The Identification and Structural Characterization of C7orf24 as γ-Glutamyl Cyclotransferase

AN ESSENTIAL ENZYME IN THE γ-GLUTAMYL CYCLE

Received for publication, May 12, 2008, and in revised form, May 30, 2008. Published, JBC Papers in Press, May 30, 2008, DOI 10.1074/jbc.M803623200

Aaron J Oakley‡, Tetsuo Yamada§, Dan Liu†, Marjorie Coggan†, Alan G. Clark§, and Philip G. Board‡†

From the ¦Research School of Chemistry and the ‡John Curtin School of Medical Research, Australian National University, Canberra, Australian Capital Territory 2601, Australia and the §School of Biological Sciences, Victoria University of Wellington, Wellington 6140, New Zealand

The hypothetical protein C7orf24 has been implicated as a cancer marker with a potential role in cell proliferation. We have identified C7orf24 as γ-glutamyl cyclotransferase (GGCT) that catalyzes the formation of 5-oxoproline (pyroglutamic acid) from γ-glutamyl dipeptides and potentially plays a significant role in glutathione homeostasis. In the present study we have identified the first cDNA clones encoding a γ-glutamyl cyclotransferase. The GGCT gene is located on chromosome 7p14-15 and consists of four exons that span 8 kb. The primary sequence is 188 amino acids in length and is unlike any protein of known function. We crystallized functional recombinant γ-glutamyl cyclotransferase and determined its structure at 1.7 Å resolution. The enzyme is a dimer of 20,994-Da subunits. The topology of GGCT is unrelated to other enzymes associated with cyclotransferase-like activity. The fold was originally classified as “BtrG-like,” a small family that only includes structures of hypothetical proteins from Mus musculus, Escherichia coli, Pyrococcus horikoshii, and Arabidopsis thaliana. Since this is the first member of this family with a defined function, we propose to refer to this structure as the γ-glutamyl cyclotransferase fold. We have identified a potential active site pocket that contains a highly conserved glutamic acid (Glu98) and propose that it acts as a general acid/base in the reaction mechanism. Mutation of Glu98 to Ala or Gln completely inactivates the enzyme without altering the overall fold.

We have identified C7orf24 as γ-glutamyl cyclotransferase (GGCT), a poorly studied enzyme in the γ-glutamyl cycle. The γ-glutamyl cycle involves the synthesis of glutathione and its utilization in the uptake of amino acids, particularly in the renal tubule and across the blood-brain barrier in the choroid plexus in the brain (4, 5) (Fig. 1). GGCT (EC 2.3.2.4) is an essential enzyme in this pathway and catalyzes the general reaction,

\[ \text{L-γ-Glutamyl-L-amino acid} \rightarrow 5\text{-oxoproline} + \text{L-amino acid} \]

Deficiency of glutathione synthetase can lead to chronic 5-oxoprolinuria as a result of the uncontrolled synthesis of the dipeptide γ-glutamylcysteine and its subsequent hydrolysis to cysteine and 5-oxoproline (pyroglutamic acid) by GGCT (6, 7) (Fig. 1). In some patients, the generation and urinary excretion of 5-oxoproline has been reported to be up to 50 g/day. This extraordinary production of 5-oxoproline and associated acidosis has been considered to be a significant factor in the pathology associated with this deficiency. The inhibition of GGCT in cases of glutathione synthetase deficiency would block the degradation of γ-glutamylcysteine and allow its accumulation to a level where it may partially substitute for glutathione in redox and detoxification reactions. Thus, there is potential therapeutic value in designing specific inhibitors of GGCT. Deficiency of 5-oxoprolinase, the enzyme that catalyzes the ATP-dependent decyclization of 5-oxoproline to glutamate, has been reported in several patients and also leads to 5-oxoprolinuria (4–10 g/day) (7). This indicates that there is a significant turnover of glutathione and γ-glutamyl-amino acid dipeptides via GGCT and the γ-glutamyl cycle under normal metabolic conditions. The position of GGCT in the γ-glutamyl cycle suggests that it could play a significant role in regulating the synthesis of glutathione by limiting the availability of γ-glutamylcysteine; however, this possibility has not been critically investigated.

Several previous studies have purified GGCTs from a range of mammalian species (8–12), and multiple isoforms have been detected in rat and human tissues. An early study using starch gel electrophoresis identified a genetic polymorphism in mouse GGCT and mapped the locus by linkage to Lyt-2 on mouse

‡ This work was supported by a grant from the Australian Research Council (to P. B.) The research was also supported by the Australian Synchrotron Research Program and facilitated by access to the Australian Proteome Analysis Facility; both were established under the Australian Government’s Major National Research Facilities Program. 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.

The atomic coordinates and structure factors (codes 2PN7, 2RH, and 3CRV) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

§ To whom correspondence should be addressed: John Curtin School of Medical Research, P.O. Box 334, Canberra, Australian Capital Territory 2601, Australia. Tel.: 61-2-61254714; Fax: 61-2-61254712; E-mail: Philip.Board@anu.edu.au.

1 The abbreviations used are: GGCT, γ-glutamyl cyclotransferase; PEG, polyethylene glycol; EST, expressed sequence tag.
Structure of \(\gamma\)-Glutamyl Cyclotransferase

An IMAGE consortium kidney cDNA clone (BC019243) was used as a template with the primers hGCTSF 5'-GTACCGCG-GTGGCATGGCCAACTCGGGCTG and hGCTNR 5'-AATT-GCGGGCGGCCACATAGAATACCCTTAG to amplify the GGCT coding sequence by PCR. The amplified product (629 bp) was cloned into pGEM-T (Promega) and sequenced to confirm the absence of amplification errors. All DNA sequencing was on an ABI 3730 sequencer (Australian Cancer Research Foundation Biomolecular Resource Facility, John Curtin School of Medical Research, Australian National University) following the manufacturer’s protocol (Applied Biosystems). The cloned cDNA was excised from the pGEM-T vector with KspI and NotI ligated into the same sites in the pHUE vector (15) and transfected into Escherichia coli BL21(DE3). Expression of recombinant enzymes in pHUE allows the rapid purification of the expressed protein by Ni\(^{2+}\)-agarose chromatography and the subsequent removal of all additional NH\(_2\)-terminal residues by cleavage with the catalytic domain of mouse ubiquitin-specific protease 2, as previously described in detail (15). Some minor contaminating proteins remained, so the enzyme was subjected to an additional step of ion exchange chromatography. The protein was dialyzed into 10 mM potassium phosphate, pH 6.5, fractionated on a 1 × 12-cm column of DEAE-Sepharose with a linear gradient to 0.2 M potassium phosphate, pH 6.5, containing 1 M NaCl. The eluted fractions containing GGCT activity were pooled, and if not used immediately, the enzyme was stored frozen at −20 °C.

