a,b,d} Eugene C. Petrella,^{a,e}
Michael J. Pucci,^{f,g} Laura E. Zawadzke,^{f,h} Brian A. Dougherty,^f Jeffrey A. Tredup,ⁱ
James W. Bryson,ⁱ Joseph Yanchunas, Jr.,ⁱ Michael L. Doyle,ⁱ Mark R. Witmer,ⁱ Marina I. Nelen,^a
Renee L. DesJarlais,^a Edward P. Jaeger,^a Heather Devine,^a Eric D. Asel,^a Barry A. Springer,^a
Roger Bone,^a F. Raymond Salemme,^{a,j} and Matthew J. Todd^{a,k}

From the ^aJohnson & Johnson Pharmaceutical Research & Development, L.L.C., Exton, Pennsylvania 19341, ⁱBristol-Myers Squibb PRI, Princeton, New Jersey 08543, and ^fBristol-Myers Squibb PRI, Wallingford, Connecticut 06492

The protein product of an essential gene of unknown function from *Streptococcus pneumoniae* was expressed and purified for screening in the ThermoFluor® affinity screening assay. This assay can detect ligand binding to proteins of unknown function. The recombinant protein was found to be in a dimeric, native-like folded state and to unfold cooperatively. ThermoFluor was used to screen the protein against a library of 3000 compounds that were specifically selected to provide information about possible biological functions. The results of this screen identified pyridoxal phosphate and pyridoxamine phosphate as equilibrium binding ligands ($K_d \sim 50 \mu\text{M}$, $K_d \sim 2.5 \mu\text{M}$, respectively), consistent with an enzymatic cofactor function. Several nucleotides and nucleotide sugars were also identified as ligands of this protein. Sequence comparison with two enzymes of known structure but relatively low overall sequence homology established that several key residues directly involved in pyridoxal phosphate binding were strictly conserved. Screening a collection of generic drugs and natural products identified the anti-fungal compound canescin A as an irreversible covalent modifier of the enzyme. Our investigation of this protein indicates that its probable biological role is that of a nucleoside diphospho-keto-sugar aminotransferase, although the preferred keto-sugar substrate remains unknown. These experiments demonstrate the utility of a generic affinity-based ligand binding technology in decrypting possible biological functions of a protein, an approach that is both independent of and complementary to existing genomic and proteomic technologies.

It is estimated that 40–60% of prokaryotic and eukaryotic genes have unknown or tentatively assigned biological functions (1). The protein products encoded by these genes are potentially valuable targets for therapeutic intervention. For example, understanding the functions of previously uncharacterized bacterial proteins could lead to the development of new classes of antibiotics (2). Such antibiotics are urgently needed for treating the growing number of bacterial strains resistant to currently available therapies. The advent of new tools for proteomics and bioinformatics has facilitated the identification of protein function, achieved through classification of functional domains, patterns of expression, and binding partners. These tools are essential for defining biological function in a cellular context.

The majority of known microbial drug targets contain sites of interaction with low-molecular weight ligands, for example, cofactors and metabolites. Discovering and understanding these “drug-able” sites on new proteins is of vital importance (3). Sequence data alone is often insufficient to describe the molecular functions of proteins and their sites of interaction with small molecules. Even when tentative assignment of molecular function is possible, biochemical characterization of the expressed protein product is essential to confirm and elaborate function. High throughput biochemical methods have been developed to address this need and take full advantage of the vast quantities of biochemical data in commercial data bases. For example, chemically reactive probes have been used to study multiple members of a family of enzymes (4). Affinity-based screening of ligands immobilized in microarrays has been employed to identify small molecule binding partners of proteins (5). Other affinity-based technologies such as surface plasmon resonance and capillary electrophoresis have also been used to study small molecule binding when protein function is unknown (6).

The utility of many methods is limited by their requirement for covalent modification of the target protein and/or the use of libraries of specialized molecules that may not be representative of the vast diversity of biochemical space. For screening using any available sample of soluble expressed protein and any commonly accessible chemical library, we have developed an affinity-based screening technology, ThermoFluor® (7). ThermoFluor measures the enhanced thermal stability conferred by the binding of ligands to the native state of proteins and is readily adaptable to high-throughput screening in 384-well plates. No knowledge of substrates or binding partners is required to detect binding in ThermoFluor, a feature that renders it ideal for characterizing the binding of biological ligands

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

^b These authors contributed equally to this work.

^c Current address: Ceptor Corp., 200 International Circle, Suite 5100, Hunt Valley, MD 21030-1350.

^d To whom correspondence may be addressed: Johnson & Johnson Pharmaceutical Research & Development, L.L.C., 665 Stockton Dr., Exton, PA 19341. E-mail: Mccumming@prduus.jnj.com.

^e Current address: Daiamed Pharmaceuticals, 1 Kendall Square, Bldg. 700, Cambridge, MA 02139.

^f Current address: Achillion Pharmaceuticals, 300 George St., New Haven, CT 06511.

^h Current address: Pfizer, Inc., 1 Eastern Point Rd., Groton, CT 06340.

^j Current address: Linguagen Corp., 2005 Eastpark Blvd., Cranbury, NJ 08512.

^k To whom correspondence may be addressed: Johnson & Johnson Pharmaceutical Research & Development, L.L.C., 665 Stockton Dr., Exton, PA 19341. E-mail: Mtodd3@prduus.jnj.com.

to proteins of unknown function. In this work, we have screened CFE97¹ from *Streptococcus pneumoniae* (8), a protein of unknown function, against a functional probe library comprising commercially available biomolecules.

In an earlier study, 113 genes were shown to be essential for growth of *S. pneumoniae* using targeted gene disruption (8). These genes were selected to have a significant level of homology to related genes from at least two of four other bacterial species (40% global amino acid sequence identity to genes found in *Bacillus subtilis*, *Enterococcus faecalis*, *Escherichia coli*, or *Staphylococcus epidermis*) (8). A number of these genes have unknown or poorly characterized functions. One of these orphan proteins, CFE97, the protein product of gene *sp1837* from *S. pneumoniae*, was cloned, expressed, purified, and screened using ThermoFluor. Information obtained from the screening hits enabled further biochemical characterization of the target, as well as the design of biochemical assays tailored to the putative function of CFE97. The results of these experiments, in combination with comparative analyses of the protein sequence, provide a clearer picture of the possible biochemical functions of CFE97.

EXPERIMENTAL PROCEDURES

Materials—All materials were of the highest available quality and were purchased from Sigma unless otherwise specified. Canescin A was obtained from Microsource (Gaylordsville, CT) and Dpx was obtained from Molecular Probes (Eugene, OR).

Cloning, Expression, and Purification of Protein—The *sp1837* gene was inserted into a pET vector with a His₆ tag as described previously (8). The plasmid was transformed into the BL21(DE3) pLys-S expression strain of *E. coli* (Novagen) and 2 liters of culture were grown. The cell paste was lysed using the Rannie homogenizer and the supernatant from ultracentrifugation of the lysate was filtered through a 1.2- μ m filter. The filtrate was loaded onto an 80-ml Ni-MCC (Pharmacia Chelating Sepharose Fast Flow) column, and the protein was eluted with a 25–500 mM imidazole gradient. Fractions were pooled based upon protein amount and purity in SDS-PAGE. The final purification step was an Amersham Biosciences Superdex-200 (26/60) size exclusion column with a running buffer of 20 mM HEPES, pH 7.5, 150 mM NaCl, and 1 mM dithiothreitol. The final protein was >95% pure as judged by SDS-PAGE using a 4–20% Tris glycine gel.

