Complete Reconstitution of the Human Coenzyme A Biosynthetic Pathway via Comparative Genomics*

Received for publication, February 20, 2002, and in revised form, March 18, 2002
Published, JBC Papers in Press, March 28, 2002, DOI 10.1074/jbc.M201708200

Matthew Daugherty‡, Boris Polanuyer‡, Michael Farrell‡, Michael Scholle‡, Athanasios Lykidis‡, Valérie de Crécy-Lagard‡, and Andrei Osterman‡‡

From ‡Integrated Genomics, Incorporated, Chicago, Illinois 60612 and the §Scripps Research Institute, La Jolla, California 92037

The biosynthesis of CoA from pantothenic acid (vitamin B₅) is an essential universal pathway in prokaryotes and eukaryotes. The CoA biosynthetic genes in bacteria have all recently been identified, but their counterparts in humans and other eukaryotes remained mostly unknown. Using comparative genomics, we have identified human genes encoding the last four enzymatic steps in CoA biosynthesis: phosphopantothenoylcysteine synthetase (EC 6.3.2.5), phosphopantothenoylcysteine decarboxylase (EC 4.1.1.36), phosphopantetheine adenyltransferase (EC 2.7.7.3), and dephospho-CoA kinase (EC 2.7.1.24). Biological functions of these human genes were verified using a complementation system in Escherichia coli based on transposon mutagenesis. The individual human enzymes were overexpressed in E. coli and purified, and the corresponding activities were experimentally verified. In addition, the entire pathway from phosphopantothenate to CoA was successfully reconstituted in vitro using a mixture of purified recombinant enzymes. Human recombinant bifunctional phosphopantetheine adenyltransferase/dephospho-CoA kinase was kinetically characterized. This enzyme was previously suggested as a point of CoA biosynthesis regulation, and we have observed significant differences in mRNA levels of the corresponding human gene in normal and tumor cells by Northern blot analysis.

CoA is an indispensable cofactor in all living organisms, where it functions as an acyl carrier and carbonyl-activating group in a number of central biochemical transformations, including the tricarboxylic acid cycle and fatty acid metabolism. It has been estimated that ~4% of all known enzymes utilize CoA as an obligate cofactor (1). Many bacteria as well as plants and yeast are capable of de novo CoA biosynthesis from aspartate and ketovalerate pantothenic acid. In contrast, animals and some pathogenic microbes lack a de novo route, and they are totally dependent on scavenging exogenous pantothenic acid (pantothenate, vitamin B₅). Dietary vitamin B₅ deficiency in mammals causes many systemic effects such as impaired motor response, depressed heme synthesis, altered growth and maturation of the small intestines during neonatal and prenatal periods, and increased prenatal mortality. Levels of CoA are affected during metabolic stress such as starvation, alcoholism, and diabetes as well as in certain tumors (2). Recently, the first gene of the universal CoA biosynthetic pathway was linked with the neurodegenerative Hallervorden-Spatz syndrome in humans (3) and with aberrant mitosis and meiosis in flies (4), setting a precedent for the importance of this pathway in many aspects of cellular and metabolic development.

The biosynthesis of CoA from pantothenate contains five universal steps (see Fig. 1). Pantothenate is first phosphorylated by pantothenate kinase (EC 2.7.1.33). 4’-Phosphopantetheine is conjugated with cysteine by phosphopantetheinylcysteine synthetase (PPCS; 6.3.2.5) and then converted to 4’-phosphopantetheine by phosphopantetheinylcysteine decarboxylase (PPDC; EC 4.1.1.36). The final two steps are adenylation by phosphopantetheine adenyltransferase (PPAT; EC 2.7.7.3) to form dephospho-CoA and phosphorylation by dephospho-CoA kinase (DPCK; EC 2.7.1.24) to form CoA. Mechanistic details of CoA biosynthesis were studied in bacteria, yeast, and mammals, and most of the corresponding enzymes were at least partially purified and characterized (for the last reviews, see Refs. 1 and 2). However, as neither CoA nor any of its phosphorylated precursors can be transported across the cell membrane, classical auxotrophic mutant analysis methods could not be used, and the identification of CoA biosynthetic genes was significantly delayed.

Pantothenate kinase, the first gene in this pathway, was initially identified in Escherichia coli (coaA) (5). In the last 2 years, E. coli genes were identified for the last four enzymatic steps: coaBC (previously dfp), encoding a bifunctional PPCS/PPDC (6, 7); coaD (previously kdtB), encoding PPAT (8); and coaE (previously yacE), encoding DPCK (9). Orthologs of all of these genes are preserved in all bacteria, except some obligate intracellular parasites such as Mycoplasma, Rickettsia, and Chlamydia, suggesting that these organisms have developed a specialized CoA (or dephospho-CoA) transport system.

When we began this study, only the gene for pantothenate kinase, which revealed no sequence similarity to bacterial CoaA, had been identified in the mammalian CoA biosynthetic pathway (10, 11). Our goal was to identify, via a comparative genomics approach, human genes encoding the four remaining enzymatic steps in CoA biosynthesis to enable further genetic and expression studies of this critically important pathway in health and disease.

