

# Glutathione synthetase homologs encode $\alpha$ -L-glutamate ligases for methanogenic coenzyme $F_{420}$ and tetrahydrosarcinapterin biosyntheses

Hong Li, Huimin Xu, David E. Graham, and Robert H. White\*

Department of Biochemistry (0308), Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

Edited by Stephen J. Benkovic, Pennsylvania State University, University Park, PA, and approved June 30, 2003 (received for review June 4, 2003)

Proteins in the ATP-grasp superfamily of amide bond-forming ligases have evolved to function in a number of unrelated biosynthetic pathways. Previously identified homologs encoding glutathione synthetase, D-alanine:D-alanine ligase and the bacterial ribosomal protein S6:glutamate ligase have been vertically inherited within certain organismal lineages. Although members of this specificity-diverse superfamily share a common reaction mechanism, the nonoverlapping set of amino acid and peptide substrates recognized by each family provided few clues as to their evolutionary history. Two members of this family have been identified in the hyperthermophilic marine archaeon *Methanococcus jannaschii* and shown to catalyze the final reactions in two coenzyme biosynthetic pathways. The MJ0620 (*mptN*) locus encodes a tetrahydromethanopterins: $\alpha$ -L-glutamate ligase that forms tetrahydrosarcinapterin, a single carbon-carrying coenzyme. The MJ1001 (*cofF*) locus encodes a  $\gamma$ - $F_{420}$ -2: $\alpha$ -L-glutamate ligase, which caps the  $\gamma$ -glutamyl tail of the hydride carrier coenzyme  $F_{420}$ . These two genes share a common ancestor with the ribosomal protein S6:glutamate ligase and a putative  $\alpha$ -amino adipate ligase, defining the first group of ATP-grasp enzymes with a shared amino acid substrate specificity. As in glutathione biosynthesis, two unrelated amino acid ligases catalyze sequential reactions in coenzyme  $F_{420}$  polyglutamate formation: a  $\gamma$ -glutamyl ligase adds 1–3 L-glutamate residues and the ATP-grasp-type ligase described here caps the chain with a single  $\alpha$ -linked L-glutamate residue. The analogous pathways for glutathione,  $F_{420}$ , folate, and murein peptide biosyntheses illustrate convergent evolution of nonribosomal peptide biosynthesis through the recruitment of single-step amino acid ligases.

Organisms that carry out aerobic respiration or oxygenic photosynthesis produce significant amounts of thiol-containing cofactors in response to oxidative stress. Most eukaryotes, cyanobacteria, and  $\alpha$ ,  $\beta$ ,  $\gamma$ -proteobacteria synthesize the tripeptide glutathione for this purpose (1). Yet the evolutionary history of glutathione biosynthesis is convoluted (2), and other aerobic microorganisms use different thiol cofactors to resist oxidative damage. Low % G+C Gram-positive bacteria produce excess amounts of CoA for this purpose, whereas high % G+C Gram-positive bacteria produce mycothiol or ergothioneine (3). Methanogenic archaea contain significant amounts of coenzyme M, a thiol cofactor that like glutathione has been recruited into pathways to degrade oxidized metabolites (4). To understand how glutathione evolved, it will be necessary to discern how the peptide biosynthetic enzymes  $\gamma$ -glutamylcysteine synthetase ( $\gamma$ -L-glutamyl:L-cysteine ligase) and glutathione synthetase ( $\gamma$ -L-glutamyl-L-cysteine:glycine ligase) evolved from separate ligase families (2, 5).

Two unrelated peptide-bond forming enzymes are required to produce the tripeptide coenzyme glutathione.  $\gamma$ -Glutamylcysteine synthetase catalyzes peptide bond formation between the  $\gamma$ -carboxyl of L-glutamate and the amino group of L-cysteine by using a  $\gamma$ -glutamylphosphate intermediate (6). Sequence alignments suggest that this protein's structure is similar to that of the C-terminal (kinase) domain of glutamine synthetase (7). Next in

the pathway, glutathione synthetase catalyzes the ligation of glycine to  $\gamma$ -glutamylcysteine, again via an acyl phosphate intermediate. Homologs of bacterial glutathione synthetase include D-alanine:D-alanine ligase (DdlB; ref. 8), the highly diverged and circularly permuted eukaryotic glutathione synthetase (GSH2; ref. 9), the bacterial ribosomal protein S6:glutamate ligase (RimK; ref. 10), and a putative  $\alpha$ -amino adipate ligase that could be involved in lysine biosynthesis (LysX; ref. 11). Although these homologs share little sequence similarity, all belong to the ATP-grasp superfamily of enzymes that share a noncanonical nucleotide-binding fold (including several flexible loops), and each has a common reaction mechanism involving an activated acyl phosphate intermediate (12, 13). Each member of this superfamily specifically recognizes only a narrow range of physiologically relevant substrates. This breadth of possible substrates confounds attempts to assign functions to homologs identified in genome sequences.

The complete genome sequence of the hyperthermophilic euryarchaeon *Methanococcus jannaschii* contains two homologs of the *Escherichia coli* rimK gene, which encodes an ATP-grasp enzyme-ribosomal protein S6:glutamate ligase (14). In *E. coli*, this RimK protein catalyzes the posttranslational addition of up to four glutamate residues to the carboxyl terminus of ribosomal protein S6 (10). *M. jannaschii*, like other archaea, has no homolog of the bacterial ribosomal protein S6 gene. Nor do the conserved residues at the C terminus of any *M. jannaschii* ribosomal protein resemble those recognized by RimK in *E. coli* ribosomal protein S6 (10). Therefore, we predicted that the proteins encoded by the *M. jannaschii* rimK homologs catalyze different amino acid ligation reactions, by using the conserved mechanism of enzymes in the ATP-grasp superfamily. The *Methanococcales* contain neither glutathione nor the peptide-linked murein cell wall produced by other ATP-grasp members (1). Orthologs of the two *M. jannaschii* genes are found only in the archaeal orders *Methanococcales* and *Methanosarcinales*, suggesting that the enzymes catalyze peptide bond formation in molecules unique to organisms in those lineages.