Description of the ThermoFluor Assay—The ThermoFluor assay was performed as described previously (7) using instrumentation built at 3-Dimensional Pharmaceuticals.² The protein (2 μ l at 2 μ M in 50 mM HEPES, 100 mM NaCl, 100 μ M Dpx) was dispensed into black Abgene polypropylene ThermoFast-384 plates (Epsom, UK) containing an equal volume of test compound in either water or 4% Me₂SO. Wells contained 4 μ l (protein + compound) and were overlaid with 1 μ l of polydimethylsiloxane DC200 oil (Sigma) to prevent evaporation. For each plate, the temperature was increased from 30 to 75 °C in 1 °C increments. At each temperature the plates were equilibrated for 3 min and then rapidly cooled to 25 °C, after which four images were recorded using a 12-bit, cooled CCD camera (10 s exposure time). An average intensity per well was calculated by integrating pixel intensities per well and averaging the four exposures. Approximately 3000 compounds comprising the functional probe library (contents described below) were screened against CFE97, at a concentration between 10 μ M and 1 mM (compound dependent).

Analysis of Ligand Binding Using ThermoFluor—The analysis of protein unfolding curves and the derivation of equilibrium binding constants has been described in detail elsewhere (7, 9). Briefly, the temperature dependence of the increase in fluorescence because of protein denaturation, $y(T)$, can be described by the following equation,

$$y_T = y_f + \frac{y_u - y_f}{1 + e^{\frac{-\Delta_U G(T)}{RT}}} = y_u + \frac{y_f - y_u}{1 + e^{\frac{-\Delta_U G(T)}{RT}}} \quad (\text{Eq. 1})$$

where y_f is fluorescence intensity when protein is folded (native), y_u is fluorescence intensity when protein is non-native, and the exponential represents the probability of a protein with unfolding free energy $\Delta_U G(T)$, being folded. The midpoint of the unfolding is defined as the T_m . The equilibrium constant for unfolding (K_{eq}) is temperature-dependent, as given by the following expression (10, 11),

$$K_{eq} = e^{\frac{-\Delta_U G_T}{RT}} = e^{\frac{-(\Delta_U H_T + \Delta_U C_p(T - T_r) - T(\Delta_U S_T + \Delta_U C_p \ln \frac{T}{T_r}))}{RT}} \quad (\text{Eq. 2})$$

where $\Delta_U G(T_r)$, $\Delta_U H_T$, and $\Delta_U S_T$ are the Gibbs free energy, enthalpy, and entropy of protein unfolding, respectively, at a reference temperature taken to be T_m . The thermal stability at any temperature, T , can be calculated from the expected change of each thermodynamic parameter relative to a reference temperature, T_r , using standard equations (Equation 2) that include the heat capacity of protein unfolding, $\Delta_U C_p$, which is assumed to be temperature independent.

The equilibrium constant for ligand binding (K_b) at any temperature can be calculated from its value at a reference temperature by a similar function,

$$K_b = e^{\frac{-\Delta_b G_T}{RT}} = e^{\frac{-(\Delta_b H_T - T\Delta_b S_T)}{RT}} = e^{\frac{-(\Delta_b H_T + \Delta_b C_p(T - T_0) - T(\Delta_b S_{T_0} + \Delta_b C_p \ln \frac{T}{T_0}))}{RT}} \quad (\text{Eq. 3})$$

where $\Delta_b G_T$, $\Delta_b H_T$, $\Delta_b S_T$, and $\Delta_b C_p$ are, respectively, Gibbs free energy, enthalpy, entropy, and heat capacity of ligand binding at any temperature, T , relative to that at a reference temperature, T_0 (again, the heat capacity is assumed to be temperature independent). For protein unfolding, the reference temperature was set to the T_m in the absence of added ligand; for ligand binding, the reference temperature was 37 °C. All reported binding affinities were measured at T_m and then extrapolated to 37 °C, assuming $\Delta_b H_{T_0=37^\circ\text{C}} \sim -5$ kcal mol⁻¹ and $\Delta_b C_p \sim -250$ cal mol⁻¹ K⁻¹. The accuracy of T_m as determined across 384 identical melts was ± 0.2 °C; ligand induced changes in thermal stability greater than 0.4 °C were considered significant.

The above expressions for K_{eq} and K_b can be combined to calculate the effect of ligand concentration on the observed protein T_m using the following equation,

$$L_t = (1 - K_{eq}) \left(\frac{P_t}{2} + \frac{1}{K_{eq} K_b} \right) \quad (\text{Eq. 4})$$

where L_t is the total ligand concentration (free + bound), and P_t is the total protein concentration (free + bound) (9).

Ligand binding affinity, K_b , was determined by measuring the effect of varying ligand concentrations on protein T_m (9). Compound was serially diluted in either aqueous buffer or Me₂SO. These dilutions were mixed with protein and dye in the assay buffer, and the T_m was measured at each compound concentration, as described above. In cases where ligand binding was found to be reversible, the plot of T_m versus ligand concentration was simulated using Equation 4 by fitting unfolding curves to Equation 1 to obtain an observed T_m and comparing these values to curves simulated by inserting different values of K_b into Equation 4.

Composition of the Functional Probe Library—The functional probe library comprised a proprietary set of molecules selected from four general classes of compounds: known bioactive molecules (substrates, cofactors, and receptor ligands), generic drugs, natural products, and compounds that are inhibitors or antagonists of biological reactions and interactions. Portions of this library were purchased in plates from vendors; the remainder consisted of compounds that were purchased individually and assembled into plates based upon annotations in data bases of biological ligands (e.g. BRENDA (12)). The assembled library was dispensed into 384-well plates as concentrated stock solutions in Me₂SO or water, and the plates were stored at -80 °C.

Absorbance Measurements and Assays—A Hewlett Packard model 8453 diode-array spectrophotometer (Agilent Technologies, Inc.) was used to record absorbance spectra using 100- μ l quartz cuvettes (1-cm path length). Protein was dialyzed exhaustively to remove excess cofactor prior to the analyses shown in Fig. 6A. Initial velocities for the half-reaction in which pyridoxal phosphate (PLP) is converted into pyridoxamine phosphate (PMP) were calculated from the increase in absorbance at 330 nm versus time using a Tecan Safire plate reader at

¹ The abbreviations used are: CFE97, cloned for expression protein 97; Dpx, dapoxy sulfoxide; Me₂SO, dimethyl sulfoxide; CCD, charge coupled device; AHBA, 3-amino-5-hydroxybenzoic acid; PLP, pyridoxal phosphate; PMP, pyridoxamine phosphate; PIPES, 1,4-piperazinediethanesulfonic acid.

² The ThermoFluor assay was developed by 3-Dimensional Pharmaceuticals, Inc., which was merged into Johnson & Johnson Pharmaceutical Research & Development, L.L.C. "ThermoFluor" is a registered trademark in the United States and certain other countries.

37 °C. Reactions (25 mM PIPES, 100 mM NaCl, 2.5 mM MgCl₂, 250 μM PLP plus 0–10 mM amino acid) were initiated by adding enzyme (to 5 μM) and the appearance of PMP was monitored for 90 min. Initial rates were calculated during the linear portion of the assay, and kinetic constants were determined using double-reciprocal plots.

Biophysical Characterization of CFE97—Purified CFE97 (20 mM HEPES, 150 mM NaCl, pH 7.5) was characterized by several biophysical techniques. Helicity of purified protein was estimated by circular dichroism (CD model 202, Aviv Biomedical, Lakewood, NJ), using protein desalted into 25 mM PIPES, 100 mM NaCl, 2.5 mM MgCl₂, pH 7.0. The CD spectrum indicated that CFE97 contained a considerable amount of α -helix as judged by minima near 208 and 222 nm. The temperature dependence of the CD signal showed a marked decrease in ellipticity on heating, with a T_m near 50 °C. Sedimentation equilibrium data were measured on a Beckman XL-I analytical ultracentrifuge at 20 °C. Samples were centrifuged until equilibrium was attained, as judged by an unchanging absorbance *versus* radial position profile. The data were analyzed with either a single homogeneous species model or a monomer-dimer model as described previously (13). Buffer density was 1.00499 g ml⁻¹ and the partial specific volume was 0.7392 ml g⁻¹. The enthalpy of protein unfolding was measured using differential scanning calorimetry (VP-DSC, MicroCal, L.L.C., Northampton, MA). Protein was desalted into 25 mM PIPES, 100 mM NaCl, 2.5 mM MgCl₂, pH 7.0, using a PD-10 desalting column (Amersham Biosciences) equilibrated with the same buffer. Degassed protein at 0.4 mg/ml was loaded into a thoroughly washed, thermally conditioned calorimeter cell. Differential excess enthalpy in the protein cell *versus* an exact buffer match in the reference cell was monitored as the temperature was ramped (10 to 90 °C, at 1 °C/min).