Selenomethionine-labeled GGCT was expressed using the same vector but grown under modified conditions, as described by Van Duyne et al. (16). The labeled GGCT was purified as described above and transferred into 20 mM Tris/HCl, 0.1 mM dithiothreitol, pH 8.0, while being concentrated after the DEAE ion exchange chromatography step.

**Determination of GGCT Activity and Kinetic Constants**

A continuous spectrophotometric method that links the production of free alanine to the oxidation of NADH via the action of added glutamate-pyruvate transaminase and lactate dehydrogenase was used, as previously described (14). The kinetic constants were determined by varying the \(\gamma\)-glutamyl-L-alanine concentration between 0.5 and 30 mM over seven steps in triplicate. The rate data were fitted to the Michaelis-Menten equation by the MacCurveFit version 1.5 program (Kevin Raner Software).

Western blots were performed as described by Towbin et al. (17). The antiserum against GGCT was raised in rabbits against purified human erythrocyte GGCT or recombinant human GGCT expressed in E. coli. The antibodies were diluted (1:1000) in PBS with 1% skim milk. After a 1-h incubation with the primary antibody and washing, the membrane was incubated with peroxidase-conjugated goat anti-rabbit immunoglobulins (1:8000; DAKO) for 1 h and washed three times with phosphate-buffered saline. Bound antibodies were visualized by using enhanced chemiluminescence (ECL Amersham Biosciences).
Crystallization and Data Collection

Wild Type Structure—A 2-μl sample of selenomethionine containing GGCT (15 mg/ml in 20 mM Tris/HCl, 0.1 mM dithiothreitol, pH 8.0) was mixed with equal volumes of γ-Glu-Glu dipeptide (40 mM) and well solution. The well contained 1 ml of a solution comprising PEG 4000 (22.5% w/v), 100 mM sodium formate buffer, pH 3.9, 200 mM ammonium acetate. Crystals appeared as flat rods within 7 days and reached maximum size (longest edge 200 μm) after 20 days. For cryoprotection, crystals were transferred to artificial mother liquor containing 32.5% PEG 4000 prior to placement in a cryostream at 100 K.

After collection and integration of peak and inflection anomalous data (Table 1), the structure was solved automatically with HKL3000 (18) as follows. Eight selenium sites were found by SHELXD (19). (This is consistent with a dimer in the asymmetric unit, since GGCT contains four methionine residues not including the NH₂-terminal methionine.) RESOLVE (20) detected noncrystallographic 2-fold symmetry based on the selenium sites. The heavy atom search was followed by 10 cycles of solvent flattening in SHELXE (19) using map contrast and connectivity to assign the correct enantiomorph. Data between 50 and 3.0 Å resolution was used for obtaining initial MAD phases. After heavy atom refinement and phase calculation in MLPHARE (21), the mean figure of merit for phases in this resolution range was 0.27. The program DM (22) was used to improve phases through solvent flattening, histogram mapping, noncrystallographic symmetry averaging, and phase extension. The resulting electron density maps were sufficient for RESOLVE to automatically build 80% of the protein structure. The “peak” x-ray data were used for these rounds of structure refinement. All reflection data were used for electron density map generation, model building, and structure refinement.

Ala98 Mutant Structure—Crystals were grown at 4 °C. A 2-μl sample of protein (15 mg/ml) was mixed with equal volumes of 20 mM γ-Glu dipeptides (either γ-Glu-Phe, γ-Glu-Glu, γ-Glu-Cys, γ-Glu-Gly, or γ-Glu-Ala) and well solution before being suspended over the well in a hanging drop configuration. The well solution was composed of 0.1 M potassium acetate buffer, pH 4.25, 20% PEG 4000, and 0.2 M ammonium acetate. The largest crystals grew from solutions containing γ-Glu-Phe (longest edge 200 μm) and appeared after 3 days and grew to maximum dimensions after 1 week.

For x-ray data collection, crystals were cryoprotected by transfer to artificial mother liquor with PEG 4000 concentration increased to 30%. Crystals were snap-frozen using an Oxford cryostream at 100 K. Data were collected using a Mar 165-mm CCD at the Australian Synchrotron. Data were processed with Mosflm (23) and scaled with Scala (24). Structure refinement commenced with the Ala98 mutant structure.

For all three models, model building and refinement was conducted using REFMAC5 (25) and O (26). Ramachandran plot statistics were derived from PROCHECK (27). Statistics for the final models are given in Table 1.

RESULTS

Identification of a GGCT cDNA Clone—Human GGCT was partially purified from human erythrocytes by ion exchange chromatography and ammonium sulfate precipitation. The presence of the enzyme was confirmed by a specific spectrophotometric assay and detection on Western blots with an antibody raised against highly purified human erythrocyte GGCT in an earlier study (11). Although the partially purified preparation contained several components, a relatively strong ~21-kDa Coomassie Blue-stained band that corresponded to the most prominent band detected on Western blots was selected for further investigation. Trypsin digest and mass spectrum analysis revealed five peptides that gave 46% coverage of a previously undescribed protein known as C7orf24 (chromosome 7 open reading frame 24) (accession number HS.530024). The C7orf24 protein appeared to be a good candidate, since its predicted molecular mass is around 21 kDa, and its position on chromosome 7 is syntenic with the previously mapped location of the mouse Ggt gene on chromosome 6 (13).