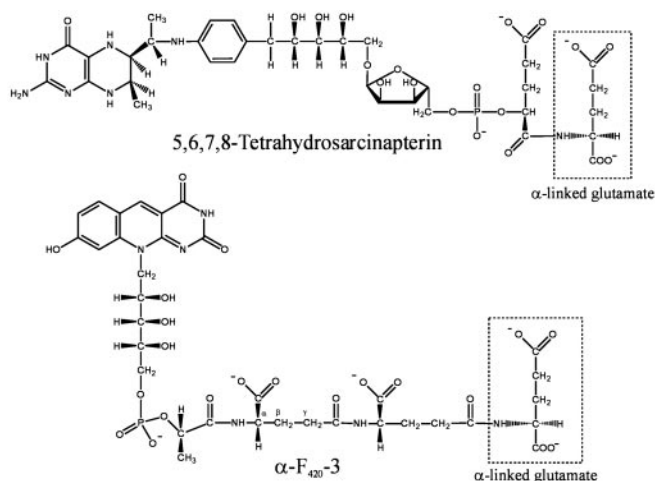
Four coenzymes from *M. jannaschii* contain amide-bonded amino acids: methanofuran, 5,6,7,8-tetrahydrosarcinapterin, coenzyme  $F_{420}$ , and coenzyme B (15). Both methanofuran and  $F_{420}$  contain  $\gamma$ -linked L-glutamate residues. We have recently identified the enzyme that catalyzes the addition of up to two  $\gamma$ -linked glutamates to  $F_{420}$ -0, forming  $\gamma$ - $F_{420}$ -2 (42). However, sarcinapterin and  $\alpha$ - $F_{420}$ -3 isolated from *M. jannaschii* cells contain terminal  $\alpha$ -linked glutamate residues (Fig. 1; refs. 16 and 17). To test whether the two *M. jannaschii* rimK homologs encode coenzyme:glutamate ligases, we have cloned and heterologously expressed both genes in *E. coli*.

Here, we report that purified protein expressed from the gene at *M. jannaschii* locus MJ1001 catalyzes the ATP-dependent

This paper was submitted directly (Track II) to the PNAS office.

Abbreviation: MPT, methanopterins.

\*To whom correspondence should be addressed. E-mail: rhwhite@vt.edu.



**Fig. 1.** Structures of the 5,6,7,8-tetrahydrosarcinapterin and  $\alpha$ -F<sub>420</sub>-3 coenzyme products of two homologous *M. jannaschii* enzymes that add terminal  $\alpha$ -linked glutamates. The MptN enzyme is a tetrahydromethanopterin: $\alpha$ -glutamate ligase that produces 5,6,7,8-tetrahydrosarcinapterin. The Coff enzyme is a  $\gamma$ -F<sub>420</sub>-2: $\alpha$ -L-glutamate ligase that produces coenzyme  $\alpha$ -F<sub>420</sub>-3.

**ligation of coenzyme  $\gamma$ -F<sub>420</sub>-2 and L-glutamate to form  $\alpha$ -F<sub>420</sub>-3.** As in glutathione biosynthesis, two unrelated amino acid ligases in F<sub>420</sub> biosynthesis sequentially form  $\gamma$ -glutamate and  $\alpha$ -carboxyl peptide bonds. The paralogous gene from *M. jannaschii* locus MJ0620 catalyzes the ATP or GTP-dependent ligation of tetrahydromethanopterin and L-glutamate to form tetrahydrosarcinapterin. A phylogeny inferred from glutathione synthetase homologs identifies orthologs of both genes in the partial genome sequence of *Methanococcus maripaludis* (which contains  $\alpha$ -F<sub>420</sub>-3 and sarcinapterin) and orthologs of the MJ0620 gene in members of the *Methanosarcinales*, which are known to produce sarcinapterin. Furthermore, this phylogeny suggests that all of the ATP-grasp enzymes that add  $\alpha$ -linked dicarboxylic amino acids share a common ancestor. The limited phylogenetic distribution of most members of the ATP-grasp amino acid ligase families indicates that the genes have been frequently recruited to catalyze analogous reactions in new pathways.

## Materials and Methods

**Preparation of Cell-Free Extracts.** Frozen cells of *Methanobacterium thermoautotrophicum* strain  $\Delta$ H (*Methanothermobacter thermoautotrophicus*) grown in complex medium containing yeast extract and tryptone were kindly supplied by R. S. Wolfe (University of Illinois, Urbana). Frozen cells of *M. maripaludis* strain S2 (LL) were generously supplied by W. B. Whitman (University of Georgia, Athens). Frozen cells of *M. jannaschii* JAL-1 (*Methanocaldococcus jannaschii*) were purchased from the University of Illinois Fermentation Facility (Urbana). Acetate-grown, frozen *Methanosarcina barkeri* cells were a generous gift from D. A. Grahame (Uniformed Services University of the Health Sciences, Bethesda).

Cells were lysed by sonication under argon in buffer containing 50 mM *N*-Tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid (TES)/NaOH, 10 mM MgCl<sub>2</sub> and 20 mM  $\beta$ -mercaptoethanol (pH 7.0). Crude extracts were stored under argon at  $-70^{\circ}\text{C}$ . The total protein concentrations of cell extracts ranged from 20 to 40 mg/ml, determined by the bicinchoninic acid (BCA) protein assay (Pierce) using BSA as a standard.

**Generation of Substrates and Reference Compounds.** The substrates and products used for this work were purified from frozen cells or cell extracts. Samples of methanopterin (MPT),  $\gamma$ -F<sub>420</sub>-2, and

methaniline were prepared by the aerobic extraction of *Methanobacterium thermoautotrophicum* cells with 60% methanol/water whereas sarcinapterin was obtained by extraction of *M. jannaschii* cells with the same solvent. In a typical extraction, 250 mg of cell pellet was mixed with 1.5 ml of 60% methanol/water and heated at  $100^{\circ}\text{C}$  for 5 min in a sealed tube. After centrifugation ( $14,000 \times g$ , 5 min at room temperature) the clear liquid was separated, and the pellet was reextracted by the same procedure. After evaporation of the solvent, the desired compounds were isolated by different chromatographic procedures. Samples of F<sub>420</sub> and MPT were obtained by chromatography of the extracts on a DEAE column, eluted with an ammonium bicarbonate gradient (18). F<sub>420</sub> isolated by this procedure was analyzed by atmospheric pressure chemical ionization MS operated in negative ion mode. This technique produced an (M-H)<sup>−</sup> ion at 772.2 *m/z* corresponding to the expected molecular formula for F<sub>420</sub>-2 of C<sub>29</sub>H<sub>36</sub>O<sub>18</sub>N<sub>5</sub>P. MPT was isolated from these extracts by preparative TLC of the extracts on silica gel 60 F<sub>254</sub> plates (Merck) by using acetonitrile/water/formic acid (88%) (40:10:5, vol/vol/vol) as the developing solvent. In this solvent system, MPT and sarcinapterin had *R<sub>f</sub>* values of 0.059 and 0.094, respectively. MPT and sarcinapterin were identified by their fluorescence and UV absorbances on the TLC plate. The isolated pterin compounds had maximal UV absorbance at 339 nm in water.

