Cloning, structure and mRNA expression of human Cctg, which encodes the chaperonin subunit CCT γ

Neal A. WALKLEY,* Andrew G. DEMAINE† and Afshan N. MALIK*‡

*School of Molecular and Medical Biosciences, University of Wales College of Cardiff, Cardiff, and †Department of Medicine, Faculty of Postgraduate Medicine, University of Plymouth, Plymouth, U.K.

We describe the cloning, DNA sequence analysis and mRNA expression analysis of human Cctg (HsCctg), a gene that encodes the γ -subunit of the eukaryotic cytosolic 'chaperonin-containing TCP-1' (CCT). Partial clones representing the 3' region of HsCctg cDNA were isolated from a human kidney cDNA library, and the missing 5' region was amplified directly from human kidney cDNA. The Cctg mRNA transcript is expressed in numerous human and mouse tissues and, like Tcp-1/Cctg, Cctg mRNA is expressed at higher levels in mouse testis when compared with kidney and brain. Southern-blot analysis has also revealed the Cctg gene to be highly conserved in mouse, rat, sheep and frog. The 1901 bp HsCctg cDNA has a coding region

of 1635 bp and encodes a predicted 60 kDa protein (544 amino acids). The predicted HsCCT γ amino acid sequence shares a high degree of sequence similarity with γ -subunits from the mouse *Mus musculus* (98 % similarity), the yeast *Saccharomyces cerevisiae* (75 % similarity) and the protozoan *Tetrahymena pyriformis* (76 % similarity) as well as with other members of the TF55/TCP-1 family, such as human TCP-1/CCT α (55 % similarity) and TCP-20/CCT ζ (54 % similarity). HsCCT γ also shares conserved domains previously identified in the TF55/TCP-1 family of chaperonins and more distantly related chaperonins such as GroEL and Hsp60.

INTRODUCTION

Chaperonins are a family of ubiquitously and abundantly expressed sequence-related molecular chaperones that mediate the ATP-dependent folding of monomeric and oligomeric subunits into fully functional native proteins within the cell [1]. Two major families of chaperonins have been identified: the GroEL family found in eubacteria, plastids and mitochondria [1], and the TF55/TCP-1 family found in archaeobacteria and the eukaryotic cytosol [2]. Some chaperonins, such as GroEL, may be heat-induced, but studies at the genetic level indicate that chaperonins are essential for protein folding at all temperatures, and they have been described as essential for cell viability (for reviews, see [3–5]).

T complex polypeptide 1 (TCP-1), the first member of the TF55/TCP-1 family of chaperonins to be identified, was proposed as a candidate eukaryotic chaperonin on the basis of weak sequence similarities with GroEL [6,7]. Subsequently TCP-1 was found to share greater sequence similarity with TF55, an archaeobacterial stress protein which displays chaperonin-like properties [2].

The eukaryotic chaperonin complex of approx. 900 kDa termed 'TCP-1 ring complex' (TRiC) or 'chaperonin containing TCP-1' (CCT) has been shown to consist of eight or nine structurally related subunits of 52–65 kDa [8,9]. The subunits are arranged radially and form two stacked rings with a central cavity [10]. TCP-1 was identified as one of the constituent subunits of both bovine [8] and murine [9] chaperonin complexes.

Kubota et al. [9] reported the cloning and DNA sequencing of a further seven murine cDNA clones which represented CCT

subunits β , γ , δ , ϵ , ζ , η and θ [9,11]. The subunits measured between 531 and 548 residues in length and shared 25–36 % sequence identity with one another. The cDNA clone representing murine CCT γ (*MmCCT γ*) was cloned on the basis of sequence similarity to the partial human cDNA clone designated HumAP4 which was originally isolated during attempts to clone the Na⁺/H⁺ antiporter gene [12]. Recently γ -subunits have been identified in the yeast *Saccharomyces cerevisiae* and the protozoan *Tetrahymena pyriformis* which share approx. 32 % identity with TCP-1 at the amino acid level [13,14].