We report here the identification and experimental verification of human genes encoding PPCS, PPDC, PPAT, and DPCK, thereby complementing the CoA biosynthetic pathway in...
humans. The human PPAT domain was identified based on biochemical evidence of a bifunctional PPAT/DPCK protein and reveals no sequence similarity to the previously described bacterial CoaD enzyme family. This is the first representative of the novel PPAT family present in eukaryotes and archaea. For this reason and the previous implication that PPAT/DPCK is involved in regulation of CoA biosynthesis (2), we have studied this protein in more detail. Steady-state kinetic parameters were obtained, and the possibility of transcriptional regulation of PPAT/DPCK gene expression was addressed using Northern blot analysis of various human healthy tissues and cancer cell lines.

Fig. 1. Universal coenzyme A biosynthetic pathway. PANK, pantothenate kinase (coaA in E. coli); PPCS, coaB; PPCDC, coaC; PPAT, coaD; DPCK, coaE.

Fig. 2. Structural comparison of CoA biosynthetic genes across all three kingdoms and reconstruction of a full-length cDNA encoding human PPAT/DPCK. A, a comparison of the predicted human PPCDC and PPCS with the homologous proteins in lower eukaryotes (represented by S. cerevisiae), the predicted bifunctional homolog in archaea (M. jannaschii), and the confirmed PPCS/PPCDC protein in E. coli. Amino acid residues identical to the human proteins are shown as black bands, and the overall score (P-score) is shown below each protein or domain. B, construction of a full-length cDNA and a predicted open reading frame (ORF) for human PPAT/DPCK. Boxes represent protein coding sequences, and black lines represent nucleotide sequences. All numbers indicate nucleotide numbers, with 1 representing the first base pair of the complete reconstructed human cDNA encoding PPAT/DPCK. Labels indicate identifiers in GenBankTM with their respective annotations. C, comparison of the predicted human three-domain PPAT/DPCK protein with representatives of the other three kingdoms as described for A. The highly conserved nucleotidyltransferase (NTase) motif (HXXH) region is shown. No comparison could be made with archaeal DPCK, as this gene, as well as the gene for pantothenate kinase, remains missing in all archaea.
Identification of Human CoA Biosynthetic Genes

EXPERIMENTAL PROCEDURES

Strains, Plasmids, and Other Reagents—E. coli strains DH5α and BL21 (Stratagene, La Jolla, CA) were used for cloning and protein overexpression, and DH10B (Invitrogen) was used as a host strain for complementation analysis by transposon mutagenesis. For expression of all genes in E. coli, the vector pPROEX-HTa (Invitrogen), containing the trp promoter, a His6 tag, and a tobacco etch virus protease cleavage site, was used. Human brain cDNA and multiple-tissue Northern blots were from CLONTECH (Palo Alto, CA). The DNA probe radiolabeling kit was from Invitrogen, and Microspin columns for removal of unincorporated nucleotides were from Amersham Biosciences. Enzymes for PCR and DNA manipulations were from New England Biolabs Inc. (Beverly, MA), MBI Fermentas (Vilnius, Lithuania), and CLONTECH.

PCR Amplification and Cloning—The three predicted complete coding regions were amplified using the three sets of primers indicated below. Introduced restriction sites (NcoI or BspHI for the 5′-end and SacI for the 3′-end) are shown in boldface; nucleotides not present in the original sequence are shown in lowercase. PPCS (gi 13638573): no mutations introduced, 5′GCGGAAATGGATCCGGTAGCC and 5′GTGGTGCTTTCACACCGTGCCAG; PPCD (gi 14042206): AGT (Ser) codon inserted after the first codon, 5′-ggggatccgATGGAAAGATGTCGACGAGGG and 5′-ggggatccCTAGTTTCTGTCACCTATAAAAGCTGTGTGTCGAG; PPAT/DPCK (gi 17981024): second codon mutated from GCC to ACC (Ala to Thr), 5′-gggatccagtGAACCAAAGGCTCGTCCACAGCTG and 3′-gggatccagTCAACTCTGCTGGAGACACCTGTGCTG; and PPAT/DPCK (gi 17981024): double codon mutated from GCC to ACC, 5′-gggatccagTCATGCATCGCCACCATCGTCT and 3′-gggatccagTCATCGGACCTGGTCAAGGATCCACAGCTG.

Purification kits were from Invitrogen, and Microspin columns for removal of unincorporated nucleotides were from Amersham Biosciences. Enzymes for PCR and DNA manipulations were from New England Biolabs Inc. (Beverly, MA), MBI Fermentas (Vilnius, Lithuania), and CLONTECH.

CoA biosynthetic pathway (coaA, coaBC, coaD, and coaE) were revealed as essential. Here, we performed the same analysis in the presence and absence of expression plasmids containing the predicted human CoA biosynthetic genes. Functional expression of human CoA biosynthetic genes was expected to complement the loss of function that would result from transposon insertion into the corresponding E. coli genes as revealed by PCR mapping.