Analysis of the EST data base revealed many C7orf24 cDNA clones, and we selected one (BC019243) from human kidney for further study. The longest open reading frame within the selected EST clone was amplified and cloned in the pHUE expression vector.

Characterization of Recombinant Human GGCT—Subsequent expression and purification of recombinant protein from the C7orf24 cDNA generated a protein with a molecular size on SDS-PAGE of ~21 kDa, and a Western blot revealed that the expressed protein cross-reacted strongly with the original antibody raised against purified human erythrocyte GGCT (Fig. 2). To further confirm that this protein was GGCT, we measured its activity with γ-glutamyl-L-Alanine as a substrate. The recombinant enzyme was very active and had a specific activity of 50.3 μmol/min/mg. Some kinetic properties of the recombinant enzyme are shown in Table 2. Thus, as a result of the immunological cross-reactivity and the enzymatic activity, we conclude that the EST clone identified in this study encodes a functional GGCT enzyme.

The Coding Sequence of Human GGCT—The cDNA sequence and the deduced amino acid sequence of human GGCT is shown in Fig. 3. The coding region encodes a peptide of 188 amino acids and has a predicted molecular mass of 20,994 Da that is in agreement with the apparent subunit size of the recombinant enzyme determined by SDS-PAGE (Fig. 2).
### TABLE 1
X-ray and structure data

| Parameter | Value |
|-----------|-------|
| **Native structure** | | |
| Protein Data Bank code | 2PN7 |
| Space Group | P2$_1$2$_1$2$_1$ |
| Unit cell (Å, degrees) | $a = 35.2$, $b = 98.7$, $c = 100.7$ |
| Wavelength (Å) | 0.97934 |
| Resolution (Å) | 50.0-2.40 (2.49-2.40)* |
| $I/|I|$ | 11.14 (2.09) |
| $R$-merge (%) | 13.3 (48.6) |
| Completeness (%) | 98.3 (90.2) |
| Unique reflections | 13,898 (1257) |
| Average redundancy | 6.1 (3.3) |
| Atoms | | |
| Protein | 2681 |
| Water | 34 |
| $R$-factor, $R$-free (%) | 19.0, 25.1 |
| Root mean square deviations from ideal values | | |
| Bond lengths (Å) | 0.015 |
| Bond angles (degrees) | 1.506 |
| Chiral center restraints (Å³) | 0.101 |
| Ramachandran plot | | |
| Residues in most favorable region | 90.5% |
| Residues in additionally allowed region | 9.5% |
| Mean $B$-factors | | |
| Protein | 45.4 |
| Water | 40.8 |
| **E98A mutant** | | |
| Protein Data Bank code | 2RBH |
| Space group | P2$_1$2$_1$2$_1$ |
| Unit cell (Å, degrees) | $a = 34.9$, $b = 99.4$, $c = 100.0$ |
| Wavelength (Å) | 1.54 (Cu Kα) |
| Resolution (Å) | 50.0-2.10 (2.17-2.10) † |
| $I/|I|$ | 10.04 (3.93) |
| $R$-merge (%) | 8.3 (26.8) |
| Completeness (%) | 78.8 (78.0) |
| Unique reflections | 16603 (1613) |
| Average redundancy | 2.2 (2.1) |
| Atoms | | |
| Protein | 2673 |
| Water | 70 |
| $R$-factor, $R$-free (%) | 20.1, 25.5 |
| Root mean square deviations from ideal values | | |
| Bond lengths (Å) | 0.010 |
| Bond angles (degrees) | 1.231 |
| Chiral center restraints (Å³) | 0.084 |
| Ramachandran plot | | |
| Residues in most favorable region | 93.5% |
| Residues in additionally allowed region | 6.5% |
| Mean $B$-factors | | |
| Protein | 26.3 |
| Water | 26.5 |
| **E98Q mutant** | | |
| Protein Data Bank code | 3CRY |
| Space group | P2$_1$2$_1$2$_1$ |
| Unit cell (Å, degrees) | $a = 34.8$, $b = 98.1$, $c = 100.5$ |
| Wavelength (Å) | 0.95361 |
| Resolution (Å) | 70-1.70 (1.8-1.7) |
| $I/|I|$ | 12.6 (3.7) |
| $R$-merge (%) | 9.2% (41.9%) |
| Completeness (%) | 99.4 (97.7) |
| Unique reflections | 38628 (5449) |
| Average redundancy | 4.6 (4.4) |
| Atoms | | |
| Protein | 2681 |
| Water | 399 |
| Acteate | 12 |
| $R$-factor, $R$-free (%) | 18.0%, 22.3% |
| Root mean square deviations from ideal values | | |
| Bond lengths (Å) | 0.013 |
| Bond angles (degrees) | 1.241 |
| Chiral center restraints (Å³) | 0.093 |
| Ramachandran plot | | |
| Residues in most favorable region | 92.9% |
| Residues in additionally allowed region | 7.1% |
| Mean $B$-factors | | |
| Protein | 13.2 |
| Water | 26.7 |

*Values in parentheses refer to highest resolution bin.

$R$-merge = $\sum |I_{o} - \langle I_{o} \rangle| / \sum I_{o}$

$R$-factor, $R$-free was calculated from 5% of the diffraction data not used in refinement.
BLAST (28) searches were used to identify related sequences in publicly available EST and protein data bases. The sequences were aligned by the program ClustalW (29), and the percentage identities are shown in Table 3. Although we found related sequences in a range of species from Caenorhabditis elegans to mammals, the best matches were in higher species, and further studies will be needed to determine if the diverged proteins identified in species such as Drosophila melanogaster, Strongylocentrotus purpuratus (echinoderm), and Caenorhabditis elegans encode proteins with GGCT activity. The BLAST search did not identify any similar proteins in plants, although a protein with GGCT activity has been detected in tobacco (30). It was particularly notable that the GGCT sequence showed no similarity to other previously described proteins of known function, including glutaminyl cyclases that catalyze the formation of pyroglutamate at the amino termini of peptides (31).