Mass Spectrometry—Protein samples were incubated alone or with a 10-fold molar excess of PLP or canescin A. Following incubation, samples were dialyzed in PIPES buffer to remove excess compound, desalted into 70% acetonitrile/water (v/v), and injected onto a Micromass Q-Tof® mass spectrometer. Mass was determined using electrospray ionization in positive ion mode. Data were analyzed using the formula $(m/z) \cdot z - z = \text{mass}$, where m/z is a mass to charge ratio given by the instrument for a peak and z is an estimated charge. Given the correct charge estimate, all peaks will arrive at the same mass \pm 0.01%.

RESULTS AND DISCUSSION

Analysis of the Oligomeric State of CFE97—For orphan proteins, by definition, there is no function-based standard by which to evaluate the state of the protein in solution. It is therefore important to determine whether the expressed protein is in a homogeneous state consistent with native, folded protein. Equilibria between intermediates on a protein unfolding pathway can complicate protein melting curves, and the interpretation of ligand effects on stability. Thus it is standard practice to undertake a thorough biophysical characterization of the target protein prior to any ThermoFluor screening. In the case of an unknown protein the biophysical data may also be useful in decryption of function.

Although purified CFE97 showed significant α -helicity by circular dichroism and no apparent light scattering, a discrepancy in the native molecular weight was seen when comparing results obtained with different measurements. Analytical size exclusion chromatography of the purified CFE97 protein yielded a homogeneous elution profile indicating an apparent mass of 41 kDa when compared with a series of protein standards, consistent with a monomeric form of the protein (expected molecular mass is 46.6 kDa, based on sequence). However, dynamic light scattering analysis yielded an apparent mass of 93 kDa, suggesting CFE97 was present as a dimer. To resolve this discrepancy, we conducted analytical ultracentrifugation experiments.

Fig. 1 shows sedimentation equilibrium data for CFE97 conducted at several rotor speeds. Global analysis of the data to a single species model gave an average mass of 91 kDa, which clearly demonstrated that CFE97 is predominantly a dimer at micromolar concentrations. However, a slightly better global fit was obtained to the multiple rotor speeds data with a monomer-dimer model, yielding a dimerization equilibrium constant

of 300 nM. Sedimentation velocity experiments were also conducted with CFE97 at 1.0, 0.3, and 0.1 mg/ml. At each of these concentrations the protein sedimented as a single species with a sedimentation coefficient of \sim 4.5 S and no detectable aggregate (data not shown).

The thermal stability of CFE97 was measured using differential scanning calorimetry in the presence and absence of environmentally sensitive dyes used in ThermoFluor experiments (Fig. 2). In the absence of hydrophobic dye, the thermogram shows a single cooperative unfolding unit. Thermal stability decreased \sim 1 °C in the presence of dye, consistent with dye interacting specifically with non-native protein. DSC curves could be simulated by a monomeric equilibrium unfolding model; however, a better fit to the data was obtained assuming a monomer-dimer equilibrium unfolding model (14), where folded dimer is in equilibrium with unfolded monomer. These latter fits are shown in Fig. 2, with fit thermodynamic parameters listed in Table I. Thus, the results of several distinct analyses indicate that CFE97 exists as a dimer in solution. A $\Delta_C H_{T_m}$ of 80 kcal mol⁻¹ and a $\Delta_C C_p$ of 3000 kcal mol⁻¹ K⁻¹ were used in Equations 2 and 4 to interpret the concentration effect of ligands on T_m (below).

Functional Probe Library Screen—ThermoFluor conditions were optimized for CFE97 unfolding measurements and the effect of ligands in the functional probe library on protein thermal stability was evaluated. Of \sim 3000 compounds tested, 55 reproducibly increased the midpoint transition temperature by an amount \geq 1 °C; a subset of these is shown in Table II.

A potent and informative hit discovered in the screening panel was PLP, which increased the T_m of CFE97 by 25 °C. PLP is a cofactor for several classes of enzymes including aminotransferases, decarboxylases, and dehydratases, and the genes for these enzymes constitute as much as 1.5% of bacterial genomes (15). Interestingly, these enzymes typically exist as homodimers or some other dihedrally symmetric polymeric form (16). The PLP binding site comprises elements from both halves of the dimer (16). Our biophysical characterization of CFE97 as a dimer is thus consistent with the biochemical relevance of the observed PLP binding. PLP forms a specific Schiff base with an active site lysine in enzymes for which it is a cofactor (17). The occurrence of pyrroloquinoline quinone as a ligand (Table II) may reflect specific Schiff base formation with the same reactive lysine residue utilized by PLP, or nonspecific protein labeling. Pyrroloquinoline quinone is an amine-reactive quinone that, in addition to serving as an essential enzyme co-factor in some Gram-negative bacteria (18), has been found to bind to pyridoxal-dependent enzymes, although the significance and details of these interactions are not clear (19). In our initial characterization of this screening hit we observed atypical closing behavior (*e.g.* interference, nonspecific binding; not shown), and this lead was not pursued further. Many of the remaining hits generally occupy one of three categories: nucleotide-like (purines or pyrimidines), those resembling a sugar moiety, and other molecules.

Characterization of Pyridoxal Phosphate and Pyridoxamine Binding—The identification of PLP as a ligand in the screen prompted us to conduct additional studies to characterize its binding to CFE97. Fig. 3A shows the thermal transitions measured using ThermoFluor in the presence and absence of 75 μM PLP or PMP. PLP shifts the midpoint in the unfolding transition from 48 to 73 °C, an increase of 25 °C, consistent with tight binding of this potential co-factor; PMP under the same conditions increased the T_m by 5 °C. The weaker but nevertheless significant affinity of CFE97 for PMP is consistent with a co-factor role for PLP/PMP. The concentration dependence of T_m as a function of PLP and PMP was simulated using Equa-

FIG. 1. **Analytical ultracentrifugation data for CFE97.** *Top panel*, residuals for global fitting to the dimerization model. The randomness of the residuals indicates that the dimerization model describes the data well. Conditions were 20 mM HEPES, pH 7.5, 150 mM NaCl. *Bottom panel*, sedimentation equilibrium data collected at four rotor speeds (12,000, 15,000, 18,000 and 21,000 rpm). The data are shown as absorbance versus radial position. The data were fit globally to a dimerization model and the best-fit dimerization constant (K_d) was estimated to be 300 nM (± 78 nM). The monomer molecular mass was constrained in the analysis to the mass spectrometry value of 47,108 Da.