Identification of Human CoA Biosynthetic Genes

Gene (Human PPAT/DPCK) or with and without complementation by the corresponding human genes. A, schematic of genetic footprinting based on nested PCR. Transposon insertion into a gene of interest (GenA) is detected and mapped using two sets of nested primers, indicated as double arrows: (i) specific for a corresponding locus in the E. coli chromosome and (ii) specific for transposon flanking regions. Using these sets of primers for the analysis of the total chromosomal DNA from a population of surviving transposon mutants allows the mapping of the positions of transposon insertion throughout the corresponding locus. B, agarose gel analysis of PCR fragments in genetic footprinting. The results of the analysis are shown for the two loci of the E. coli chromosome containing genes coaD (PPAT) and coaE (DPCK). PCR amplification with the corresponding sets of primers was performed without complementation (Vector Control) or with functional complementation by the corresponding human gene (Human PPAT/DPCK) provided in the expression vector. The regions on the agarose gel corresponding to the boundaries of insertions within the CoA biosynthetic genes are boxed: coaD gene (solid boxes) and coaE gene (dashed boxes) boundaries. Each condition is represented by two lanes corresponding to the detection of both possible orientations of the transposon by two different sets of transposon-specific primers. Lane M contains 1-kb DNA ladder markers. C, mapping of the transposon insertions within E. coli Coa biosynthetic genes. Genetic footprinting and agarose gel analysis of PCR fragments were performed in the three loci of the E. coli chromosome containing genes coaBC (PPCS/PPCDC), coaD (PPAT), and coaE (DPCK). The positions of detected transposon insertions mapped in the corresponding loci are shown by vertical lines. CoA biosynthetic genes are shown as black arrows, and the adjacent genes are shown as white arrows. Insertions in the experiment without complementation (Vector Control) are shown above the respective genes. The three genes coaBC, coaD, and coaE are essential for E. coli growth and survival, and they do not contain transposon insertions in the experiment without complementation. Coexpression of the two human enzymes PPCS and PPCDC provided on the plasmids complements for the loss of function of coaBC, as evidenced by the appearance of transposon insertions within the boundaries of this gene (shown below the locus representation). Similarly, the vertical lines below the coaD and coaE loci reflect functional complementation of these genes in the presence of human bifunctional PPAT/DPCK.

Available at www.ncbi.nlm.nih.gov/Genbank/index.html.

Available at genome-www.stanford.edu/Saccharomyces. Available at www.bioinformatics.org/avf/.

Available at www.scp.berkeley.edu/.

Available at www.bbc.co.uk/.

Available at www.sciencemag.org/

Available at www.astrobiology.org/.

Available at www.surfacephysics.org/.

Available at www.genome.gov/.

Available at www.davideprojects.org/.
Identification of Human CoA Biosynthetic Genes

| Enzymatic activity and substrate | \( k_{\text{cat}} \) | \( K_{\text{m(app)}} \) | \( h_{\text{cat}}/K_{\text{m(app)}} \) |
|-------------------------------|----------|-----------------|---------------|
| Forward PPAT reaction         |          |                 |               |
| 4'-Phosphopantetheine (5 mM ATP) | 13.5 ± 0.38 | 7.6 ± 1.3 | 1.8 x 10^{10} ± 0.3 x 10^{10} |
| ATP (100 mM 4'-phosphopantetheine) | 10.3 ± 0.7 | 145 ± 29.8 | 7.1 x 10^{10} ± 1.5 x 10^{10} |
| Reverse PPAT reaction         |          |                 |               |
| Dephospho-CoA (2 mM PP)        | 26.4 ± 0.96 | 41.3 ± 3.6 | 6.4 x 10^{10} ± 0.6 x 10^{10} |
| PP (100 mM dephospho-CoA)      | 26.5 ± 1.76 | 769 ± 116 | 3.4 x 10^{10} ± 0.6 x 10^{10} |
| Forward DPCK reaction          |          |                 |               |
| Dephospho-CoA (1 mM ATP)       | 12.2 ± 0.25 | 16.7 ± 1.92 | 7.5 x 10^{10} ± 0.9 x 10^{10} |
| ATP (100 mM dephospho-CoA)     | 14.2 ± 0.59 | 34.4 ± 6.6 | 4.1 x 10^{10} ± 0.8 x 10^{10} |


\( K_{\text{m(app)}} \) was obtained by varying the substrate concentration at a constant concentration of the second substrate, as shown.

Briefly, transposons were preformed by incubating 7 ng/µl artificial transposon EZ::TN(KAN-2) DNA with 0.1 unit/µl Tn5 EZ::TN transposase (Epitcenc Technologies Corp., Madison, WI) for 30 min at 37 °C. Following dialysis against 10 mM Tris (pH 7.5) and 1 mM EDTA on 0.05-µm filters (Millipore Corp., Bedford, MA), samples were transferred by electroporation to DH10B and grown in enriched LB medium on 0.05-

**Recombinant proteins**—Recombinant proteins were overexpressed as an N-terminal fusion with a His tag and a tobacco etch virus protease cleavage site in E. coli strain BL21. Cells were grown to \( A_{600} = 0.8–1.0 \) at 37 °C, isopropyl-\( \beta \)-D-thiogalactopyranoside was added to 0.8 mM; and harvesting was performed after ~12 h of shaking at 20 °C. All proteins were purified from 50-ml cultures using nickel-nitrilotriacetic acid agarose chromatography and gel filtration on a HiLoad Superdex 200 10/30 column (Amersham Biosciences) for more detailed kinetic analysis and crystallization trials. Truncated versions of this protein (domains I and II (amino acids 1–358) and domains II and III (amino acids 180–565)) were PCR-amplified using a corresponding set of primers and cloned in the same expression vector. The C-terminally truncated version (domains I and II) was expressed with a comparable yield of the soluble protein, preserved PPAT activity, and no DPCK activity. The second, N-terminally truncated version produced inactive protein in the form of inclusion bodies.