Crystal Structure of Human GGCT—The initial structure of human GGCT was determined at 2.4 Å resolution by multienzyme wavelength anomalous dispersion. The crystals were orthorhombic and belonged to the space group P2₁2₁2₁. The unit cell parameters and space group (Table 1) were consistent with a dimer in the asymmetric unit with a Matthews coefficient of 2.08 Å³/dalton. The structure comprises a dimer with continuous 120° elbow and reversing their topological hierarchy. Dimerization interactions occur across strand 6. Dimerization interactions occur across strand 2 and strands 1, 2, and 6. Dimerization interactions occur across strand 3 and 310-helix 4. Because of the moderate resolution of the x-ray data, water molecules were built into the crystal structures, which lies adjacent to the noncrystallographic 2-fold axis, forming a 120° elbow and reversing their topological hierarchy. The helices are loosely clustered around strands β1, β2, β3, and β6. Dimerization interactions occur across strand β4, which lies adjacent to the noncrystallographic 2-fold axis, forming a continuous β-sheet with its noncrystallographic symmetry-related partner (Fig. 5). Other dimerization interactions lie between strand β3 and 3₁₀-helix 4. Because of the moderate resolution of the x-ray data, water molecules were built into electron density in a conservative manner, resulting in lower average B-factor for water compared with protein (Table 1).

Each GGCT monomer features an invagination formed by helices α₁ and α₂ and strands β1 and β2. There is electron density within this site that cannot be explained by the presence of a water molecule but cannot easily be interpreted as substrate or reaction product. We propose that this is the location of the active site. The site is lined with hydrophilic and amphipathic...
residues (Fig. 6). At the COOH-terminal end of helix α2 lies Glu98, which we propose acts as a general acid/base in the reaction (Fig. 7). Two water molecules were observed to cluster around Ser24, and the unexplained density is found at the bottom of the pocket near Tyr22. The absence of electron density for reactants and products can be explained by catalytic turnover of the substrate in the crystal and product release by the enzyme.

Glu98 is conserved in sequence homologues as distantly related as Anopheles gambiae. We mutated Glu98 to Gln or Ala and found that both mutations completely inactivate the enzyme (Table 2). This strongly supports the contention that Glu98 plays a significant role in catalysis and is consistent with the role of active site carboxylic residues in the aforementioned glutaminylcyclases.

To confirm that the loss of activity was not simply due to misfolding of the mutated protein, we crystallized the E98A and E98Q mutants and solved their structures. The crystals were isomorphous, with the wild type enzyme having the same space group and similar unit cell dimensions. A summary of the data is shown in Table 1. The structures of the wild type and the mutant enzymes superimpose with root mean square deviations of 0.212 Å (E98A) and 0.278 Å (E98Q) over 338 C-α atoms, indicating that the mutations had not significantly altered the overall structure. The E98A mutation was confirmed by residues (Fig. 6). At the COOH-terminal end of helix α2 lies Glu98, which we propose acts as a general acid/base in the reaction (Fig. 7). Two water molecules were observed to cluster around Ser24, and the unexplained density is found at the bottom of the pocket near Tyr22. The absence of electron density for reactants and products can be explained by catalytic turnover of the substrate in the crystal and product release by the enzyme.

Glu98 is conserved in sequence homologues as distantly related as Anopheles gambiae. We mutated Glu98 to Gln or Ala and found that both mutations completely inactivate the enzyme (Table 2). This strongly supports the contention that Glu98 plays a significant role in catalysis and is consistent with the role of active site carboxylic residues in the aforementioned glutaminylcyclases.

To confirm that the loss of activity was not simply due to misfolding of the mutated protein, we crystallized the E98A and E98Q mutants and solved their structures. The crystals were isomorphous, with the wild type enzyme having the same space group and similar unit cell dimensions. A summary of the data is shown in Table 1. The structures of the wild type and the mutant enzymes superimpose with root mean square deviations of 0.212 Å (E98A) and 0.278 Å (E98Q) over 338 C-α atoms, indicating that the mutations had not significantly altered the overall structure. The E98A mutation was confirmed by residues (Fig. 6). At the COOH-terminal end of helix α2 lies Glu98, which we propose acts as a general acid/base in the reaction (Fig. 7). Two water molecules were observed to cluster around Ser24, and the unexplained density is found at the bottom of the pocket near Tyr22. The absence of electron density for reactants and products can be explained by catalytic turnover of the substrate in the crystal and product release by the enzyme.

Glu98 is conserved in sequence homologues as distantly related as Anopheles gambiae. We mutated Glu98 to Gln or Ala and found that both mutations completely inactivate the enzyme (Table 2). This strongly supports the contention that Glu98 plays a significant role in catalysis and is consistent with the role of active site carboxylic residues in the aforementioned glutaminylcyclases.

To confirm that the loss of activity was not simply due to misfolding of the mutated protein, we crystallized the E98A and E98Q mutants and solved their structures. The crystals were isomorphous, with the wild type enzyme having the same space group and similar unit cell dimensions. A summary of the data is shown in Table 1. The structures of the wild type and the mutant enzymes superimpose with root mean square deviations of 0.212 Å (E98A) and 0.278 Å (E98Q) over 338 C-α atoms, indicating that the mutations had not significantly altered the overall structure. The E98A mutation was confirmed by residues (Fig. 6). At the COOH-terminal end of helix α2 lies Glu98, which we propose acts as a general acid/base in the reaction (Fig. 7). Two water molecules were observed to cluster around Ser24, and the unexplained density is found at the bottom of the pocket near Tyr22. The absence of electron density for reactants and products can be explained by catalytic turnover of the substrate in the crystal and product release by the enzyme.

Glu98 is conserved in sequence homologues as distantly related as Anopheles gambiae. We mutated Glu98 to Gln or Ala and found that both mutations completely inactivate the enzyme (Table 2). This strongly supports the contention that Glu98 plays a significant role in catalysis and is consistent with the role of active site carboxylic residues in the aforementioned glutaminylcyclases.