The involvement of the CCT complex in ATP-dependent protein folding has been demonstrated *in vitro*. Newly translated tubulin subunits were shown to associate with an approx. 900 kDa complex, which by immunoblot analysis was shown to contain TCP-1. After several minutes the subunits were found either as free monomers or α / β -tubulin heterodimers [15]. CCT has also been shown to be involved in the folding of actin [16] and firefly luciferase [8].

To date only two members of the human CCT complex, namely TCP-1/CCT α [17] and TCP-20/CCT ζ [18], have been identified and fully DNA-sequenced. A partial 119 bp cDNA representing a third subunit, CCT, has also been recently cloned and sequenced [11]. In the present paper we describe a fourth human subunit, namely HsCCT γ . We describe the cloning of HumAP4 and the corresponding full-length HsCctg cDNA, expression patterns of HsCctg mRNA in human tissues, elevated expression of *MmCctg* mRNA in mouse testis, the conservation of this gene in a divergent group of organisms and we also present the full-length HsCctg cDNA sequence along with the predicted HsCCT γ peptide sequence.

Abbreviations used: CCT, chaperonin containing TCP-1; TCP-1, T complex polypeptide-1; TRiC, TCP-1 ring complex; CCT γ , γ -subunit of CCT; Cctg, gene encoding CCT γ ; HsCctg/HsCCT γ , Cctg/CCT γ from human; *MmCctg/MmCCT γ* , Cctg/CCT γ from *Mus musculus*; *TpCctg/TpCCT γ* , Cctg/CCT γ from *Tetrahymena pyriformis*; *XlCctg/XlCCT γ* , Cctg/CCT γ from *Xenopus laevis*; NHA, Na⁺/H⁺ antiporter; Rubisco, ribulose biphosphate carboxylase.

‡ To whom correspondence should be addressed.

The nucleotide sequence of human Cctg cDNA has been deposited with the EMBL/GenBank/DBJ Nucleotide Sequence Databases under the accession number X74801.

MATERIALS AND METHODS

RNA and DNA isolation

Total RNA was isolated from cultured human skin fibroblasts, mouse and rat tissues, human foetal tissues (MRC Tissue Bank) and adult kidneys (post-mortem material; Kings College Hospital, London) by using a guanidinium isothiocyanate/phenol method [19] with the modifications described previously [20]. Genomic DNA was prepared from human placenta, mouse liver, rat kidney, sheep liver, frog liver, stone-loach liver and pike liver [21]. Plasmid DNA was prepared using Wizard Minipreps (Promega). Single-stranded M13 DNA was prepared using a poly(ethylene glycol) precipitation method [21] and PCR products were excised from agarose gels and purified using the QIAquick gel extraction kit (Qiagen).

Amplification of NHA5-7

Double-stranded cDNA was synthesized from 10 µg of total RNA isolated from cultured human skin fibroblasts and used as template for PCR amplification. Primers NHA1-5 and NHA1-7 (Table 1), designed against the 3' cytoplasmic region of the Na⁺/H⁺ antiporter gene (NHA1) [22], were used to amplify the PCR product NHA5-7 in a Perkin-Elmer/Cetus thermal cycler by denaturation at 94 °C for 1 min, annealing at 55 °C for 2 min and extension at 72 °C for 4 min for a total of 35 cycles.

Cloning of HumAP4 cDNA from human kidney

A human kidney cDNA library was constructed (cDNA synthesis and cloning kits; Amersham International) in the cDNA cloning vector lambda gt10 [23] and approx. 200 000 unamplified cDNA clones were probed with [α -³²P]CTP-labelled NHA5-7 [24]. Hybridization was carried out at 65 °C in HYBN buffer [6 × SSC (1 × SSC is 0.15 M NaCl/0.015 M sodium citrate)/5 × Denhardt's/0.5% SDS/50 µg/ml sonicated and denatured salmon sperm DNA]. The filters were stringently washed (0.1 × SSC/0.1% SDS, 65 °C, 40 min) prior to autoradiography. Positive clones were picked and purified by three further rounds of screening. Phage DNA was prepared [25] and digested with

the restriction endonuclease *Eco*RI. The cDNA inserts were subcloned into the *Eco*RI site of the phage vectors M13mp18 and M13mp19 and/or pBluescript SKII+ for DNA sequencing.