**Coupled Enzyme Assays**—All coupled assays were performed in 500 µl of 50 mM Tris (pH 8.0) and 2 mM MgCl₂ using a Beckman DU-640 spectrophotometer to monitor the change in absorbance at 340 nm in a six-cuvette assembly thermostatted at 37 °C. Additionally, the results of CoA biosynthetic enzyme assays were verified by direct HPLC analysis as described below, monitoring formation and/or consumption of ATP, ADP, AMP, dephospho-CoA, and CoA at 254 nm.

**PPCS Activity Assay**—The release of PP was coupled to the oxidation of NADH to NAD and detection at 340 nm as modified (14). Briefly, 1.0 mM 4'-phosphopantetheine was incubated with 5 mM cysteine, 1 mM either ATP or CTP, -0.5 µm partially purified human PPCS, and 200 µl of pyrophosphate reagent.

**PPDC Activity Assay**—An assay was developed in which formation of 4'-phosphopantetheonate was coupled to consecutive enzymatic Coa formation and release of PP, which was detected utilizing the same technique as described above. 4'-Phosphopantetheonolycysteine was incubated with 50 µg/ml partially purified PPCDC, 25 µg/ml purified PPAT/DPCK, 5 mM ATP, and 200 µl of pyrophosphate reagent. Control samples lacking PPCDC showed no signal above the background rate.

**Forward PPAT Activity Assay**—PPAT activity was verified by detecting the release of PP, as described above. Briefly, 5–500 µM 4'-phosphopantetheine was incubated with 5 mM ATP and 0.15–0.3 µg/ml purified human PPAT/DPCK in the presence of pyrophosphate reagent detection mixture.

**Reverse PPAT Activity Assay**—The reverse PPAT reaction assay was performed by coupling the release of ATP to the reduction of NADP to NADPH and detection at 340 nm as described (14). Briefly, 5–500 µM dephospho-CoA was incubated with 2 mM PP, 5 mM glucose, 1 mM NADP, 2 units of hexokinase, 1 unit of glucose-6-phosphate dehydrogenase, and 0.5–1.0 µg/ml purified human PPAT/DPCK.

**DPCK Activity Assay**—The DPCK activity of the full-length PPAT/DPCK protein was determined by a standard technique: coupling the release of ADP to the oxidation of NADH to NAD as described (12). Briefly, 5–100 µM dephospho-CoA was incubated with 1 mM ATP, 2 mM phosphoenolpyruvate, 0.3 mM NADH, 2.5 units of lactate dehydrogenase, 1.25 units of pyruvate kinase, and 0.15–0.3 µg/ml purified human PPAT/DPCK.

To obtain kinetic parameters for the bifunctional PPAT/DPCK enzyme, the apparent \( K_{m} \) of one substrate was determined by varying the concentration of that substrate while keeping the second substrate at a constant saturating concentration. An NADH or NADPH extinction coefficient of 6.22 mM⁻¹ cm⁻¹ was used for rate calculations.

**In Vitro Reconstitution of CoA Biosynthesis from 4'-Phosphopantetheonate**—Conversion of 4'-phosphopantetheonate to CoA using purified recombinant enzymes was determined by HPLC. The reaction mixture contained 50 mM Tris (pH 8.0), 10 mM MgCl₂, 20 mM KCl, 1 mM dithiothreitol, 0.5 mM 4'-phosphopantetheonate, 2 mM ATP, 1 mM cysteine, and 1–10 µM concentrations of all three purified recombinant proteins: PPCS, PPDC, and PPAT/DPCK. Following overnight incubation at 37 °C, reactions were filtered through Microcon 10-MW (Millipore Corp.), and samples were diluted 1:10 in 100 µM fresh dithiothreitol to reduce dephospho-CoA disulfide. Samples were analyzed using a Gilson HPLC system with a Model 322 pump, UV/VIS-152 detector, and Model 234 auto-sampler. Isocratic separation was performed at 1.5 ml/min on a Supelcosil LC-18 column (5 µm, 4.6 cm, 5 µm Supelco Inc., Bellefonte, PA) with UV detection at 250 nm. The eluent was 100 mM sodium phosphate (pH 5.5), 8 mM tetrabutylammonium bromide, and 10% acetonitrile. Molar areas were determined using calibration with standard solutions of dephospho-CoA and AMP under the same chromatographic conditions.

**Northern Blot Analysis**—Blots with 2 µg of mRNA from normal human tissues and human cancer cell lines were analyzed by hybridization with a radiolabeled DNA probe following the manufacturer’s protocol (CLONTECH). Briefly, the DNA probe (PCR fragment corresponding to the central domain (PPAT domain) of human PPAT/DPCK) was radiolabeled with \([\alpha-\text{32P}]\text{dATP}\) using a commercially available kit (Inno-Genetix), and unincorporated nucleotides were removed using Microspin G-50 columns. The normal cell multiple-tissue blot and multiple-cell cancer cell line blot (CLONTECH) were preincubated for 30 min in ExpressHyb solution supplied by the manufacturer. The labeled DNA probe was mixed with fresh ExpressHyb solution and incubated with the membranes for 1 h at 68 °C. The blots were rinsed several times as suggested by the manufacturer and then exposed to x-ray film at -70 °C. The blots were subsequently washed and rehybridized with radiolabeled β-actin to normalize for mRNA loading levels.
RESULTS

Prediction of Human CoA Biosynthetic Genes Using Comparative Genomics—The universal CoA biosynthetic pathway is presented in Fig. 1. Previous biochemical analysis established the activity of the CoA biosynthetic enzymes in humans (for review, see Ref. 2). However, before we began this study, only the gene for human pantothenate kinase was known (3). The bacterial genes for the CoA biosynthetic enzymes have all been recently elucidated (for review, see Ref. 1). Given the conservation at the functional level of this pathway between humans and bacteria, we attempted to project the genes from bacteria to humans using comparative genomics.