Tetrahydromethanopterin (H<sub>4</sub>MPT) was isolated from *Methanobacterium thermoautotrophicum* cells extracted under anaerobic conditions with methanol/water. Thus, 0.5 g of cell pellet under argon was mixed with 1 ml of anaerobic 60% methanol/water and heated at  $100^{\circ}\text{C}$  for 5 min. After centrifugation to remove insoluble material ( $14,000 \times g$ , 10 min at room temperature), the extract was evaporated to dryness with a stream of nitrogen gas, and the resulting residue was dissolved in the 100  $\mu$ l of anaerobic buffer containing 50 mM TES/Na<sup>+</sup>, 10 mM MgCl<sub>2</sub>, and 20 mM DTT. To determine the concentration of tetrahydromethanopterin in the extracts, portions of these extracts were oxidized with I<sub>2</sub> and cleaved with Zn/HCl, and the Bratton-Marshall azo dye products were formed as described (19). These extracts were found to contain 0.77 mM H<sub>4</sub>MPT based on a previously reported extinction coefficient for azo dye derivatives of 48,000 M<sup>−1</sup>·cm<sup>−1</sup> at 562 nm (19).

Methaniline or  $\alpha$ -glutamylmethaniline were obtained from extracts containing MPT or sarcinapterin by reductive cleavage with Zn/HCl (19). These compounds were purified by using the preparative TLC procedure described above, and they were observed by spraying a portion of the TLC plate with 1,2-naphthoquinone-4-sulfonic acid in dilute acetic acid followed by heating (20). Methaniline was identified by the development of a pink color corresponding to a spot with an *R<sub>f</sub>* of 0.17 in this TLC solvent. In each case, the TLC support was scraped from the TLC plate, and the desired compound was recovered by elution with 50% methanol in water. Samples of  $\alpha$ -glutamylmethaniline were prepared from the tetrahydrosarcinapterin present in anaerobically prepared cell extracts of *M. jannaschii* and *Methanosarcina barkeri* by precipitation of the proteins with methanol, oxidation with I<sub>2</sub>, and reductive cleavage by Zn/HCl (19).

**Cloning and Expression of the *M. jannaschii* MJ0620 and MJ1001 Genes.** The *M. jannaschii* genes at loci MJ0620 and MJ1001 (Swiss-Prot accession numbers Q58037 and Q58407, respectively) were amplified by PCR from genomic DNA by using oligonucleotide primers (Invitrogen). The MJ0620 gene was amplified by using the primers MJ0620-Fwd (5'-GGTCA-TATGAACTTGGTATAATTACC-3') and MJ0620-Rev (5'-GCTGGATCCTTAAGATTAAAC-3'). The primers for MJ1001 were MJ1001-Fwd (5'-GGTCATATGGTAAATAACC-3') and MJ1001-Rev (5'-GCTGGATCCTCAG-GACTTTGC-3'). The translational start site chosen for the



recombinant MJ1001 gene was located six nucleotides downstream from the previously annotated start site of MJ1001 (14), which overlapped with a putative ribosome-binding site. PCR amplifications were performed as described (21) by using a 50°C annealing temperature for MJ0620 and MJ1001. PCR products were purified by using a QIAquick spin column (Qiagen, Valencia, CA) and digested with *Nde*I and *Bam*HI enzymes and then ligated into compatible sites in plasmid pT7-7 to make the recombinant plasmids pMJ0620 and pMJ1001. DNA sequences were verified by dye-terminator sequencing at the University of Iowa DNA facility. The resulting plasmids were transformed into *E. coli* strain BL21-Codon Plus (DE3)-RIL (Stratagene). Heterologous expression of the MJ0620 and MJ1001 proteins was performed by standard methods (21).

**Analysis of Protein Expression and Protein Purification.** *E. coli* cells containing the heterologously expressed proteins were sonicated, and the insoluble material was removed by centrifugation (21). The apparent masses of the expressed, denatured proteins were determined by comparing protein migration with the migration of low molecular weight protein standards (Bio-Rad) separated by electrophoresis on an SDS/polyacrylamide gel (12% T, 2.7% C acrylamide) with a Tris/glycine buffer system. SDS/PAGE analyses of the whole cells as well as the sonicated extracts were used to establish the extent of protein expression and the solubility of the heterologously expressed protein. MJ0620 and MJ1001 proteins were purified by heating, anion-exchange chromatography, and gel-filtration chromatography (22). Analytical gel-filtration chromatography was performed as described (22).

**Analysis of MJ1001 Reaction Products and Enzymatic Activities.** The purified MJ1001 protein was tested for  $\gamma$ -F<sub>420</sub>-2:glutamate ligase activity in a standard assay. MJ1001 protein (1  $\mu$ g) was incubated with  $\gamma$ -F<sub>420</sub>-2 (0.5  $\mu$ M) in a 50- $\mu$ l reaction mixture containing 50 mM 2-(*N*-cyclohexylamino)ethanesulfonic acid/Na<sup>+</sup> buffer (pH 8.5), 0.2 M KCl, 10 mM L-glutamate, 5 mM ATP, 2 mM DTT, 5 mM MgCl<sub>2</sub>, and 5 mM MnCl<sub>2</sub> at 50°C for 20 min. The reaction was initiated by adding  $\gamma$ -F<sub>420</sub>-2 and terminated by the addition of EDTA to a final concentration of 10 mM. Reaction products were analyzed and quantified by using a Shimadzu SCL-6B HPLC with a C-18 reversed phase column (AXXI-Chrom octyldecyl silane column, 5  $\mu$ m, 4.6-mm internal diameter). Chromatography was performed by using an isocratic elution with 25 mM sodium acetate (pH 6.0), 0.02% (wt/vol) sodium azide, and different concentrations of methanol for each substrate and product pair. Eluting compounds were detected by using an RF-551 fluorescence HPLC monitor (Shimadzu).