To confirm that the loss of activity was not simply due to misfolding of the mutated protein, we crystallized the E98A and E98Q mutants and solved their structures. The crystals were isomorphous, with the wild type enzyme having the same space group and similar unit cell dimensions. A summary of the data is shown in Table 1. The structures of the wild type and the mutant enzymes superimpose with root mean square deviations of 0.212 Å (E98A) and 0.278 Å (E98Q) over 338 C-α atoms, indicating that the mutations had not significantly altered the overall structure. The E98A mutation was confirmed by residues (Fig. 6). At the COOH-terminal end of helix α2 lies Glu98, which we propose acts as a general acid/base in the reaction (Fig. 7). Two water molecules were observed to cluster around Ser24, and the unexplained density is found at the bottom of the pocket near Tyr22. The absence of electron density for reactants and products can be explained by catalytic turnover of the substrate in the crystal and product release by the enzyme.

Glu98 is conserved in sequence homologues as distantly related as Anopheles gambiae. We mutated Glu98 to Gln or Ala and found that both mutations completely inactivate the enzyme (Table 2). This strongly supports the contention that Glu98 plays a significant role in catalysis and is consistent with the role of active site carboxylic residues in the aforementioned glutaminylcyclases.
Substrate Binding Model—Based on the structural data and the mutagenesis of Glu\(^98\), we have developed a model of substrate binding shown in Fig. 8. In this model, the γ-Glu group is buried in the proposed active site pocket with the amine nitrogen engaged in a salt bridge with Glu\(^98\). The carboxylate group accepts hydrogen bonds from the backbone amide groups of Tyr\(^{22}\) and Gly\(^{23}\) and the side chain hydroxyl of Tyr\(^{19}\). The amide oxygen in the substrate resides in an oxyanion hole formed by the side chain hydroxyl groups of Tyr\(^{139}\) and Ser\(^{24}\) and by the backbone amide nitrogen of Ser\(^{24}\). The carboxyl group of the second residue in the substrate engages in a salt bridge with Arg\(^{30}\). The side chain of the second residue lies in a groove on the surface of the enzyme formed by Ile\(^{68}\) and Trp\(^{64}\). It is noteworthy that the γ-Glu carboxylate and γ-peptide group overlap closely with the acetate groups seen in the E98Q structure (Fig. 4).

Active Site Mutations—Prior to the development of our preferred substrate binding model, we made additional mutations around the proposed active site pocket. These mutant proteins (G23A, Y105F, and T125F) were kinetically characterized, and the data are presented in Table 2. The G23A mutation caused a 2-fold reduction in \(k_{\text{cat}}\) and a 4-fold increase in the \(K_m\) that results in a marked reduction in the catalytic efficiency \(k_{\text{cat}}/K_m\) and in the specific activity. These changes are consistent with Gly23 contributing to the conformation of the substrate binding pocket. Although the hydrogen bond between the Gly\(^{23}\) NH and the γ-Glu carboxyl would be maintained with the Ala\(^{23}\) substitution, the additional side chain is likely to affect local conformation and substrate binding. The Y105F mutation caused a significant reduction in the \(k_{\text{cat}}\) and a 3-fold increase in the \(K_m\). These changes result in a marked reduction in the catalytic efficiency \(k_{\text{cat}}/K_m\) and in the specific activity. These changes are consistent with Gly\(^{23}\) contributing to the conformation of the substrate binding pocket. Although the
In contrast, the Y125F mutation had only minor effects on the enzyme’s reaction kinetics. This is also consistent with the preferred substrate binding model. Although Tyr^{125} lies in the vicinity of the active site pocket (Fig. 8), its side chain hydroxyl is not positioned where it can contribute directly to substrate binding or catalysis.

**Genomic Organization of the Human GGCT Gene**—A BLAST search of the human genome with the GGCT cDNA sequence revealed matching sequences on chromosomes 5, 7, and 20. The sequences on chromosomes 5 (NT_023148) and 20 (NT_011387) contain many base substitutions and do not contain introns and are therefore likely to represent reverse transcribed pseudogenes. In contrast, the sequence on chromosome 7p15-p14 (NT_079592; AACC02000087) extends over 8 kb and is divided by four introns. The position and size of the introns within the coding sequence is provided in Fig. 3 and Table 4. All of the intron boundaries conform to the GT/AG rule. The first 23 bp (underlined in Fig. 3) of the cDNA studied in detail here (BC019243) do not appear to occur in close proximity to the rest of the gene and could not be located by BLAST searches. A similar extension was also noted in a small number of other cDNA clones from other tissues, suggesting the possibility of variable splicing within the 5’-end of the gene. In contrast, many ESTs contain about 18 alternate 5’ bases that correspond exactly to the genomic sequence, suggesting that the normal start of transcription may lie around 121 bp upstream of the ATG that initiates translation (Fig. 3).

**Expression of GGCT in Human Tissues**—To determine the level of GGCT expression, we used the UniGene site with the EST Profile Viewer to examine the number of GGCT transcripts in different tissues represented in the human EST data base. For ease of comparison, the data in Table 5 have been standardized to show the number of GGCT transcripts/million ESTs in each tissue. Clearly, GGCT is widely expressed in many tissues. On the basis of previous studies and the proposed role of GGCT and the γ-glutamyl cycle in the reabsorption of amino acids in the kidney, it was expected that GGCT expression would be relatively high in this tissue. However, the data suggest that expression levels in the kidney are similar to those in other major organs and tissues, such as the liver and skeletal muscle. This analysis revealed extremely high levels of gene expression in the bladder and salivary gland. The level of expression was even higher in bladder tumor, confirming the original observation. A number of other tumors in kidney, mammary gland, and prostate showed an increase in expression levels compared with normal tissue.