PSI-BLAST searches allowed us to identify three proteins in the human cDNA sequence database available from GenBank™ as strong homologs of *E. coli* CoA biosynthesis enzymes. One homolog was found for PPCDC (gi 14042206, P-score ~ 10^{-12}), and two homologs were found for DPCK (gi 13623688 and gi 13376838, P-score ~ 10^{-27} and 10^{-15}, re-
respectively). No reliable homologs could be found for *E. coli* PPCS or PPAT.

The predicted human PPCDC appeared to be a monofunctional enzyme. This is in contrast to most bacteria, in which PPCDC is fused with PPCS, forming a bifunctional CoaBC protein. Early biochemical data indicate that human PPCDC and PPCS activities in fact reside in two separate proteins (15, 16). Enzymatic activity was recently verified for the plant PPCDC ortholog (17), which was previously described as a halotolerance protein (18) and characterized at the three-dimensional structure level (19). Homologs of human PPCDC were found in all analyzed eukaryotic genomes. Among prokaryotic genomes, only streptococci and enterococci contain monofunctional PPCDC genes. Bacterial monofunctional PPCDC from these organisms shows the highest sequence similarity to human monofunctional PPCDC. In the same bacterial genomes, PPCS is also monofunctional and is found in the same operon with PPCDC. Using this unique monofunctional PPCS from *Streptococcus pneumoniae* (gi 14972712), we were able to identify a candidate for human monofunctional PPCS (gi 13638573) with a reliable similarity (P-score = 10^-7). The marginal similarity of PPCS domains of bacterial bifunctional proteins to putative human monofunctional PPCS (P-score = 0.02 for the *E. coli* PPCS domain) was insufficient to predict this candidate with any degree of certainty. Homologs of the putative human PPCS are readily identified in other eukaryotes, including *Saccharomyces cerevisiae* (YIL083C) and *Arabidopsis thaliana* (hypothetical protein T7H20.130). The YIL083C gene was found to be essential for yeast viability in a systematic gene knockout study (Saccharomyces Genome Database) (see Ref. 20), providing additional support for this functional prediction. A comparison of eukaryotic (human and yeast), archaeal (*Methanococcus jannaschii*), and bacterial (*E. coli*) PPCS/PPCDC structural organizations is schematically presented in Fig. 2A.

At this stage of analysis, the only remaining gene “missing” in human CoA biosynthesis was PPAT. Previous biochemical analysis in rat and pig suggested the existence of a non-dissociable complex, potentially a bifunctional fusion protein, of PPAT and DPCK (14). As mentioned above, uncharacterized DPCK homologs can be found in the human genome and in all analyzed eukaryotic genomes. Based on the biochemical evidence of PPAT/DPCK fusion, additional searches in the human expressed sequence tag database were performed, revealing that one of the predicted human DPCK open reading frames (gi 13376838) was potentially 5'-truncated. A larger contiguous
Identification of Human CoA Biosynthetic Genes

Expression and Characterization of Human Recombinant CoA Biosynthetic Enzymes—The three human genes were overexpressed in E. coli, and the corresponding recombinant proteins were purified using standard affinity tag techniques. Human recombinant PPCDC is a bright yellow protein, with a UV spectrum having two maxima at 382 and 458 nm, characteristic of flavins. Although earlier results had classified mammalian PPCDC as a pyruvoyl-dependent enzyme (24), our results are consistent with those for the E. coli bifunctional PPCS/PPCDC protein, where the tight binding of FMN was directly proven by mass spectrometry (7) and led to the proposal of a novel redox mechanism of decarboxylation (25). Our data indicating approximately equimolar FMN binding in the evolutionarily distant human monofunctional PPCDC support the mechanistic importance of the FMN cofactor for the whole family.

Due to the structural dissimilarity between human PPAT/DPCK and the previously characterized bacterial PPAT and the implication that PPAT/DPCK is a point of regulation in CoA biosynthesis, we chose to study this protein in more depth. Kinetic parameters were obtained for both the forward PPAT reaction (4'-phosphopantetheine + ATP → dephospho-CoA + PP) and the reverse reaction (dephospho-CoA + PP, → 4'- phosphopantetheine + ATP) as well as the DPCK reaction (Table I). The PPAT reaction equilibrium state is established at roughly equimolar concentrations of all substrates and products, beginning from either direction, as revealed by HPLC analysis after extensive incubation with C-terminally truncated human protein depleted of DPCK activity. In the presence of full-length PPAT/DPCK, the adenylyltransferase reaction becomes irreversible due to the consecutive phosphorylation of dephospho-CoA and formation of the final product, CoA. All attempts to express an isolated DPCK domain failed to produce any soluble protein, and the kinase activity was characterized using the full-length PPAT/DPCK protein and dephospho-CoA as a substrate. Moreover, although the function of the N-terminal domain remains unknown, it is absolutely required for the proper expression and folding of the protein when overexpressed in E. coli.