The glutamate linkage in the modified F<sub>420</sub> was examined by enzymatic hydrolysis. The MJ1001 reaction mixture (2  $\mu$ l) was treated with 5  $\mu$ l of a solution of carboxypeptidase Y (baker's yeast; Sigma; 16 units/100  $\mu$ l in 50 mM TES/Na<sup>+</sup>, pH 7.5, 10 mM MgCl<sub>2</sub>) or 5  $\mu$ l of carboxypeptidase G (*Pseudomonas* sp., Sigma; 5 units/100  $\mu$ l in 50 mM TES/Na<sup>+</sup>, pH 7.5, 10 mM MgCl<sub>2</sub>) for 2 h at room temperature. The products were separated by HPLC and identified by their retention times compared with known F<sub>420</sub> compounds (17).

**Analysis of MJ0620 Reaction Products and Enzymatic Activities.** The purified MJ0620 protein was tested for glutamate ligase activity in a standard assay. MJ0620 protein (12  $\mu$ g) was incubated with methaniline, MPT (27  $\mu$ M), or H<sub>4</sub>MPT (86  $\mu$ M) in a 100- $\mu$ l reaction mixture containing 50 mM TES/Na<sup>+</sup> buffer (pH 7.0), 0.11 M KCl, 5.5 mM L-glutamate, 5.5 mM ATP or GTP, 5 mM DTT, and 10 mM MgCl<sub>2</sub>, at 60°C for 30 min. At the end of the reaction period, different procedures were used to derivatize the products, which were analyzed by HPLC as described above. In

the case of methaniline and  $\alpha$ -glutamylmethaniline, either the arylamines were converted into their fluorescamine derivatives (23) and monitored by fluorescence (407 nm excitation, 519 nm emission) or the products were converted into their azo-dye derivatives and purified by Biogel P-4 chromatography before HPLC (19). Eluting methanopterin and sarcinapterin were detected by their fluorescence (355 nm excitation, 440 nm emission). The H<sub>4</sub>MPT samples were cleaved before formation of the Bratton-Marshall azo-dye adducts and followed by absorbance at 462 nm by using an SPD-6AV UV-Vis spectrophotometric detector (Shimadzu). In each case, the amount of product was measured based on the ratio of starting material to product after separation by HPLC.

**Phylogeny of Glutathione Synthetase Homologs.** The translated sequences of the *M. jannaschii* MJ0620 and MJ1001 genes were used to query the nonredundant protein database at the National Center for Biotechnology Information by using the BLASTP program (Version 2.2.6; ref. 24) with the BLOSUM62 matrix and default gap costs for existence-11 and extension-1. Additional homologs were identified by using similar methods to search partial genome sequences. A subset of homologs was assembled by choosing representative orthologs of previously identified enzymes and including all recognizable homologs from *M. jannaschii* and *M. maripaludis*. Orthologs of the bacterial D-alanine:D-alanine ligase were included as an out-group. Protein sequences used in this analysis (and their database accession numbers) were from *Aquifex aeolicus* (DdlB, sp|O66806), *Bacillus subtilis* (DdlB, sp|P96612), *E. coli* (DdlB, sp|P07862; GshB, sp|P04425; RimK, sp|P17116), human (ref|XP\_048675.2), *Methanococcoides burtoni* (www.jgi.doe.gov/), *M. maripaludis* (www.genome.washington.edu/UWGC/methanococcus/), *Methanosarcina acetivorans* (MptN, gb|AAM06639.1; gb|AAM04777.1), *Methanosarcina mazei* (MptN, gb|AAM29807.1), *Pyrococcus furiosus* (LysX, NP\_579411), *Sulfolobus solfataricus* (LysX, tpg|DAA00053.1), *Synechocystis* sp. PCC6803 (GshB, sp|P73493), *Takifugu rubripes* (SINFRUP00000085858), *Thermus thermophilus* (LysX, dbj|BAC67243.1), and *Trichodesmium erythraeum* (RimK, gb|ZP\_00073799.1).

These diverged homologs were aligned piecemeal to minimize alignment artifacts. Three D-alanine:D-alanine ligase protein sequences and 19 glutathione synthetase homologs were aligned separately by using the CLUSTALW program (Version 1.83; ref. 25), and then the alignments were combined manually, guided by a structural alignment of *E. coli* proteins GshB (PDB ID code 1gsa) and DdlB (PDB ID code 1iow) from the Dali server (26). From the full alignment of 22 sequences, 286 aa positions that were deemed to be confidently aligned were selected for phylogenetic analysis.

The phylogeny of the glutathione synthetase homologs was inferred by protein distance and maximum likelihood methods. The protein distance method used the PROTDIST and NEIGHBOR programs (Version 3.6a3; ref. 27) with the Jones, Taylor, and Thornton model of amino acid changes and a  $\gamma$  distribution of positional rates of change (coefficient of variation = 0.64). Bootstrap proportions were calculated by using SEQBOOT, PROTDIST, NEIGHBOR, and CONSENSE programs (27) to create and evaluate 500 resampled alignments. An alternative tree was identified by the protein maximum likelihood criteria by using the PROML program (version 3.6a3; ref. 27). The PROML program was used with the same parameters used for the PROTDIST program.

***M. maripaludis* Cofactor Analysis.** To test whether the F<sub>420</sub> and pterin coenzymes from *M. maripaludis* contain  $\alpha$ -linked glutamate(s), coenzymes were extracted from *M. maripaludis* cells (0.6 g wet weight) by heating for 10 min at 100°C with 1.0 ml of

methanol. After centrifugation ( $14,000 \times g$ , at room temperature), the supernatant of the solution was evaporated, and the residue was dissolved in deionized water.  $F_{420}$  species were separated and identified by HPLC, as described (17). The glutamate linkage was tested by enzymatic hydrolysis, described above. The sarcinapterin present in the extract was cleaved with  $Zn/HCl$ , and the resulting arylamine was converted into the azo dye derivative and analyzed as described above.