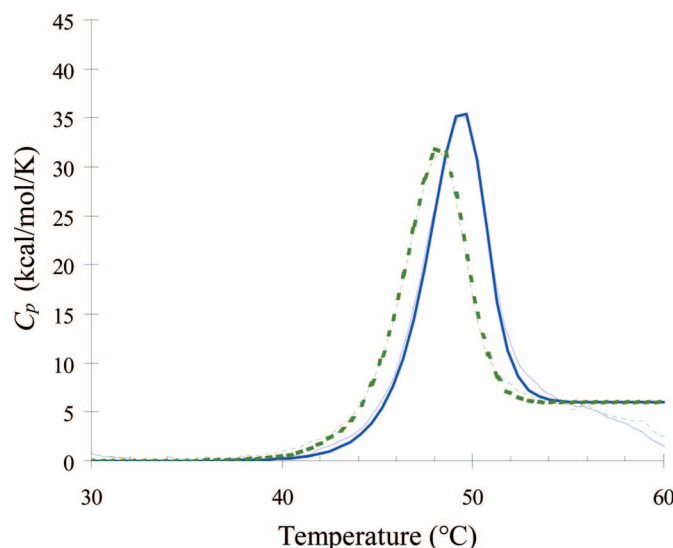
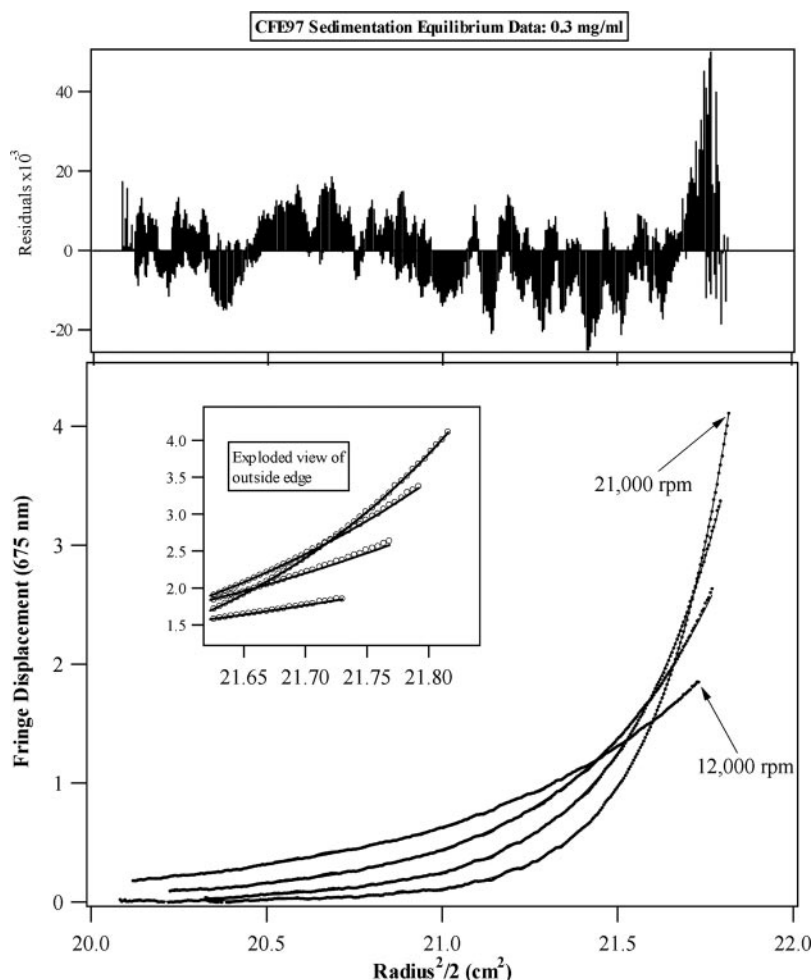


FIG. 2. **Thermal stability of CFE97 as measured by differential scanning calorimetry (DSC).** Desalted protein was diluted to 8 μ M (monomer) and analyzed by DSC as described under “Experimental Procedures” (raw data, *thin blue line*; fit, *thick blue line*). Protein stability was also examined in buffer + 2% Me_2SO , + 50 μ M Dpx (reference cell contained the same additives), and the conditions used for ThermoFluor assays (raw data, *thin green line*; fit, *dashed green line*). Data were fit to give the values listed in Table I.

tion 4 with calorimetrically determined $\Delta_U H_{T_m}$ and $\Delta_U C_p$ values (Table I). The effect of ligand concentration on T_m is not only highly dependent on these thermodynamic parameters (9)

TABLE I
Thermodynamic parameters for CFE97 unfolding

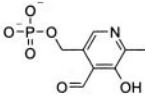
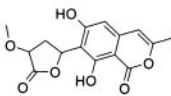
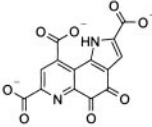
Parameter	Value no Me_2SO	Value + Dpx + Me_2SO^a
Monomeric unfolding model		
T_m	49.2 °C	
$\Delta_U H_{T_m}$	163 kcal mol ⁻¹	
$\Delta_U C_p$	6 kcal mol ⁻¹ K ⁻¹	
Dimeric unfolding model ^b		
T_m	56.5 °C	55.7 °C
$\Delta_U H_{T=56.5^\circ\text{C}}$	182 kcal mol ⁻¹	174 kcal mol ⁻¹
$\Delta_U S_{T=56.5^\circ\text{C}}$	552 cal mol ⁻¹ K ⁻¹	529 cal mol ⁻¹ K ⁻¹
$\Delta_U C_p$	6 kcal mol ⁻¹ K ⁻¹	6 kcal mol ⁻¹ K ⁻¹

^a 50 μ M Dpx, 2% Me_2SO .

^b Using dimeric unfolding models, the fit T_m (defined as where $\Delta_U G_{T_m} = 0$) is dependent on protein concentration. Thus the midpoint of a thermal transition using 4 μ M dimeric protein (Fig. 2) is significantly lower than this calculated T_m .

but is also model dependent. Equation 4 assumes a binding stoichiometry of one ligand bound per homodimer. A much better description of the dependence of T_m on ligand concentration was achieved using values for unfolding energetics equal to half of those measured, as might be expected if each homodimer bound two PLP molecules. These latter simulations, yielding K_d estimates of 50 pM (PLP) and 2.5 μ M (PMP), are shown in Fig. 3B. The difference in the binding affinities of the two forms of the co-factor is probably because of additional stabilization energy conferred by Schiff base formation upon binding of PLP. The shape of the PLP titration curve, in which the T_m continuously increases and fails to plateau even at the highest PLP concentrations, is consistent with reversible covalent binding (9). Mass spectrometry measurements for CFE97

TABLE II
Selected hits from a ThermoFluor screen of the functional probe library against CFE97

Description	Structure	Assay Concentration (mM)	ΔT_m (°C)	std. dev. (°C)
Pyridoxal 5'-Phosphate		0.1	25	0.3
Canescin A		0.1	21	0.3
Sodium Citrate		50	6.6	0.0
Nickel Sulfate		50	4.9	0.1
Pyrroloquinoline Quinone		0.5	4.4	0.2
D-Ribose-5-Phosphate		50	4.3	0.9
kinetin riboside		0.1	4	1
Potassium Tartrate		0.1	3.6	0.0
Pentamidine		2.1	3.1	0.3
EGTA		40	2.4	0.1
6-Furfuryl adenine (kinetin)		0.1	2.3	0.6
Cyclohexylamine		5	2.3	0.3
Sulfobenzoic acid		5	1.9	0.1
D-Erythrose-4-Phosphate		5	1.7	0.1
Inosine 5'-diphosphate		50	1.6	0.2
D-Glucose-6-Phosphate		50	1.5	0.1
O-Phospho-L-Serine		50	1.3	0.1
D-Ribulose 1,5-Phosphate		12.5	1.1	0.1
Uridine 5'-Diphosphoglucuronic Acid		12.5	1.1	0.1
D(-)Erythrose		50	1.0	0.1
(-)-Byssochlamic acid		0.1	0.9	0.1

treated with 1 mM PLP, then dialyzed, did not reveal an increase in molecular weight, confirming reversibility. The thermal stabilities of more than 20 other proteins have been measured in the presence and absence of PLP, and no instances of nonspecific, high-affinity PLP binding have been observed in these control studies. The binding of PMP, although weaker than that of PLP, is likely because of a specific interaction with the protein at the pyridoxal binding site. Taken together, these results suggest that the interaction between PLP/PMP and CFE97 is specific and may reflect a requirement for enzymatic function.