**DISCUSSION**

In this study, we have identified the hypothetical human gene known as C7orf24 (chromosome 7 open reading frame 24) as a gene encoding γ-glutamyl cyclotransferase. This enzyme catalyzes an essential step in the γ-glutamyl cycle (5), and its activity contributes to the excessive generation of 5-oxyproline and the associated pathology in some patients with glutathione synthetase deficiency (7). The identification of C7orf24 as the GGCT gene was based on the immunological identity of recombinant protein expressed from the corresponding cDNA and the demonstration that the recombinant protein has GGCT activity. The identification of the gene was also supported by the observation that the locus at 7p14-15 is syntenic with the region on mouse chromosome 6 where the mouse Ggc7 gene had been localized by linkage analysis (13).

Several independent studies aimed at identifying proteins that are tumor markers have recently shown that C7orf24 is overexpressed in a range of cancers (1–3). In addition, some data indicate that C7orf24 expression is directly related to cancer cell proliferation. Our identification of C7orf24 as GGCT and the determination of its structure has opened the way for further studies of its role in cell proliferation and its evaluation as a novel cancer drug target.

The enzymatic properties of recombinant human GGCT are very similar to the properties of GGCT purified from human eryth-
TABLE 4
DNA sequences of the exon-intron boundaries of human GGCT
Exon sequence is indicated by capital letters, and intron sequence is indicated by lowercase letters. gt/ag splice junctions are underlined. The start of exon 1 is taken as the ATG start of translation.

| Exon | Exon size (bp) | Splice donor | Intron size (bp) | Splice acceptor |
|------|----------------|--------------|-----------------|-----------------|
| 1    | 140            | GCCCTGCAggtgagtgc | 1            | 3886   ctttgtagGATTTTA |
| 2    | 146            | CTGATGATgaggtg   | 2            | 1598   ggttcgGCAAGAAC |
| 3    | 136            | ATAAAAGATgagta   | 3            | 1568   atttccGATTTTA |
| 4    | 240            |               |                |                  |

TABLE 5
Expression profiles derived from GGCT expressed sequence tags

| Tissue                  | GGCT transcripts per 10^6 ESTs |
|-------------------------|-------------------------------|
| Bladder                 | 429                           |
| Bladder tumor           | 577                           |
| Blood                   | 53                            |
| Bone marrow             | 122                           |
| Brain                   | 18                            |
| Cervix                  | 0                             |
| Colon                   | 49                            |
| Colon tumor             | 46                            |
| Esophagus               | 0                             |
| Heart                   | 22                            |
| Kidney                  | 75                            |
| Kidney tumor            | 100                           |
| Liver                   | 76                            |
| Lung                    | 37                            |
| Mammary gland           | 70                            |
| Mammary gland tumor     | 97                            |
| Muscle                  | 78                            |
| Ovary                   | 37                            |
| Pancreas                | 63                            |
| Placenta                | 10                            |
| Prostate                | 64                            |
| Prostate tumor          | 83                            |
| Salivary gland          | 685                           |
| Skin                    | 37                            |
| Small intestine         | 0                             |
| Spleen                  | 18                            |
| Stomach                 | 48                            |
| Testis                  | 11                            |
| Tonsil                  | 58                            |
| Uterus                  | 62                            |

erythrocytes (Table 2). The higher specific activity of the recombinant protein is to be expected because of the rapidity of its purification from freshly cultured bacteria. In contrast, the naturally occurring enzyme was prepared from expired blood bank samples and required a larger number of chromatographic steps and longer time to purify (11). The subunit molecular weight of the cloned enzyme (25,250 Da) (11). However, the human erythrocyte enzyme was reported to be a monomer, which contrasts with the dimeric structure determined by crystallography. In other studies not shown, we found that glutaraldehyde cross-linking of the recombinant protein generated dimers on SDS-PAGE, thus supporting the crystallographic structure. Further studies are clearly required to confirm the quaternary structure of GGCT in vivo.

The amino acid sequence and the crystal structure of GGCT are unlike those of any other functionally characterized protein. The topology is unrelated to other known enzyme structures associated with cyclotransferase-like activity, viz. papaya glutaminyl cyclase (32) and human glutaminyl cyclase (31). The fold is classified as the “BtrG-like” family in SCOP (33), a small family that includes hypothetical proteins from Mus musculus (Protein Data Bank code 1VKB), E. coli (Protein Data Bank code 1XHS), Pyrococcus horikoshii (Protein Data Bank code 1V30), and Arabidopsis thaliana (Protein Data Bank code 2G0Q). The structures of the five known BtrG-like proteins are compared in Fig. 9. Structural conservation of GGCT with these proteins spans strands B1–B5. All of these proteins feature a pocket in a topologically conserved location (proposed to be the active site in GGCT); however, only 1VKB and 2G0Q have a residue equivalent to Glu98. In 1XHS, Glu98 is replaced with an arginine, and there is no equivalent residue in 1V30 due to helix α2 being foreshortened in that structure. It is therefore unlikely that the catalytic activity of GGCT is conserved or even similar in all structural homologues. Residues Tyr22 and Tyr125 (GGCT sequence) line the binding pocket and are conserved in all cases. Tyr125 is conserved or conservatively substituted (for tryptophan in 1V30). A short loop connecting strand B1 with helix 310 (Tyr22–Ser24 in GGCT) contains backbone amide nitrogen atoms that are oriented into the pocket and interact with substrate in our model (Fig. 8). This loop is structurally conserved in all homologues. 1V30 has a 2-[N-cyclohexylamino]ethane sulfonic acid group bound in its pocket with the SO3 group overlapping the acetate group that interacts with Asn22 and Tyr25 in the GGCT E9Q structure. The 1VKB structure has a formate group bound in its pocket, corresponding to the aforementioned SO3 and acetate groups in 1V30 and GGCT, respectively. This suggests that the pocket may have a conserved binding function in this structural family of proteins. The loops surrounding the proposed active site are poorly conserved and may reflect a diversity of function.