## Results

### *rimK* Homologs in *M. jannaschii* Are Expressed as Monomers or Dimers.

The two genes from *M. jannaschii* that were previously annotated as “ribosomal protein S6 modification proteins” were cloned into plasmids in *E. coli*. Heterologously expressed proteins from these genes (*M. jannaschii* loci MJ0620 and MJ1001) were thermostable and highly soluble in *E. coli* cell-free extract. A three-step purification scheme produced protein preparations that each showed single predominant bands when analyzed by SDS/PAGE. The denatured MJ0620 protein migrated with an apparent molecular mass of 35 kDa, close to its predicted molecular mass of 33,278 Da. The native MJ0620 protein eluted with an apparent molecular mass of 72 kDa from an analytical size-exclusion column, consistent with this protein forming a homodimer. The denatured MJ1001 protein migrated with an apparent molecular mass of 35 kDa during SDS/PAGE, which is close to its predicted molecular mass of 33,112 Da. When loaded on an analytical size-exclusion column, MJ1001 protein eluted with an apparent molecular mass of 37 kDa and therefore forms monomers.

**CofF Protein Is a  $\gamma$ - $F_{420}$ -2: $\alpha$ -L-Glutamate Ligase.** When incubated with  $F_{420}$ -2, L-glutamate, and ATP, the MJ1001 protein catalyzed the addition of one L-glutamate molecule to  $\gamma$ - $F_{420}$ -2 to form  $\alpha$ - $F_{420}$ -3 (Fig. 1). HPLC analysis showed that this  $F_{420}$ -3 product coelutes with  $\alpha$ - $F_{420}$ -3, the predominant form of  $F_{420}$  isolated from *M. jannaschii* (17). Carboxypeptidase Y (a peptidyl-L-amino acid hydrolase that specifically removes carboxyl-terminal  $\alpha$ -linked amino acids; ref. 28) cleaved this  $F_{420}$ -3 product, regenerating the  $\gamma$ - $F_{420}$ -2 substrate. In contrast, carboxypeptidase G (an exopeptidase that releases terminal L-glutamate residues from folate and  $\gamma$ - $F_{420}$ -2; refs. 17 and 29) did not cleave the  $F_{420}$ -3 product. Therefore, the product of CofF is identical to the  $\alpha$ - $F_{420}$ -3 species that has been identified in *M. jannaschii* (17). No  $F_{420}$ -3 compound or other products were detected in the reaction mixture when MJ0620 protein was incubated with  $\gamma$ - $F_{420}$ -2 under the same conditions as with MJ1001. Because the MJ1001 protein specifically adds an  $\alpha$ -linked glutamate to  $\gamma$ - $F_{420}$ -2 produced by the *M. jannaschii* CofE protein (42), this enzyme will be designated the CofF protein.

Without 2 mM DTT or 2-mercaptoethanol in the reaction mixture, CofF activity was up to 5-fold lower. The CofF protein strictly required a divalent metal ion for ligase activity. Similar to the CofE  $F_{420}$ -0: $\gamma$ -glutamyl ligase, CofF enzyme incubated with 10 mM  $MgCl_2$  had  $\approx 33\%$  of the activity supported by 10 mM  $MnCl_2$ . But unlike CofE, CofF activity is inhibited by KCl. In the presence of 0.1 or 0.2 M KCl, activities were 40% and 15% relative to reactions without KCl. Initial reaction rates for  $F_{420}$ -2: $\alpha$ -glutamate ligase activity were measured in standard assays at various concentrations of  $F_{420}$ -2 (0–2  $\mu M$ ), and activity data were fit to the Michaelis-Menten pseudofirst-order rate equation. The apparent kinetic parameters for CofF were  $K_m = 1.2 \pm 0.3 \mu M$   $F_{420}$ -2, and  $V_{max} = 0.72 \pm 0.1$  nmol/min per mg.

The CofF protein preferred ATP to other purine nucleotide triphosphates: GTP (5 mM) gave  $\approx 25\%$  of the activity observed with 5 mM ATP. The phosphonate nucleotide analogs  $\alpha,\beta$ - $CH_2$ -ATP and  $\beta,\gamma$ - $CH_2$ -ATP supported no ligase activity at concen-

trations of 5 mM. However, when added to a reaction mixture containing an equal amount of ATP,  $\alpha,\beta$ - $CH_2$ -ATP inhibited 30% of the activity whereas no activity was observed with  $\beta,\gamma$ - $CH_2$ -ATP. These results supported a phosphoryl transfer rather than nucleotidyl transfer mechanism for the enzyme, similar to other members of the glutathione synthetase family. None of the following amino acids or analogs supported ligase activity at 10-mM concentrations: D-glutamate,  $\beta$ -glutamate, L-aspartate, L-glutamine, L- $\alpha$ -amino adipate, or D,L-2-amino-4-phosphono-butyrates. These data suggest that CofF is an L-glutamate-dependent ligase.

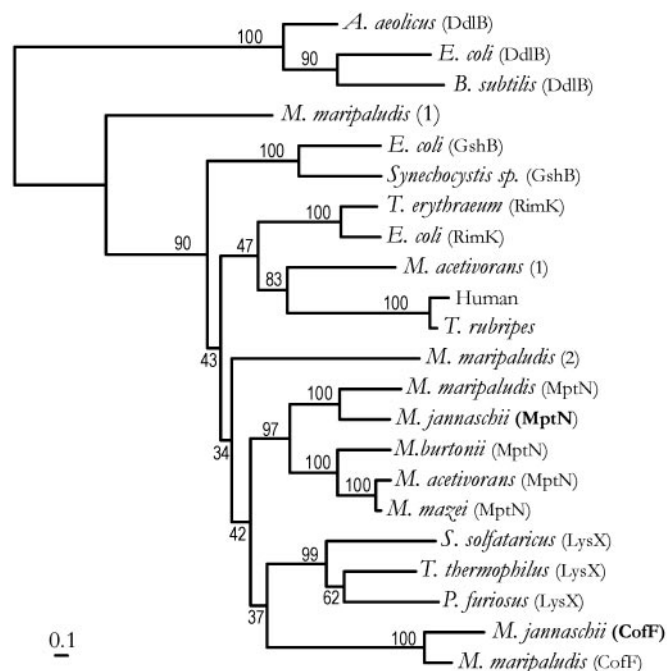
**MptN Protein Is an  $H_4$ MPT: $\alpha$ -L-Glutamate Ligase.** Incubations of the MJ0620 protein with L-glutamate, ATP, or GTP and either methaniline or MPT produced no  $\alpha$ -glutamylmethaniline or sarcinapterin. Incubation with  $H_4$ MPT readily produced tetrahydrosarcinapterin (Fig. 1), which was assayed as the azo dye derivative of  $\alpha$ -glutamylmethaniline after oxidation and reductive cleavage. Because the MJ0620 protein specifically adds an  $\alpha$ -linked glutamate to MPT, which is produced by the Mpt proteins from *M. jannaschii* (15), this enzyme is designated the MptN protein. Under rate-limiting assay conditions, MptN produced 12.7 nmol of sarcinapterin per min per mg of protein in the presence of GTP and  $K^+$ . The enzyme did not discriminate between ATP and GTP: with ATP, the enzyme had a specific activity of 12.0 nmol/min per mg of protein. Neither did MptN require  $K^+$ , as its specific activity without potassium was 11.2 nmol/min per mg of protein.