Canescin A Binding—Another apparently tight-binding ligand detected in the ThermoFluor screen was canescin A, a natural product that inhibits fungal growth by an unknown mechanism (Table II) (20). This compound is one of the substances in the functional probe library, such as natural products and generic drugs, that have biological activity for which the target of interaction is not precisely known. For such compounds, ThermoFluor can be used as a tool to identify potential targets for compounds, a type of “reverse screening” similar to other methods of target identification (21). The concentration dependence of the effect of canescin A on CFE97 thermal stability was obtained to estimate a binding constant. The titra-

tion curve reached a plateau after addition of 1 molar eq of ligand (Fig. 3B). This saturating effect on T_m is diagnostic of irreversible covalent ligands, because according to Equation 4, molecules that bind reversibly should continue to elevate the T_m well beyond a molar eq (as was observed for PLP and PMP). CFE97 could not simultaneously bind both PLP and canescin A, because no additional enhancement in thermal stability was observed when PLP was added to a preparation of CFE97 that had been preincubated with saturating amounts of canescin A (data not shown). Irreversible covalent binding of canescin A to CFE97 was confirmed by mass spectrometry. The molecular mass of the untreated protein was 47,108.3 daltons, whereas the treated, dialyzed protein yielded a molecular mass of 47,415.6 daltons. The difference, 307.3 daltons, was close to the molecular mass of canescin A, 306 daltons. Thus, canescin A binds covalently to CFE97. Our data may also be suggestive of competition between PLP and canescin A for a common binding site, but this remains to be conclusively established. This result may indicate that the protein target for this antifungal compound is a PLP-utilizing enzyme.

Sequence and Structure-based Analysis of CFE97 and Related PLP-binding Enzymes—A BLAST search (22) of the com-

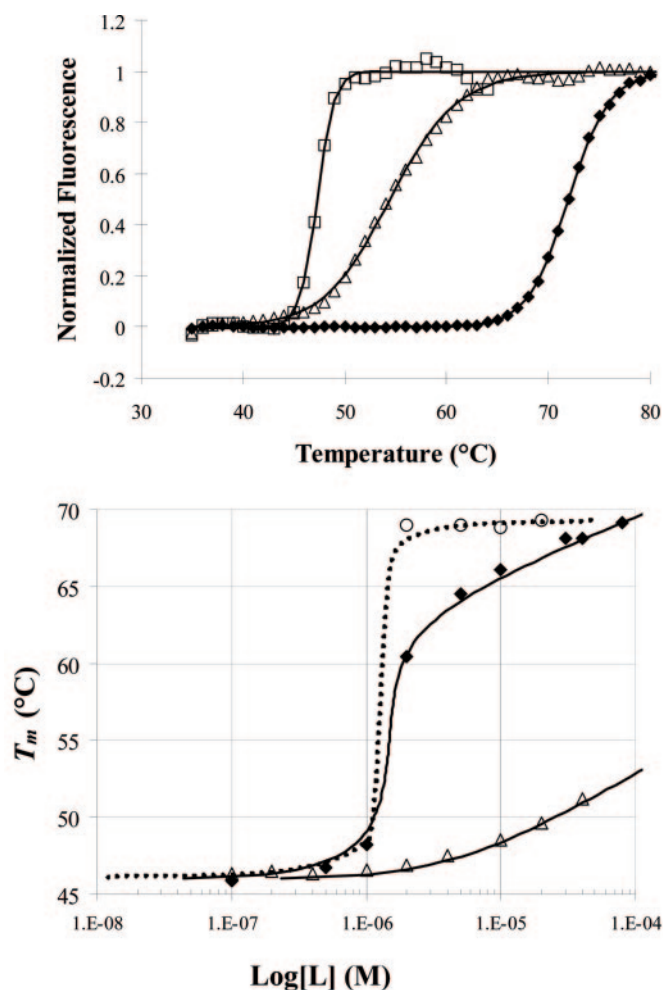


FIG. 3. CFE97 thermal stability increases in the presence of equilibrium binding ligands. *A*, thermal stability of CFE97 (2 μ M in 25 mM HEPES, 50 mM NaCl, 2.5 mM MgCl₂, 50 μ M Dpx, pH 7.5) was measured in the absence of ligand (open squares), 75 μ M PLP (closed diamonds), or 75 μ M PMP (open triangles). Theoretical curves fitted using Equation 1 are shown as solid lines. *B*, dependence of CFE97 T_m upon ligand concentration. T_m values shown were midpoints of thermal unfolding transitions, obtained from ThermoFluor data collected as in *A*. In parallel, thermal stability was determined using 11 concentrations of PLP (closed diamonds), PMP (open triangles), or canescin A (open circles). The solid lines are drawn using K_d values simulated according to Equation 4 using thermodynamic parameters for protein stability as determined in Fig. 2. The T_m is constant at all concentrations of canescin A above the protein concentration, because canescin A binds in an irreversible, covalent manner.

prehensive non-redundant sequence data base yielded many hits with greater than 30% sequence identity to CFE97 (not shown). The most similar protein found was Cps7G, a capsular protein from *Streptococcus suis* (23) with 74% sequence identity to CFE97. Encouragingly, in light of our initial ThermoFluor result showing tight PLP binding, many of these hits had known or putative aminotransferase functions. For example, UDP-bacillosamine synthetase from *Bacillus cereus* (24), the second most similar sequence hit (not shown), is 61% identical to CFE97. Overall, these results indicate that CFE97 has significant sequence homology with several classes of pyridoxal-utilizing enzymes, including those involved in the synthesis of nucleotide sugars.

A BLAST search (22) of the sequences of the protein structures available from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (25) yielded relatively few hits, and our analysis focused on the two most similar: 3-amino-5-hydroxybenzoic acid (AHBA) synthase from *Amycolatop-*

sis mediterranei (Protein Data Bank entries 1B9H and 1B9I (26)), and ArnB aminotransferase from *Salmonella typhimurium* (Protein Data Bank entries 1MDO, 1MDX, and 1MDZ (27)). AHBA synthase is involved in polyketide biosynthesis, serving an essential function in rifamycin B formation (26, 28). ArnB aminotransferase serves a key role in lipid A-mediated resistance to cationic peptide antimicrobials, transferring an amino group from glutamate to form UDP-4-amino-4-deoxy-L-arabinose (27). Interestingly, AHBA synthase was previously noted to be the closest known structural homolog of ArnB aminotransferase (27). Both of these enzymes use PLP as co-factor. Of particular relevance to the present study, a structure is available for each of these enzymes with the PLP co-factor bound (Protein Data Bank entries 1B9H (26) and 1MDX (27)).

Overall, CFE97 shares 25% sequence identity and 44% sequence similarity with *A. mediterranei* AHBA synthase, and 32% identity and 50% similarity with *S. typhimurium* ArnB aminotransferase (Fig. 4). These relatively low levels of sequence identity increase dramatically for the residues near the PLP binding site. 68% of the 25 residues closest to bound PLP in the ArnB aminotransferase crystal structure are identical to the aligned residues of CFE97; the corresponding number is 56% for AHBA synthase (Fig. 4). Furthermore, a tighter definition of the co-factor binding site revealed almost complete conservation of this functional site. For AHBA synthase, 7 of the 8 residues previously noted (26) as being involved in key PLP binding contacts are strictly conserved in CFE97, with the single change being a Phe to Tyr substitution (Fig. 4). Two additional binding site residues contributed by the adjacent monomer of the functional homodimer, Asn²³⁴ and Arg²³⁶, are not strictly conserved (Fig. 4), but in concert the two alterations (Asn to Lys and Arg to Asn) may allow for conservation of the relevant PLP binding interactions (see also Fig. 8 in Ref. 27). Thus, virtually all significant PLP binding contacts observed for AHBA synthase including hydrogen bonds, van der Waals interactions, and charge-charge contacts (26) appear to be conserved in CFE97. ArnB aminotransferase shows similar conservation of this more tightly defined PLP binding site. Two residue differences, Met to Leu and Trp to Tyr, are observed between CFE97 and ArnB aminotransferase over the 10-residue positions noted above (Fig. 4). Similar to AHBA synthase, it appears that all direct PLP binding interactions are conserved between CFE97 and ArnB aminotransferase. For both AHBA synthase and ArnB aminotransferase this includes strict conservation of the active site Lys that forms a Schiff base with PLP (Fig. 4; Lys¹⁸⁸ in both AHBA synthase and ArnB aminotransferase). Clearly there is a high degree of sequence conservation in the pyridoxal co-factor binding sites of these three enzymes. The results of this sequence and structure-based analysis support the conclusion that the tight binding of PLP observed in ThermoFluor reflects specificity for Schiff base formation between PLP and the probable active-site lysine of CFE97 (Fig. 4).