It is also of interest to consider the possible catalytic role of the BtrG protein that gave rise to the “BtrG-like” name that was originally proposed for this family of proteins (34). Butirosins are naturally occurring aminoglycoside antibiotics produced by Bacillus circulans. The synthesis involves the γ-glutamylation and deglutamylation of an acyl carrier protein in what appears to be an example of naturally occurring protective group chemistry. The butirosin synthesis is catalyzed by the products of a gene cluster (btrA–btrW), and the specific function of many of these genes has been assigned (34). Although it has not been experimentally demonstrated, it has been proposed that the BtrG protein may catalyze either the transfer of an acyl chain to ribostamycin or the deglutamylation of the acyl-protein intermediate. In view of our present results, the latter reaction seems most likely, since we have shown that GGCT has a fold similar to that of the BtrG-like proteins and catalyzes the cleavage of γ-glutamyl dipeptides.

GGCT appears to be the first protein of the “BtrG-like” folding family to have an unequivocal function assigned to it, and we propose that these structures should now be regarded as having the γ-glutamyl cyclotransferase fold.

The glutaminyl cyclases catalyze the post-translational formation of NH2-terminal pyroglutamate on a number of bioactive neuropeptides, hormones, and cytokines (31). This is a reaction similar to that carried out by GGCT, but the structure of GGCT is clearly unrelated to that of the glutaminyl cyclases (31, 32). The glutaminyl cyclases are structurally related to the zinc-dependent exopeptidases that catalyze a reaction that is also reminiscent of the cleavage of the γ-glu-
Despite their lack of structural homology with GGCT, we considered the possibility that there may be similarities in the reaction mechanism. The glutaminyl cyclases and the related zinc-dependent exopeptidases employ both a glutamyl residue as a general acid base and a zinc atom in the catalytic cycle (35, 36). We identified Glu\textsuperscript{98} within an accessible pocket as a potential active site residue. Mutation of Glu\textsuperscript{98} to Ala or Gln completely eliminated activity, strongly suggesting that this residue is involved in catalysis. The pocket also contains some electron density that may be a reaction product but cannot be unequivocally identified at this resolution. It also seems unlikely that this electron density represents a metal ion, since there are insufficient negatively charged residues in the pocket to provide the appropriate coordination. Previous studies have indicated that guinea pig GGCT in various tissues was stimulated by optimal concentrations of K\textsuperscript{+} and Mg\textsuperscript{2+} (37). However, human GGCT purified from erythrocytes was not stimulated or inhibited by up to 15 mM K\textsuperscript{+}, Mn\textsuperscript{2+}, or Mg\textsuperscript{2+} (11). The reaction mechanism proposed in Fig. 7 does not require a metal ion, and mutagenic studies are consistent with the view that Glu\textsuperscript{98} is the primary catalytic residue.

Despite many attempts under a variety of conditions, we were unable to obtain a structure containing a substrate or product in the active site. Based on the wild type and mutant structures as well as the mutagenesis results, we developed a model of substrate binding shown in Fig. 8. This model is consistent with the mutagenesis data and explains the absence of observed substrate binding in the mutants. In our model, the negatively charged Glu\textsuperscript{98} is needed to complement the positively charged amine group of the substrate. Mutation to neutral Ala or Gln would require the energetically unfavorable burial of a positive charge upon substrate binding. The model also removes the need for a metal ion to stabilize the oxyanion in the intermediate. From our model, the amino group of γ-Glu is expected to attack the \textit{Re}-face of the peptide link to cysteine, creating a chiral center in the intermediate with an \textit{R}-configuration.
Since this is the first identification of a cDNA encoding a GGCT enzyme, we used a bioinformatics approach to search for related sequences in humans and other species. This analysis revealed related sequences in a range of species from *C. elegans* to mammals. It was notable that although GGCT activity has been detected in plants (30), related sequences were not evident in plants, yeast, or other microbial species. However, since the BtrG protein from *B. circulans* appears to catalyze a similar reaction (34) yet shares no significant sequence identity with GGCT, it is possible that the activity in plants may be derived from a protein belonging to the same folding family but with a highly diverged sequence. A previous study of GGCT from human erythrocytes revealed multiple isoenzymes that could have been the products of separate genetic loci (12). Searches of the human EST database have only identified additional examples of the same cDNA coding sequence, suggesting that there is only one GGCT enzyme expressed in humans. This view is supported by the observation that only one functional gene could be identified within the human genome, and the sequence of exons within that gene corresponded exactly with the cDNA sequence. Although related sequences were identified on chromosomes 5 and 20, the lack of introns and the number of base changes suggests that these are nonfunctional reverse transcribed pseudogenes. It is possible that other members of the same folding family with distantly related sequences may occur and could catalyze a similar reaction. In this regard, we identified a human homologue (accession BU156875) of the mouse BtrG-like protein (1VKB) that has less than 10% identity with GGCT. Further studies are required to determine if this protein has GGCT-like activity. Alternate splicing can also result in the expression of multiple isoenzymes from a single gene. In this case, we found evidence of potential alternate splicing within the 5′-untranslated region (see Fig. 3), but this is unlikely to cause a change in the translated enzyme. The results so far suggest that the multiple GGCT isoenzymes found in human erythrocytes are most likely to be due to a post-translational modification. A similar post-translational modification may also occur in other species, since multiple isoenzymes were also detected in rat and mouse tissues (10, 13).

Previous studies have identified GGCT in human brain (8), human erythrocytes (11), and a range of rat and mouse tissues (9, 10, 13). To gain a greater understanding of the tissues in which GGCT might be expressed, we analyzed the occurrence of GGCT transcripts in a wide range of tissues represented in the EST database. This analysis (Table 5) revealed moderate levels of GGCT mRNA expression in a wide range of tissues, including the liver and kidney, which had shown the highest levels of protein expression in rats (10). In contrast, there were relatively very high (5×) levels of GGCT mRNA expression in the bladder and in the salivary gland. It is not immediately obvious what these two tissues have in common that requires this high level of expression. Because GGCT is an essential step in the γ-glutamyl cycle, we considered it likely that there may be similar levels of expression of other γ-glutamyl cycle enzymes in the same tissues. However, there were no transcripts of γ-glutamyltranspeptidase in the EST database from either the bladder or the salivary gland. This observation suggests that there may be another source of γ-glutamyl dipeptides in those tissues or that GGCT has another as yet undiscovered activity or function.