### Evolution of Glutamate Ligases in the Glutathione Synthetase Family.

The phylogeny of members of the glutathione synthetase family was inferred from an alignment of 22 homologs (Fig. 2). Because the *ddlB* gene encoding D-alanine:D-alanine ligase seems to have been vertically inherited in most bacterial lineages and was probably required by bacterial ancestors to produce a murein sacculus, DdlB protein sequences were used to root the tree. Tertiary structure information from crystal structure models of the *E. coli* GshB and DdlB proteins was used to align the sequences of these highly diverged proteins ( $<20\%$  amino acid sequence identity; refs. 8 and 26). All members of this family share a conserved Mg-ATP-binding domain that is believed to undergo a conformational change on substrate binding (30).

Phylogenetic trees that were inferred by protein distance (Neighbor-Joining) and maximum likelihood methods have almost identical topologies (Fig. 2). They differ only in the placement of one unidentified gene from *M. maripaludis*. The quality of the sequence alignment had a much greater effect on the inferred phylogeny (results not shown). Bootstrap replicates strongly support the clustering of orthologous genes, although many internal nodes lack significant support. Therefore, the relationships among most individual gene clusters remain unresolved. Nevertheless, homologs that catalyze peptide bond formation with an  $\alpha$ -linked glutamate residue (RimK, MptN, CofF) or  $\alpha$ -amino adipate (proposed for LysX) do group together. Recognition and transfer of an  $\alpha$ -amino dicarboxylate may be an ancestral property of these enzymes.

***M. maripaludis* Contains  $\alpha$ - $F_{420}$ -3 and Sarcinapterin.** The incomplete genome sequence of *M. maripaludis* contains genes that are orthologs of the *M. jannaschii* *cofF* and *mptN* genes (Fig. 2). If these genes are expressed, then that organism's  $F_{420}$  and methanopterin-related cofactors should have  $\alpha$ -linked glutamate modifications similar to those observed in *M. jannaschii*. HPLC analysis of  $F_{420}$  extracted from *M. maripaludis* cells showed three major fluorescent peaks (% of total fluorescence):  $F_{420}$ -2 ( $\gamma$ -linked, 4.86%),  $F_{420}$ -3 ( $\alpha$ -linkage for the third glutamate, 89.9%), and  $F_{420}$ -4 ( $\alpha$ -linkage for the fourth glutamate, 5.2%).



**Fig. 2.** Phylogeny of glutathione synthetase homologs inferred by the neighbor-joining method. Sequences (described in *Materials and Methods*) represent groups of amino acid ligases including glutathione synthetase (GshB), D-alanine:D-alanine ligases (DdlB), ribosomal protein S6 modification proteins (RimK), tetrahydromethanopterin: $\alpha$ -glutamate ligases (MptN), putative  $\alpha$ -amino acid ligases (LysX), and  $\gamma$ -F<sub>420</sub>:2: $\alpha$ -glutamate ligases (CofF). Uncharacterized paralogs from several methanogens are distinguished by numbers in parentheses after their species name. The tree was rooted by using the DdlB sequences as an outgroup. Bootstrap percentages are indicated for branches supported by a plurality of bootstrap replicates. (Bar = 0.1 aa replacements per position.)

The glutamyl linkages of these F<sub>420</sub> compounds were confirmed by hydrolysis with carboxypeptidases. This distribution of F<sub>420</sub> species is very similar to that of *M. jannaschii* (17). Preparations of MPT-related cofactors from *M. maripaludis* confirmed that that organism contains tetrahydrosarcinapterin, as observed in *M. jannaschii*.

## Discussion

Most folates have  $\gamma$ -linked L-glutamate tails. Although these poly- $\gamma$ -glutamyl chains trap the coenzymes inside the cell and promote specific binding to pteroylpolyglutamate-dependent enzymes, their functions and the rules governing the length of polyglutamate tails are poorly understood (31). The *Methanococcales* are unusual in having F<sub>420</sub> and MPT-related coenzymes that contain  $\alpha$ -linked glutamate residues. Among the bacteria, only folypolyglutamates of *E. coli* K-12 str. W1485 were reported to contain both  $\gamma$ - and  $\alpha$ -linked glutamates (32).  $\gamma$ -Folypolyglutamate: $\alpha$ -L-glutamate ligase activity has also been partially purified from cell-free *E. coli* extract (33). Nevertheless, no protein or gene has yet been linked to that activity. Similarly, the *Methanosarcinales* are known to produce sarcinapterin, the same  $\alpha$ -linked glutamate-modified form of methanopterin found in the *Methanococcales* (16). Until the present identification of the MptN protein, no enzyme had been shown to be responsible for this modification.