Identification of Potential Amino-donor Substrates for CFE97—Most PLP-dependent enzymes either catalyze aminotransferase reactions as their primary function or exhibit a secondary aminotransferase activity, *i.e.* they are “catalytically promiscuous” (15). We therefore designed ThermoFluor experiments to detect potential substrates for aminotransferase activity. Fig. 5 shows the results of an experiment that was designed to probe a panel of amino acids for potential amine donors. In this experiment, CFE97 was preincubated with PLP to form a complex, which was then assayed in the presence of potential amino donors. A successful amino donor should cause the T_m to decrease because of conversion of the PLP to PMP, which binds with lower affinity.

FIG. 4. Sequence alignment of CFE97 with 3-amino-5-hydroxybenzoic acid synthase and ArnB aminotransferase. Protein Data Bank entry 1B9H is *A. mediterranei* AHBA synthase complexed with PLP (26); Protein Data Bank entry 1MDX is *S. typhimurium* ArnB aminotransferase (27). Residues in red or blue are within 6 Å of the PLP bound to the corresponding protein monomer in the 1B9H crystal structure; residues in red were noted as making important direct contacts with bound PLP (26). The aligned positions of Lys²⁴¹ and Asn²⁴³ of ArnB aminotransferase (Asn²³⁴ and Arg²³⁶ in AHBA synthase) are highlighted in green (discussed in the text), and the 5-residue spacing between Ser¹⁸³ and Lys¹⁸⁸ of AHBA synthase is indicated with *hyphens above* the two positions (discussed in the text).

```

CFE97 -----MPNYNIPFSP--PDITEAEIAEVADTLRSG--WITTPGPKTELERRLSLYTQTPK
1MDX  MAEGKMMSDFLPFSR--PAMGAEELAAVKTVLDSG--WITTPGKQNELEAAFCRLTGNQY 56
1B9H  -----MNAKKAPEFFPAWPQYDAAERNGLVRALEQQQWWRMGDEVNSPEREFAAHHGAAH 55

CFE97  TVCLNSATAALELILRLVLEVGPGDEVIVPAMTYTASCSVITHVGATPVMVDIQADTFEMD
1MDX  AVAVSSATAGWHIALMALGIGEGDEVITPSMTWVSTLNMI VLLGANPVMVDVDRDTLMVT 116
1B9H  ALAVTNGTHALELALQVMGVGPGTEVIVPAFTFISSSQAAQLGAVTVPVVDVAATYNLD 115

CFE97  YDLLEQAITEKTKV IIPVELAGIVCDYDRLFQVVEKRRDFFTASSKWKQAFNRIVIVSDDS
1MDX  PEHIEAAITPQTKAIIPVHYAGAPADLDIAIYALGERYG-----IPVIEDA 161
1B9H  PEAVAAAVTPRTKVI PVHMAGLMADMDALAKISADTG-----VPLLODA 160

CFE97  ANALGSTYKQPSGSIADFTSFFHAVNFTTAWGGSATWKPANPVIDDEEMYKEFQILSL
1MDX  ANATGTSYKGRHIG-ARGTAIFSFHAINITCAEGG-----IVVTDNPFQADKLRLSKLF 214
1B9H  ANAHGARWQKGRVGLDLSIATFSFQNGRLMTAGEGG-----AVVFPDGETEKYETAFLR 214

CFE97  HGQTKDALAKMQLG-SWEYDIVTPAYKCNMTDIMASLGLVQLDRYPSLLQRRKDIVDRYD
1MDX  HGLGVDAWRDQSGGRAPQAEVLPAGYKYNLPDLNAAIALAQQLKLDALNARRAAIAAQYH 274
1B9H  HSCGRPRDDR-----RYFHKIAGSNMRLNEFSASVLAQLARLDEQIAVRDERWTLSS 267

CFE97  SGFAGSR-IHPLAHKTETVESSRHLYITRVEGVSLEE-RNLI IQELAKAGIASNVHYKPL
1MDX  QAMADLP-FQPLSLPSWEHIAWHLFIIRVDEARCGITRDALMASLKTGIGTGLHFRAA 333
1B9H  RLLGAIDGVVPGQGDVADRNRSHYAMAMFRIPGLTEER-RNALVDRLVEAGLPAFAAFRAI 328

CFE97  PLLTAYKNLGFDMTNYPKAYAFFENEITLPLHTKLSDEEVDYI IETFKTVSEKVLTLSSKLE
1MDX  HTQKYRER-FPTLTLPDTEWNSERICSLPLFPDMTESDFDRVITALHQIAGQGGSHHHHHH- 393
1B9H  YRTDAFWELGAPDESVDATARRCPNTDAISSDCVWLHHRVLLAGLPELHATAEIIADAVGRA 390

```

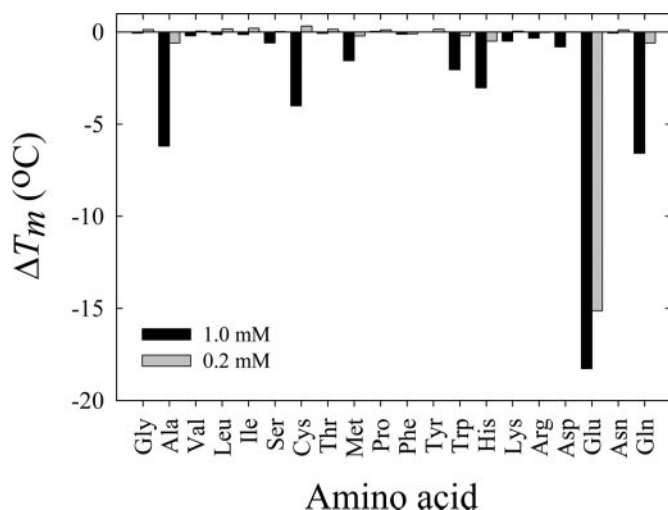


FIG. 5. Amine donors decrease CFE97-PLP melting temperature. CFE97 T_m was determined by ThermoFluor using 2 μ M protein in 25 mM HEPES, 50 mM NaCl, 2.5 mM MgCl₂, 50 μ M Dpx, 75 μ M PLP plus amino acids at 0.2 mM (gray bars) or 1 mM (black bars). The reference T_m of CFE97 + PLP was \sim 68 °C under these conditions, 22 °C higher than in the absence of PLP (Fig. 3B). Amine donors convert PLP to PMP, which is less stabilizing.

Screening the 20 L-amino acids against the pyridoxal-bound protein revealed that glutamate caused the largest decrease in the T_m (-18.5 °C), consistent with complete conversion to the enzyme-PMP complex even at the lower glutamate concentration (Fig. 5). Glutamate is a common substrate for aminotransferases, and its utilization by CFE97 supports the hypothesis that this enzyme is an aminotransferase rather than a member of another class of PLP-dependent enzyme (29).

Having identified potential amino donors using the ThermoFluor assay, an absorbance assay was used to monitor the half-reaction for enzymatic conversion of PLP to PMP. The UV-visible spectrum of purified CFE97 did not show absorbance in the visible region, as might be diagnostic of bound cofactor (Fig. 6A). Addition of PLP caused rapid appearance of an absorption peak characteristic of Schiff base formation (at 396 nm (17)) that was retained following rapid dialysis. Addition of the amino donor glutamate caused a slow change in the spectrum, inducing a peak at 335 nm (Fig. 6A), consistent with formation of a CFE97-PMP complex, as described previously

for *E. coli* aspartate transaminase (30). At low CFE97 concentrations in the presence of excess PLP and glutamate, this absorbance change could be exploited to monitor the slow conversion of PLP to PMP (Fig. 6B).