In summary, we have identified a cDNA and the gene encoding human γ-glutamylcyclotransferase. The heterologous expression of active GGCT allowed the determination of its novel crystal structure and the preliminary analysis of its reaction mechanism. These studies will finally allow the elucidation of the role of GGCT in glutathione homeostasis and will permit the design of specific inhibitors that may be of value in the treatment of patients with glutathione synthetase deficiency or in manipulating cancer cell proliferation.

Acknowledgments—We thank Dr. Youngchang Kim and staff at Structural Biology Center Collaborative Access Team for assistance with beamline operations. Use of the Structural Biology Centre beamlines at the Advanced Photon Source was supported by the United States Department of Energy, Office of Biological and Environmental Research, under Contract W-31-109-ENG-38. Some of this research was undertaken on the protein crystallography beamline at the Australian Synchrotron (Victoria, Australia).

REFERENCES

1. Zhang, C., Li, H. R., Fan, J. B., Wang-Rodriguez, J., Downs, T., Xu, X. D., and Zhang, M. Q. (2006) *BMC Bioinformatics* 7, 202
2. Kageyama, S., Iwaki, H., Inoue, H., Isono, T., Yuasa, T., Nogawa, M., Maekawa, T., Ueda, M., Kajita, Y., Ogawa, O., Toguchida, J., and Yoshiki, T. (2007) *Proteomics-Clin. Applications* 1, 192–199
3. Xu, L., Geman, D., and Winslow, R. L. (2007) *BMC Bioinformatics* 8, 275
4. Meister, A. (1973) *Science* 180, 33–39
5. Meister, A., and Anderson, M. E. (1983) *Annu. Rev. Biochem.* 52, 711–760
6. Wellner, V. P., Sekura, R., Meister, A., and Larsson, A. (1974) *Proc. Natl. Acad. Sci. U. S. A.* 71, 2505–2509
7. Ristoff, E., and Larsson, A. (1998) *Chem. Biol. Interact.* 111, 113–121
8. Orlowski, M., Richman, P. G., and Meister, A. (1969) *Biochemistry* 8, 1048–1055
9. Orlowski, M., and Meister, A. (1971) *Biochemistry* 10, 372–380
10. Orlowski, M., and Meister, A. (1973) *J. Biol. Chem.* 248, 2836–2844
11. Board, P. G., Moore, K. A., and Smith, J. E. (1978) *Biochem. J.* 173, 427–431
12. Board, P. G. (1980) *Hum. Hered.* 30, 248–250
13. Tulchin, N., and Taylor, B. A. (1981) *Genetics* 99, 109–116
14. Board, P. G., Smith, J. E., and Moore, K. (1978) *J. Lab. Clin. Med.* 91, 127–131
15. Catanzariti, A. M., Soboleva, T. A., Jans, D. A., Board, P. G., and Baker, R. T. (2004) *Protein Sci.* 13, 1331–1339
16. Van Duyn, G. D., Standera, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) *J. Mol. Biol.* 229, 105–124
17. Towbin, H., Staehelin, T., and Gordon, J. (1979) *Proc. Natl. Acad. Sci. U. S. A.* 76, 4350–4354
18. Minor, W., Cyborski, M., Otwinowski, Z., and Chruszcz, M. (2006) *Acta Crystallogr. Sect. D Biol. Crystallogr.* 62, 859–866
19. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr. Sect. D Biol. Crystallogr.* 58, 1772–1779
20. Tsvendorf, T. (2004) *J. Synchrotron Radiat.* 11, 49–52
21. Bailey, S. (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* 50, 760–763
22. Cowtan, K. D., and Main, P. (1996) *Acta Crystallogr. Sect. D Biol. Crystallogr.* 52, 43–48
23. Leslie, A. G. W. (1999) *Acta Crystallogr. D Biol. Crystallogr.* 55, 1696–1702
24. Evans, P. R. (2006) *Acta Crystallogr. D Biol. Crystallogr.* 62, 72–82
25. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D Biol. Crystallogr.* 53, 240–255
26. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr.* 30, 565–572

Structure of γ-Glutamyl Cyclotransferase

In summary, we have identified a cDNA and the gene encoding human γ-glutamylcyclotransferase. The heterologous expression of active GGCT allowed the determination of its novel crystal structure and the preliminary analysis of its reaction mechanism. These studies will finally allow the elucidation of the role of GGCT in glutathione homeostasis and will permit the design of specific inhibitors that may be of value in the treatment of patients with glutathione synthetase deficiency or in manipulating cancer cell proliferation.
Structure of γ-Glutamyl Cyclotransferase

27. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) J. Appl. Crystallogr. 26, 283–291
28. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Nucleic Acids Res. 25, 3389–3402
29. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Nucleic Acids Res. 22, 4673–4680
30. Steinkamp, R., Schweinhofen, B., and Rennenberg, H. (1987) Physiol. Plant. 69, 499–503
31. Huang, K. F., Liu, Y. L., Cheng, W. J., Ko, T. P., and Wang, A. H. (2005) Proc. Natl. Acad. Sci. U. S. A. 102, 13117–13122
32. Wintjens, R., Belrachi, H., Clantin, B., Azarkan, M., Bompad, C., Baeyens-Volant, D., Looze, Y., and Villeret, V. (2006) J. Mol. Biol. 357, 457–470
33. Hubbard, T. J. P., Murzin, A. G., Brenner, S. E., and Chothia, C. (1997) Nucleic Acids Res. 25, 236–239
34. Li, Y., Llewellyn, N. M., Giri, R., Huang, F., and Spencer, J. B. (2005) Chem. Biol. 12, 665–675
35. Bzymek, K. P., and Holz, R. C. (2004) J. Biol. Chem. 279, 31018–31025
36. Fundoiano-Herschcovitz, Y., Rabinovitch, L., Langut, Y., Reiland, V., Shoham, G., and Shoham, Y. (2004) FEBS Lett. 571, 192–196
37. Woltersberger, M. G., and Tabachnick, J. (1974) J. Invest. Dermatol. 62, 587–590