The ultimate verification of the human CoA biosynthetic pathway was achieved by the successful in vitro reconstitution of the four-step biochemical transformation of the committed precursor 4'-phosphopantethenoyl CoA cofactor. The reaction mixture contained 4'-phosphopantethenoyl, cysteine, ATP, and 0.25–2.5 μg/ml concentrations of all three purified recombinant proteins: PPCS, PPCDC, and PPAT/DPCK. The conversion to CoA approached 100%, as monitored by HPLC (Fig. 4). When the bifunctional PPAT/DPCK enzyme was replaced with a C-terminally truncated enzyme with only PPAT activity, the reaction proceeded only to formation of dephospho-CoA. Incubation with only the first two enzymes quantitatively yields phosphopantethene as evidenced by AMP formation.

Northern Blot Analysis of Human PPAT/DPCK—We used Northern blot analysis of mRNAs from various human healthy
tissues and cancer cell lines to assess the possibility of PPAT/DPCK regulation at the level of transcription (Fig. 5). The overall level of transcription of PPAT/DPCK in normal tissues varied between different cell types within a range of ~10-fold (as normalized by actin), being lowest in peripheral blood leukocytes and highest in kidney and liver, supporting potential transcriptional regulation of PPAT/DPCK activity. One may also notice a general tendency of tumor cells to have a consistently higher overall transcriptional activity of this gene, such that all cancerous tissues examined had higher expression of the PPAT/DPCK mRNA than any of the normal tissues, with the exception of skeletal muscle, liver, and kidney. In examples where tissues may be compared, such as normal lung tissue versus lung carcinoma and normal colon tissue versus colorectal adenocarcinoma, the cancerous tissues expressed 2–4 times the PPAT/DPCK mRNA of the normal tissues.

Two PPAT/DPCK transcripts ~2200 and 2600 nucleotides long were detected in all examined tissues, with the possible exception of blood leukocytes. Both transcripts are significantly larger than the full-length coding region (1692 nucleotides). In healthy tissues, a mass distribution between the two transcripts varied from almost 1:1 in the majority of tissues with a relatively low overall expression level to an ~3 times predominance of the smaller transcript in kidney and liver and up to an ~2 times bias toward the larger transcript in skeletal muscle and placenta. In all tumor cell lines, a larger transcript was the clearly predominant and in some cases the only detectable form of PPAT/DPCK mRNA.

**DISCUSSION**

With the sequencing of the human genome complete, the next great challenge has become the identification of genes involved in key cellular pathways. On the basis of comparative genome analysis, we have identified human genes responsible for the last four enzymatic reactions in CoA biosynthesis. The biological functions of these human enzymes were verified by a high-throughput complementation technique based on transposon mutagenesis and genetic footprinting in *E. coli*. Capitalizing on the previously observed essentiality of *E. coli* genes involved in these last steps of CoA biosynthesis, we used the cloned human genes (cDNAs) to complement the loss of the endogenous enzymatic functions in *E. coli*. Introduction of the complementing human CoA biosynthetic genes into the expression vector made the corresponding *E. coli* genes nonessential as revealed by the appearance of transposon insertions. This approach can be systematically applied for preliminary functional analysis of uncharacterized biosynthetic genes in a number of pathways shared between *E. coli* and humans.

The individual human CoA biosynthetic enzymes were overexpressed in *E. coli* and purified, and the predicted activities were experimentally confirmed. Moreover, we have shown that CoA is quantitatively formed *in vitro* from the committed precursor 4′-phosphopantothenate, ATP, and L-cysteine by incubation with a mixture of all four human recombinant enzymes. Based on sequence similarity, human PPCS, PPCDC, and DPCK belong to the same families as recently described bacterial enzymes (6, 7, 9). Human PPAT reveals no sequence similarity to the bacterial CoaD family, establishing the existence of two structurally distinct families of nucleotidyltransferases performing adenylation of 4′-phosphopantotheine. The prediction and verification of a novel PPAT in humans allowed us to project this function to orthologs from other eukaryotic and archaeal genomes. Identification of human PPAT in the form of a fusion protein with DPCK is in agreement with previous biochemical data (14). The function of the third (N-terminal) domain of this protein is unclear. Based on our experimental data, this domain may be involved in protein folding. Additional roles in intracellular compartmentalization to the mitochondrial matrix as well as interactions with other CoA biosynthetic enzymes or regulatory factors may be hypothesized.

Due to the structural dissimilarity between human PPAT/DPCK and previously described bacterial enzymes and the implication of PPAT as a possible point of regulation in human CoA biosynthesis, we chose to study this enzyme in more depth. The two half-reactions of the PPAT/DPCK enzymes (transfer of the adenylyl moiety and consecutive phosphorylation) reveal comparable catalytic efficiencies as estimated by \( k_{\text{cat}}/K_{\text{m(app)}} \) suggesting that *in vitro* both steps are partially rate-limiting (Table I). Previous biochemical data suggested PPAT as a rate-limiting step based on the absence of dephospho-CoA accumulation *in vivo* (10). This may be reconciled as a potential consequence of the proximity of the two active sites in the bifunctional enzyme, supporting at least a partially non-dissociable mechanism (tunneling). Further kinetic and structural studies are required to assess this possibility. Alternatively, the additional monofunctional DPCK (gi 13623688), which was identified by our homology searches, may be responsible for increasing *in vivo* efficiency of dephospho-CoA to CoA conversion. Tentative identification of this monofunctional DPCK is in agreement with biochemical data indicating that although both PPAT and DPCK activities are associated with the mitochondrial matrix, only DPCK is present in the intermembrane space and in the outer membrane fraction (28).