Steady-state Kinetics of Pyridoxal Phosphate Turnover—The rate of PLP conversion to PMP as a function of amino acid concentration was measured for each of the 20 L-amino acids. Results for substrates that yielded measurable reactions were analyzed using standard hyperbolic kinetics (Table III). Values for k_{cat} were essentially independent of the amino donor and were several orders of magnitude lower than rates typically observed for aminotransferases. This result suggests that all of the reactions have the same rate-limiting step, either reversal of CFE97 Schiff base formation or dissociation of PMP from CFE97, steps unlikely to occur in a full aminotransferase reaction. Significant variations in K_m were observed for different substrates in the half-reaction, with glutamate having the highest affinity. The preference for glutamate as a substrate is consistent with the effects measured in the ThermoFluor assay (Fig. 5). These data illustrate the utility of ThermoFluor as a generic tool for detecting ligands that modify the stability of an enzyme by converting a bound substrate to a product that binds with different affinity.

In an effort to identify molecules that could serve as amino acceptors, the functional probe library was screened *versus* CFE97 in the presence of 20 μ M PLP and 50 μ M glutamate. Under these conditions all tight-binding PLP was converted to the weak binding PMP. Compounds that elevate the T_m of CFE97-PMP may be equilibrium binding ligands, or may serve as amine acceptors in the reaction converting CFE97-PMP to CFE97-PLP. α -Ketoglutarate is the α -keto acid product formed by the half-reaction of CFE97-PLP with glutamate and is present in the library. As expected, α -ketoglutarate increased the T_m by 15 °C (not shown), presumably shifting the enzyme to the more stable CFE97-PLP form and serving as a positive control for the screen. Two other α -keto acids, α -keto adipate and oxaloacetate, were also identified as hits, producing shifts of 14 and 7 °C, respectively (not shown). Both of these α -keto acids are close structural analogs of α -ketoglutarate. Similar promiscuity was reported for ArnB aminotransferase (Ref. 27, and see below), consistent with a role for CFE97 as a sugar aminotransferase. α -Keto adipate is a substrate for an aminotransferase found in the yeast pathway for *de novo* biosynthesis of L-lysine from α -ketoglutarate; this pathway is not known to

FIG. 6. Absorption spectra of CFE97 and cofactors. A, absorption spectra of CFE97 (50 μM) in 25 mM HEPES, 50 mM NaCl, pH 7.5, with 0 (black) or 50 μM PLP (red). The blue line was obtained after adding 5 mM glutamate to protein + PLP, and is characteristic of a PMP absorbance spectrum. B, the slow conversion of PLP to PMP by CFE97. Absorbance spectra before and at various times after addition of 5 mM glutamate (prior to addition, red; ~30 s after addition, green; 10 min after addition, purple; 20 min after addition, blue).

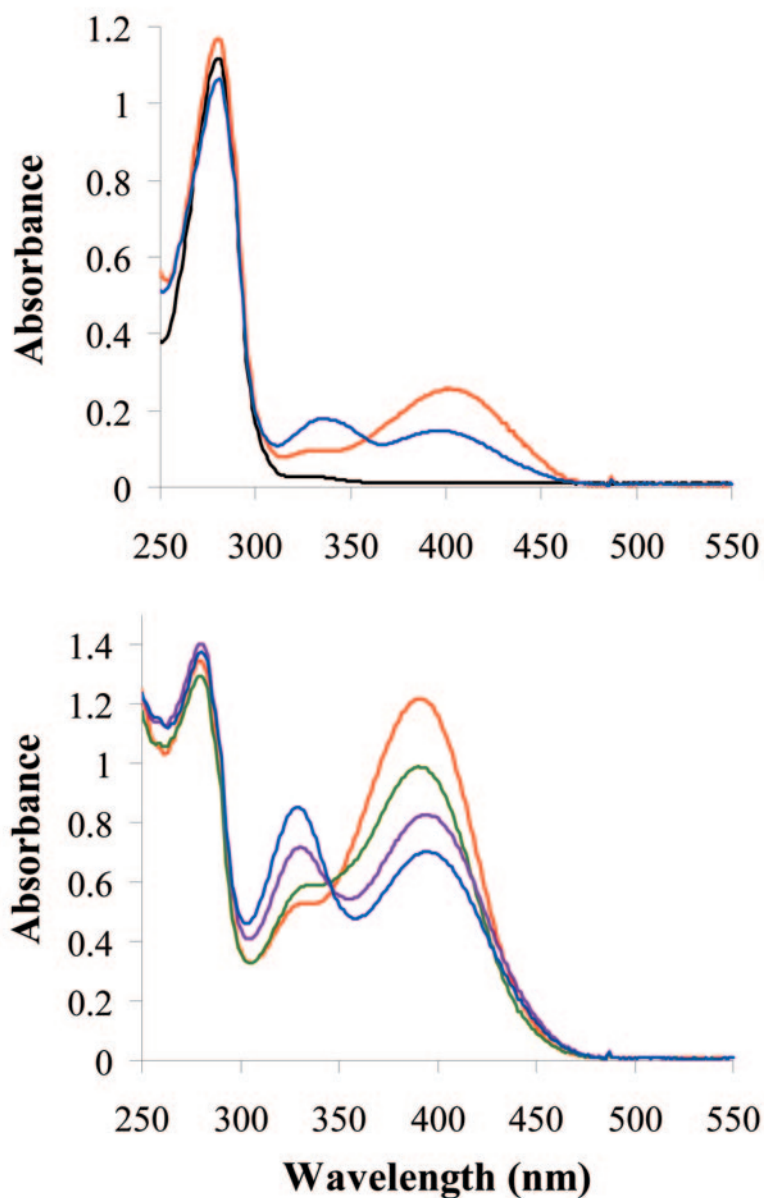


TABLE III
Kinetic parameters for substrates in the CFE97-catalyzed aminotransferase half-reaction

Amino acid	K_m mM	k_{cat} min^{-1}
Glu	0.32	1.2
Gln	3.3	1.7
Ala	2.1	1.3

TABLE IV
 ΔT_m observed for binding of uridine diphosphate and selected sugars to CFE97

UDP sugar	ΔT_m
5 mM UDP	0.6
5 mM UDP-5'-galactose	0.9
5 mM UDP-5'-glucose	0.55
5 mM UDP-5'-glucuronic acid	0.45
5 mM UDP-5'-N-acetylglucosamine	1.8

exist in bacteria. However, two similar transamination reactions occur in the bacterial synthesis of L-lysine from L-aspartate. The enzymes catalyzing these reactions utilize *N*-acetyl-L-2-amino-6-oxopimelate and *N*-succinyl-L-2-amino-6-oxopimelate as substrates (31). Amination of one or both of these intermediates has not been ruled out as a function for CFE97.

Binding of Sugars and Nucleotide Sugars to CFE97—In the initial functional probe library screen, several sugars and diphosphonucleotides were found to bind weakly (Table II). As noted above, the amino acid sequences most similar to that of CFE97 include those of several known and/or putative sugar aminotransferases. To investigate the possibility that a keto

sugar is the amino acceptor for this enzyme, a panel of 45 commercially available sugars and UDP sugar derivatives were screened with ThermoFluor at 500 μM . Several hits were found that could indicate a preferred substrate (Table IV). Because these sugars were not in their corresponding deoxy-keto forms, catalytic activity using these molecules as amino acceptor substrates could not be confirmed. The panel of commercially available sugars (primarily pentoses and hexoses) was retested at higher concentrations (10–50 mM; data not shown). We were unable to detect binding of any simple sugar to CFE97 using ThermoFluor. Only the UDP derivatives of glucose and galac-

tose yielded statistically significant T_m shifts. The results observed for UDP sugars, in addition to the observed homology to ArnB aminotransferase and AHBA synthase (see also below), may indicate that a UDP-keto-sugar or some other nucleotide diphospho-keto-sugar is the preferred amino acceptor substrate for this enzyme.