The CofF and MptN ligases are members of the large ATP-grasp superfamily of amide bond-forming enzymes. This group is one of several superfamilies that have independently evolved amino acid ligase activity. Others include the MurD peptide ligase family (folypoly- $\gamma$ -glutamate synthetase, poly- $\gamma$ -glutamate synthetase, and MurCDEF peptide ligases; ref. 34) and the glutamine synthetase family (guanido kinases and probably  $\gamma$ -glutamylcysteine synthetase; ref. 7). These enzymes share a common reaction mechanism (nucleophilic attack by an  $\alpha$ -amino group on an activated acyl phosphate intermediate), and each is proposed to undergo a conformational shift during catalysis (30, 35). Although their quaternary structures range from monomers to tetramers (Table 1), each subunit of these ligases seems to function independently. The apparent  $K_m$  values for the enzymes range from millimolar concentrations (glutathione biosynthetic enzymes) to micromolar concentrations (F<sub>420</sub>, methanopterin, and folate) consistent with cellular concentrations of the respective metabolites. Correspondingly, the specific activities of the glutathione synthetase,  $\gamma$ -glutamylcysteine synthetase and D-alanine:D-alanine ligase enzymes are much higher than those of the enzymes that use F<sub>420</sub> or folate substrates. In all of the enzymes, two Mg<sup>2+</sup> or Mn<sup>2+</sup> ions are required along with ATP (or GTP) for activity.

These enzyme families differ in their substrate specificities and their modes of evolution. Except for the RimK protein, members

**Table 1. Comparison of amino acid ligases**

Enzyme*	GshB	GSH2	DdlB	CofF	MptN	GSH1	FolC	CofE
Mol. mass, kDa	36	56	33	33	33	73	45	27
Oligomeric state	Tetramer	Dimer	Dimer	Monomer	Dimer	Hetero-dimer	Monomer	Dimer
Carboxylate substrate	$\gamma$ -Glu-Cys	$\gamma$ -Glu-Cys <sup>†</sup>	D-Ala	$\gamma$ -F <sub>420</sub> -(Glu) <sub>2</sub>	MPT	L-Glu	H <sub>4</sub> Folate	F <sub>420</sub> -O/ $\gamma$ -F <sub>420</sub> -(Glu) <sub>1</sub>
$K_m^{app}$	2.6 mM	0.5 mM	(3.3 $\mu$ M, 1.2 mM)	1.2 $\mu$ M	ND	1.6 mM	14 $\mu$ M	1.0 $\mu$ M
Amino substrate	Gly	Gly	D-Ala	L-Glu	L-Glu	L-Cys	L-Glu	L-Glu
$K_m^{app}$	2.0 mM	2.0 mM	(3.3 $\mu$ M, 1.2 mM)	ND	ND	0.3 mM	0.3 mM	ND
Specific activity <sup>‡</sup>	11	11	31	0.0004	0.013	1600	0.5	0.0006
Superfamily	ATP-grasp	ATP-grasp	ATP-grasp	ATP-grasp	ATP-grasp	ND	MurD ligase	ND
Divalent metal ions	Mg <sup>2+</sup> , Mn <sup>2+</sup> , Co <sup>2+</sup>	Mg <sup>2+</sup> , Mn <sup>2+</sup>	Mg <sup>2+</sup>	Mg <sup>2+</sup> , Mn <sup>2+</sup>	Mg <sup>2+</sup>	Mg <sup>2+</sup> , Mn <sup>2+</sup>	Mg <sup>2+</sup> , Mn <sup>2+</sup>	Mg <sup>2+</sup> , Mn <sup>2+</sup>

ND, not determined.

\*Enzymes and their sources were GshB, glutathione synthetase from *E. coli* B (37); GSH2, glutathione synthetase from *Saccharomyces cerevisiae* (38); DdlB, D-alanine:D-alanine ligase from *E. coli* (39); CofF and MptN from *M. jannaschii* (this work); GSH1,  $\gamma$ -L-glutamate:L-cysteine ligase from rat kidney (40); FolC, bifunctional tetrahydrofolate:  $\gamma$ -L-glutamate ligase/dihydrofolate synthase from *E. coli* (41); and CofE, F<sub>420</sub>-O: $\gamma$ -L-glutamate ligase from *M. jannaschii* (42).

<sup>†</sup>Kinetic values for GSH2 were measured using  $\gamma$ -L-glutamyl-L- $\alpha$ -aminobutyrate and glycine as amino acid substrates.

<sup>‡</sup>Specific activities are micromoles of substrate converted to product per min per mg of enzyme.



of the ATP-grasp and glutamine synthetase superfamilies catalyze the ligation of a single additional amino acid. Crystal structure models of the glutathione synthetase and D-alanine:D-alanine ligase proteins show that residues in the active sites interact extensively with the amino acid substrates, restricting the size of potential substrates (9). In contrast, members of the MurD peptide ligase family either catalyze multiple ligations (folypoly- $\gamma$ -glutamate synthetase and poly- $\gamma$ -glutamate synthetase) or have evolved by gene duplication to catalyze successive steps in peptide biosynthesis (MurCDEF proteins). Crystal structure models of MurD and folypoly- $\gamma$ -glutamate synthetase protein complexes show substrates bound in a large cleft that contacts only a portion of the carboxylate substrate, allowing these enzymes to accommodate expansive substrates (35).

Previous studies of the evolutionary history of glutathione synthetase proteins concluded that the bacterial and eukaryotic genes evolved separately, probably from an ancestral gene encoding an ATP-grasp type protein (2, 9). Whereas the eukaryal-type gene was vertically inherited throughout the aerobic eukaryotes, the bacterial-type glutathione synthetase gene seems to have been vertically inherited only within the  $\alpha$ , $\beta$ , $\gamma$ -proteobacterial and cyanobacterial lineages (results not shown). D-Alanine:D-alanine ligase genes have been vertically inherited throughout the Bacteria, except for the homologs that confer glycopeptide antibiotic resistance (these homologs have been laterally transferred; ref. 36). The RimK protein that modifies

ribosomal protein S6 is primarily restricted to  $\gamma$ -proteobacteria. The phylogeny of ATP-grasp genes presented here (Fig. 2), suggests that the  $\alpha$ -L-glutamate ligases evolved from a single ancestral gene, which diverged before the evolution of the modern glutathione synthetase and D-alanine:D-alanine ligase gene lineages.

Unlike the modular peptide synthetase systems that assemble complete peptides by using covalently bound substrates and products, the single-step peptide synthetases described here must carefully discriminate among diffusible intermediates. This requirement for specific carboxylate and amine substrates constrains the evolution of new peptide biosynthetic pathways. For the multistep glutathione, F<sub>420</sub>, folate, and murein peptide biosyntheses, cells have recruited individual peptide synthetases from several different superfamilies, each with suitably different substrate binding and kinetic properties. The diversity of yet-to-be-characterized ATP-grasp genes in genomes from the Archaea, Bacteria, and Eucarya underscores the genes' histories of recruitment into new biosynthetic pathways.