To assess the possibility of regulation of the CoA biosynthetic pathway at the level of PPAT/DPCK gene expression, we have analyzed mRNA levels in both normal human tissues and human cancer cell lines using Northern blot hybridization (Fig. 5). The overall expression of the PPAT/DPCK transcript and the distribution between the larger and smaller transcripts varied significantly among the tissues analyzed. As we found only a single copy of a PPAT/DPCK gene in the human genome, the observed microheterogeneity may reflect alternative splicing. Although understanding the physiological role of such differences in PPAT/DPCK gene expression cannot be achieved without further studies, a role is suggested for this enzyme in regulating tissue-specific processes. Additionally, there is an overall tendency of tumor cells to have a greater abundance of PPAT/DPCK mRNA compared with normal tissues. Differential expression of the CoA biosynthetic genes may play a significant developmental role because the level of CoA consumption may vary significantly in various types of cells. One may expect higher production and consumption of CoA in rapidly growing cells, including tumor cells, where fatty acid synthase is known to be significantly up-regulated (see, for example, Ref. 27).

On the basis of our characterization of four enzymes in human CoA biosynthesis and the previously characterized mammalian pantothenate kinase (10) and human sodium-dependent pantothenate transporter (28), the minimal set of human genes required for the uptake and utilization of vitamin B₅ is presently known. Completion of the human CoA biosynthetic pathway will allow further analysis of its role in health and disease.

The conservation of this pathway in all three kingdoms of life allowed the efficient use of comparative genome analysis to reveal previously uncharacterized human genes. Fig. 2 illustrates remarkable patterns of “mosaicism” relating the evolutionary relationship within CoA biosynthetic genes in various taxa. These patterns reflect conservation, diversification, and multiple gene fusion and “unfusion” events, which may not be easily reconciled with the “tree of life,” such as built on the basis of 16 S RNA (29). For example, the first two enzymes (PPCS and PPCDC) in the four-step transformation of 4′-phos-
phopantothenate to CoA are consistently fused in bacteria (with the exception of streptococci and enterococci) and archaea, but “unfused” in all eukaryotes. Remarkably, one of these two components (PPCDC) is significantly more conserved between eukaryotes and prokaryotes than the other component (PPCS). A similar pattern is observed for the two last steps, with DPCK being significantly more conserved than PPAT. However, with respect to PPAT sequence conservation, archaea cluster with eukaryotes, and not with bacteria. In contrast to the PPCS/PPCDC case, a fusion event for the last two enzymes in the pathway appears to be a relatively modern invention, as it occurs in mammals and flies, but not in plants, fungi, or prokaryotes. Interestingly, in humans, which have both a bi-
functional PPAT/DPCK and a monofunctional DPCK, the latter reveals a higher sequence similarity to bacterial monofunc-
tional DPCK (P-score $\sim 10^{-42}$ with DPCK from *Bacillus subtilis*) than to the DPCK domain of human bifunctional PPAT/DPCK (P-score $\sim 10^{-19}$). The observed evolutionary mo-
saicism underscores the importance of using a diverse collection of organisms to increase the predictive power of comparative genome analysis. In particular, only monofunctional PPCS proteins present in a small group of bacteria (streptococci and enterococci) allowed the prediction of a human PPCS, whereas the sequence similarity to PPCS domains present in other bacterial genomes was insufficient to do so.

Despite significant variations in the sequences of individual enzymatic components, the overall topology of the universal CoA biosynthetic pathway appears to be preserved in all taxa-
ons. Previous biochemical data prompted the assertion that an alternate CoA biosynthetic pathway exists in *S. cerevisiae* (30). However, the identification of yeast orthologs for all of the human CoA biosynthetic enzymes described here, in combination with the essentiality of the corresponding *S. cerevisiae* genes, suggests that the CoA biosynthetic route in yeast is similar to that in *E. coli* and humans and that the former interpretation needs to be revised.

With the growing number of sequenced genomes, it is becom-
ing apparent that numerous pathways and individual compo-
ents of central metabolism, such as cofactor biosynthesis, amino acid biosynthesis, fatty acid metabolism, glycolysis, etc., have a remarkable tendency to be conserved across the three kingdoms. This tendency, along with the wealth of genomic sequencing data and advanced tools for comparative analysis, has significantly improved our ability to accurately identify entire biosynthetic pathways in complex organisms, as illustra-
ted here for human CoA biosynthesis.

Acknowledgments—We are infinitely indebted to Ross Overbeek and the ERGO development team at Integrated Genomics, Inc., who made this study possible and to the other colleagues in our laboratory, S. Chowdury, and S. Gerdes for technical and psychological support. We are grateful to T. Begley and E. Strauss (Cornell University) for inspiration, help with this manuscript, discussions, and sharing results on bacterial coaBC studies as well as for providing synthetic substrates. We thank D. Drueckhammer (State University of New York at Stony Brook) for discussions and sharing results on bacterial coaE. We thank P. Schimmel for help and advice on the manuscript. We are grateful to N. Grishin and H. Zhang (University of Texas Southwestern Medical Center) for help with distant sequence similarity analysis and to S. Mirkin and M. Krasninkova (University of Chicago) for helping with Northern blot experiments.