A more detailed comparison of the sequence of CFE97 to that of AHBA synthase and ArnB aminotransferase provided further clues about CFE97 function. Examples based on each of the two structural homologs are outlined below. Eads and co-workers (26) noted the 5-residue spacing between the conserved Ser¹⁸³ and the active site Lys¹⁸⁸ of AHBA synthase, observing that a 3-residue gap is more typical overall for the aspartate aminotransferase family, whereas the 5-residue spacing is consistent with a role in the biosynthesis of antibiotic sugar moieties. The 5-residue spacing is conserved in CFE97 (Fig. 4), suggesting a role for this enzyme in antibiotic sugar biosynthesis. Lys²⁴¹ of ArnB aminotransferase appears to play a role in substrate selectivity (27). Interaction between the side chain amino group and the terminal carboxylate of L-glutamate may contribute to preference for this substrate (27). Lys²⁴¹ is conserved in CFE97 (Fig. 4), and may serve a similar function, contributing to the observed preference of CFE97 for L-glutamate as the amino donor (Fig. 5, Table III). Although it is not always straightforward to translate sequence- and structure-based similarities and differences into molecular details regarding possible substrates, inferences such as these can serve to guide and augment the selection of molecules to be tested in follow-up studies aimed at further defining enzyme function.

CONCLUSIONS

ThermoFluor, an affinity-based screening technology, was used to search for ligands that bind CFE97, the protein product of a gene (*sp1837*) of unknown function from *S. pneumoniae*. The results of this screen indicate that CFE97 binds PLP and can utilize it as a cofactor in an enzymatic reaction. Biophysical characterization of CFE97 as a dimer is consistent with the biochemical significance of PLP binding. Further experiments established that CFE97 preferentially utilizes glutamate as the amino donor in the first step of an aminotransferase reaction, and that it can bind UDP sugar derivatives. Combining this biochemical and biophysical analysis with protein sequence and structure analysis has provided a more definitive assignment of the putative function of CFE97 than would be possible using sequence comparison alone. This body of data collectively suggests that this enzyme is likely to be an aminotransferase involved in the biosynthesis of a diphosphonucleotide amino sugar.

A general approach has been demonstrated for confirming or disproving existing functional hypotheses for an orphan protein and for discovering novel, previously uncharacterized functions. Furthermore, for enzyme classes that catalyze reactions employing many different biological substrates, affinity-based functional screens are a useful method for identifying

probable substrates for a particular enzyme. These substrates provide directions that can then be experimentally investigated in more detail to identify the function of an orphan protein. This high-throughput approach to functional assignment of expressed protein products may yield a wealth of new data for understanding protein function.

Acknowledgment—We thank Prof. John C. Vederas of the Chemistry Department of the University of Alberta for providing compounds for testing.

REFERENCES

1. Yakunin, A. F., Yee, A. A., Savchenko, A., Edwards, A. M., and Arrowsmith, C. H. (2004) *Curr. Opin. Chem. Biol.* **8**, 42–48
2. Chan, P. F., Macarron, R., Payne, D. J., Zalacain, M., and Holmes, D. J. (2002) *Curr. Drug Targets Infect. Disord.* **2**, 291–308
3. Strausberg, R. L., and Schreiber, S. L. (2003) *Science* **300**, 294–295
4. Adam, G. C., Sorensen, E. J., and Cravatt, B. F. (2002) *Mol. Cell Proteomics* **1**, 781–790
5. MacBeath, G., Koehler, A. N., and Schreiber, S. L. (1999) *J. Am. Chem. Soc.* **121**, 7967–7968
6. Williams, C. (2000) *Curr. Opin. Biotechnol.* **11**, 42–46
7. Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B. A., Lane, P., and Salemme, F. R. (2001) *J. Biomol. Screen* **6**, 429–440
8. Thanassi, J. A., Hartman-Neumann, S. L., Dougherty, T. J., Dougherty, B. A., and Pucci, M. J. (2002) *Nucleic Acids Res.* **30**, 3152–3162
9. Matulis, D., Kranz, J. K., Salemme, F. R., and Todd, M. (2005) *Biochemistry*, in press
10. Brandts, J. F., Hu, C. Q., Lin, L. N., and Mos, M. T. (1989) *Biochemistry* **28**, 8588–8596
11. Brandts, J. F., and Lin, L. N. (1990) *Biochemistry* **29**, 6927–6940
12. Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G., and Schomburg, D. (2004) *Nucleic Acids Res.* **32**, D431–D433
13. Doyle, M. L., Tian, S. S., Miller, S. G., Kessler, L., Baker, A. E., Brigham-Burke, M. R., Dillon, S. B., Duffy, K. J., Keenan, R. M., Lehr, R., Rosen, J., Schneeweis, L. A., Trill, J., Young, P. R., Luengo, J. I., and Lamb, P. (2003) *J. Biol. Chem.* **278**, 9426–9434
14. Todd, M. J., Semo, N., and Freire, E. (1998) *J. Mol. Biol.* **283**, 475–488
15. John, R. A. (1995) *Biochim. Biophys. Acta* **1248**, 81–96
16. Jansonius, J. N. (1998) *Curr. Opin. Struct. Biol.* **8**, 759–769
17. Toney, M. D., and Kirsch, J. F. (1993) *Biochemistry* **32**, 1471–1479
18. Klinman, J. P., and Mu, D. (1994) *Annu. Rev. Biochem.* **63**, 299–344
19. Choi, S. Y., Khemlani, L. S., and Churchich, J. E. (1992) *Biofactors* **3**, 191–196
20. Angawi, R. F., Swenson, D. C., Gloer, J. B., and Wicklow, D. T. (2003) *J. Nat. Prod.* **66**, 1259–1262
21. Smith, C. (2004) *Nature* **428**, 225–231
22. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410
23. Smith, H. E., van Bruijnsvoort, L., Buijs, H., Wisselink, H. J., and Smits, M. A. (1999) *FEMS Microbiol. Lett.* **178**, 265–270
24. Ivanova, N., Sorokin, A., Anderson, I., Galleron, N., Candelon, B., Kapatral, V., Bhattacharyya, A., Reznik, G., Mikhailova, N., Lapidus, A., Chu, L., Mazur, M., Goltzman, E., Larsen, N., D'Souza, M., Walunas, T., Grechkin, Y., Pusch, G., Haselkorn, R., Fonstein, M., Ehrlich, S. D., Overbeek, R., and Kyrpides, N. (2003) *Nature* **423**, 87–91
25. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res.* **28**, 235–242
26. Eads, J. C., Beeby, M., Scapin, G., Yu, T. W., and Floss, H. G. (1999) *Biochemistry* **38**, 9840–9849
27. Noland, B. W., Newman, J. M., Hendle, J., Badger, J., Christopher, J. A., Tresser, J., Buchanan, M. D., Wright, T. A., Rutter, M. E., Sanderson, W. E., Muller-Dieckmann, H. J., Gajiwala, K. S., and Buchanan, S. G. (2002) *Structure (Lond.)* **10**, 1569–1580
28. Kim, C. G., Yu, T. W., Fryhle, C. B., Handa, S., and Floss, H. G. (1998) *J. Biol. Chem.* **273**, 6030–6040
29. Percudani, R., and Peracchi, A. (2003) *EMBO Rep.* **4**, 850–854
30. Inoue, Y., Kuramitsu, S., Inoue, K., Kagamiyama, H., Hiromi, K., Tanase, S., and Morino, Y. (1989) *J. Biol. Chem.* **264**, 9673–9681
31. Patte, J. C., Morand, P., Boy, E., Richaud, C., and Borne, F. (1980) *Mol. Gen. Genet.* **179**, 319–325