We thank Dr. Mehdi Ashraf-Khorassani (Department of Chemistry, Virginia Polytechnic Institute and State University) for performing atmospheric pressure chemical ionization MS analysis. This work was supported by a U.S. National Science Foundation grant awarded to D.E.G. in 2002 and National Science Foundation Grant MCB 0231319 (to R.H.W.).

1. Newton, G. L. & Fahey, R. C. (1990) in *Glutathione Metabolism and Physiological Function*, ed. Viña, J. (CRC, Boca Raton, FL), pp. 69–77.
2. Copley, S. D. & Dhillon, J. K. (2002) *Genome Biol.* **3**, 1–16.
3. Fahey, R. C. (2001) *Annu. Rev. Microbiol.* **55**, 333–356.
4. Allen, J. R., Clark, D. D., Krum, J. G. & Ensign, S. A. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 8432–8437.
5. Fahey, R. C. & Sundquist, A. R. (1991) *Adv. Enzymol. Relat. Areas Mol. Biol.* **64**, 1–53.
6. Strumeyer, D. H. & Bloch, K. (1960) *J. Biol. Chem.* **235**, PC27.
7. Abbott, J. J., Pei, J., Ford, J. L., Qi, Y., Grishin, V. N., Pitcher, L. A., Phillips, M. A. & Grishin, N. V. (2001) *J. Biol. Chem.* **276**, 42099–42107.
8. Fan, C., Moews, P. C., Shi, Y., Walsh, C. T. & Knox, J. R. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 1172–1166.
9. Polekhina, G., Board, P. G., Gali, R. R., Rossjohn, J. & Parker, M. W. (1999) *EMBO J.* **18**, 3204–3213.
10. Kang, W.-K., Icho, T., Isono, S., Kitakawa, M. & Isono, K. (1989) *Mol. Gen. Genet.* **217**, 281–288.
11. Nishida, H., Nishiyama, M., Kobashi, N., Kosuge, T., Hoshino, T. & Yamane, H. (1999) *Genome Res.* **9**, 1175–1183.
12. Yamaguchi, H., Kato, H., Hata, Y., Nishioka, T., Kimura, A., Oda, J. & Katsube, Y. (1993) *J. Mol. Biol.* **229**, 1083–1100.
13. Galperin, M. Y. & Koonin, E. V. (1997) *Protein Sci.* **6**, 2639–2643.
14. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., et al. (1996) *Science* **273**, 1017–1140.
15. Graham, D. E. & White, R. H. (2002) *Nat. Prod. Rep.* **19**, 133–147.
16. van Beelen, P., Labro, J. F., Keltjens, J. T., Geerts, W. J., Vogels, G. D., Laarhoven, W. H., Guijt, W. & Haasnoot, C. A. (1984) *Eur. J. Biochem.* **139**, 359–365.
17. Graupner, M. & White, R. H. (2003) *J. Bacteriol.* **185**, 4662–4665.
18. White, R. H. (1990) *Biochemistry* **29**, 5397–5404.
19. White, R. H. (1997) *Methods Enzymol.* **281**, 391–401.
20. Smyth, R. B. & McKeown, G. G. (1964) *J. Chromatogr.* **16**, 454–459.
21. Graham, D. E., Xu, H. & White, R. H. (2002) *J. Biol. Chem.* **277**, 13421–13429.
22. Li, H., Xu, H., Graham, D. E. & White, R. H. (2003) *J. Biol. Chem.* **278**, 11100–11106.
23. Furness, R. A. H. & Loewen, P. C. (1981) *Anal. Biochem.* **117**, 126–135.
24. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402.
25. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680.
26. Holm, L. & Sander, C. (1996) *Science* **273**, 595–603.
27. Felsenstein, J. (2001) PHYLIP, Phylogeny Inference Package (Dept. of Genetics, Univ. of Washington, Seattle), Version 3.6a2.1.
28. Hayashi, R. (1976) *Methods Enzymol.* **45**, 568–587.
29. Goldman, P. & Levy, C. C. (1967) *Proc. Natl. Acad. Sci. USA* **58**, 1299–1306.
30. Gogos, A. & Shapiro, L. (2002) *Structure (London)* **10**, 1669–1676.
31. McGuire, J. J. & Coward, J. K. (1984) in *Folates and Pterins: Chemistry and Biochemistry of Folates*, eds. Blakley, R. L. & Benkovic, S. J. (Wiley, New York), Vol. 1, pp. 135–190.
32. Ferone, R., Hanlon, M. H., Singer, S. C. & Hunt, D. F. (1986) *J. Biol. Chem.* **261**, 16356–16362.
33. Ferone, R., Singer, S. C. & Hunt, D. F. (1986) *J. Biol. Chem.* **261**, 16363–16371.
34. Eveland, S. S., Pompliano, D. L. & Anderson, M. S. (1997) *Biochemistry* **36**, 6223–6229.
35. Bertrand, J. A., Fanchon, E., Martin, L., Chantalat, L., Auger, G., Blanot, D., van Heijenoort, J. & Dideberg, O. (2000) *J. Mol. Biol.* **301**, 1257–1266.
36. Evers, S., Casadewall, B., Charles, M., Dutka-Malen, S., Galimand, M. & Courvalin, P. (1996) *J. Mol. Evol.* **42**, 706–712.
37. Gushima, H., Miya, T., Murata, K. & Kimura, A. (1983) *J. Appl. Biochem.* **5**, 210–218.
38. Meister, A. (1974) in *The Enzymes*, ed. Boyer, P. D. (Academic, New York), Vol. 10, pp. 671–696.
39. Zawadzke, L. E., Bugg, T. D. & Walsh, C. T. (1991) *Biochemistry* **30**, 1673–1682.
40. Meister, A. (1989) in *Glutathione: Chemical, Biochemical and Medical Aspects*, eds. Dolphin, D., Avramovic, O. & Poulson, R. (Wiley, New York), Vol. 3A, pp. 367–474.
41. Bogner, A. L., Osborne, C., Shane, B., Singer, S. C. & Ferone, R. (1985) *J. Biol. Chem.* **260**, 5625–5630.
42. Li, H., Graupner, M., Xu, H. & White, R. H. (2003) *Biochemistry*, in press.