REFERENCES

1. Begley, T. P., Kinsland, C., and Strauss, E. (2001) *Vitam. Horm.* 61, 157–171
2. Tahiliani, A. G., and Beinlich, C. J. (1991) *Vitam. Horm.* 46, 165–228
3. Zhou, B., Westaway, S. K., Levinson, B., Johnson, M. A., Gitschier, J., and Hayflick, S. J. (2001) *Nat. Genet.* 28, 345–349
4. Afshar, K., Gonzalez, P., DiNardo, S., and Wasserman, S. A. (2001) *Genetics* 157, 1267–1276
5. Song, W. J., and Jackowski, S. (1992) *J. Bacteriol.* 174, 6411–6417
6. Kupke, T., Uebele, M., Schmid, D., Jung, G., Blassee, M., and Steinbacher, S. (2000) *J. Biol. Chem.* 275, 31838–31846
7. Strauss, E., Kinsland, C., Ge, Y., McAlferty, F. W., and Begley, T. P. (2001) *J. Biol. Chem.* 276, 13513–13516
8. Geerlof, A., Lewendon, A., and Shaw, W. V. (1999) *J. Biol. Chem.* 274, 27105–27111
9. Mishra, P., Park, P. K., and Drueckhammer, D. G. (2001) *J. Bacteriol.* 183, 2774–2778
10. Rock, C. O., Calder, R. B., Karim, M. A., and Jackowski, S. (2000) *J. Biol. Chem.* 275, 1377–1383
11. Calder, R. B., Williams, R. S., Ramaswamy, G., Rock, C. O., Campbell, E., Unkles, G. E., Kinghorn, T. J., and Jackowski, S. (1999) *J. Biol. Chem.* 274, 2014–2020
12. Daugherty, M., Volstein, V., Overbeek, R., and Osterman, A. (2001) *J. Bacteriol.* 183, 292–300
13. O’Brien, W. E. (1976) *Anal. Biochem.* 76, 423–430
14. Worrall, D. M., and Tubbs, P. K. (1983) *Biochem. J.* 215, 153–157
15. Ishikawa, Y. (1967) *J. Biochem. (Tokyo)* 61, 290–299
16. Ishikawa, Y. (1967) *J. Biochem. (Tokyo)* 61, 300–308
17. Kupke, T., Hernandez-Acosta, P., Steinbacher, S., and Culianez-Macia, F. A. (2001) *J. Biol. Chem.* 276, 19180–19186
18. Espinosa-Ruiz, A., Belles, J. M., Serrano, R., and Culianez-Macia, F. A. (1999) *Plant J.* 20, 529–539
19. Albert, A., Martinez-Ripoll, M., Espinosa-Ruiz, A., Yenush, L., Culianez-Macia, F. A., and Serrano, R. (2006) *Structure Fold Des.* 8, 961–969
20. Winkler, E. A., Shoemaker, D. D., Astone, M., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boo, J. D., Bussey, H., Chu, A. M., Connelly, C., Davis, K., Dietrich, F., Dow, S. W., El Bakhoury, M., Foy, F., Friend, S. H., Gentalen, E., Gierweger, G., Hegemann, J. H., Jones, T., Laub, M., Liao, H., Davis, R. W., et al. (1999) *Science* 285, 901–906
21. Venkatasalam, K. V., Fuda, H., Koonin, E. V., and Strotz, C. A. (1999) *J. Biol. Chem.* 274, 2921–2934
22. Bork, P., Holm, L., Koonin, E. V., and Sander, C. (1995) *Proteins Struct. Funct. Genet.* 22, 259–266
23. Izard, T., and Geerlof, A. (1999) *EMBO J.* 18, 2021–2030
24. Scandurra, R., Morigg, M., Consalvi, V., and Politi, L. (1979) *Methods Enzymol.* 62, 245–248
25. Strauss, E., and Begley, T. P. (2001) *J. Am. Chem. Soc.* 123, 6449–6450
26. Skrede, S., and Halvorsen, O. (1988) *Eur. J. Biochem.* 171, 57–63
27. Kuhajda, F. P., Pizer, E. S., Li, J. N., Mani, N. S., Frehywot, G. L., and Townsend, C. A. (2000) *Proc. Natl. Acad. Sci. U. S. A.* 97, 3450–3454
28. Wang, H., Huang, W., Fei, Y. J., Xia, H., Yang, F. T., Lebach, F. H., Devos, L. D., Ganapathy, V., and Prasad, P. D. (1999) *J. Biol. Chem.* 274, 14875–14883
29. Olsen, G. J., Wise, C. R., and Overbeek, R. (1994) *J. Bacteriol.* 176, 1–6
30. Bucovac, E. T., Turnovski, S. J., Morrison, W. C., Macleod, R. M., Morrison, J. C., Sobhy, C. M., Rhoades, J. L., Fryer, J. E., Wakim, J. M., and Whybrew, W. D. (1980) *Mol. Cell. Biochem.* 30, 7–26