Amplification and cloning of 5' region of HsCctg from human kidney

Total RNA (10 µg in a volume of 20 µl) from foetal kidney was denatured (65 °C, 10 min) prior to adding to a reverse-transcription mixture containing 1 mM 4 × dNTPs, 1 × RT buffer [50 mM Tris/HCl (pH 8.0)/5 mM MgCl₂/5 mM dithiothreitol/50 mM KCl/BSA (µg/ml)], 975 mM dithiothreitol, 100 ng of random hexamers or oligo-dT and 50 units of Moloney-murine-leukaemia-virus reverse transcriptase (USB, Cambridge Bioscience). The volume was adjusted to 40 µl with double-distilled water and incubated (37 °C, 1 h). After enzyme inactivation by heat shock (95 °C, 5 min), the template was used for PCR.

Approx. 1/20th of the cDNA (2 µl) was used as template for amplification using degenerate 5' primers Hg-15 or Hg-16 in combination with either Hg-14 or Hg-8 (Table 1). Amplification conditions used were as follows: denaturation, 94 °C, 1 min; annealing, 57 °C, 1.5 min; and extension, 72 °C, 2 min for a total of 30 cycles in the presence of 3 mM MgCl₂ and 100 ng of each primer. PCR products were revealed on 1% agarose gels and cloned directly into the pCRII vector (Invitrogen) or the pCR-Script SK+ vector (Stratagene).

DNA sequencing and analysis of HsCctg

A number of independently isolated cDNA clones spanning the complete coding region of HsCctg were used to determine the full nucleotide sequence of HsCctg (Table 2). Sequencing was carried out for most clones with both the sense and antisense strands using the Sequenase v2.0 sequencing kit (Cambridge Bioscience), M13 forward (M13 F) and reverse (M13 R) primers and the internal Hg primers listed in Table 1. The entire HsCctg cDNA sequence was determined in both directions.

Sequence alignments and database searches to establish sequence similarities with other published sequences were carried out using the University of Wisconsin Genetic Computing group (GCG) packages [26]. Programs used included Wordsearch for

Table 1 Oligonucleotide primers used for amplification and DNA sequencing of HsCctg

All primers are shown in the 5' to 3' direction and are labelled sense (S) or anti-sense (AS). The positions of the primers are indicated with respect to either the published Na⁺/H⁺ antiporter sequence or to the full HsCctg sequence. Y represents T or C, R represents A or G, and I represents an inosine.

Primer	5'						3'		Position	S/AS	Reference
M13 F	GTT	TTC	CCA	GTC	ACG	AC					
M13 R	CAG	GAA	ACA	GCT	ATG	AC					
NHA1-5	AGA	CAC	ACG	CTG	GGG	TGG	CAG	A	1951–1970	S	[12]
NHA1-7	CAC	CAG	CCC	CAG	CAG	TT C	TG		2665–2684	AS	[12]
Hg-1	CTC	TGC	TGG	TTC	TGT	GCA	TTG		1655–1675	AS	The present paper
Hg-2	CAG	GCC	TGT	TGG	AAA	TCA	AGA		1037–1057	S	The present paper
Hg-4	CAG	CAG	TGG	AGA	CGG	CAG	TT		1520–1539	S	The present paper
Hg-5	GTT	CTC	CTG	GGT	GTG	CTT	GG		1400–1419	AS	The present paper
Hg-6	TGA	TTT	CCA	ACA	GGC	CTG			1037–1054	AS	The present paper
Hg-8	TGA	GCT	AAA	TCT	GAG	ATG	CCC	T	878–899	AS	The present paper
Hg-14	GTA	CTC	TTC	CTC	CAT	CTG			802–819	AS	The present paper
Hg-15	ATG	GGI	CAY	CGI	CCI	GT			1–17	S	[9]
Hg-16	ATG	GGI	CAY	AGR	CCI	GT			1–17	S	[9]
Hg-17	GCC	ATT	CTT	CGA	GAG	AT			190–206	S	The present paper
Hg-18	CAA	TGT	CAG	TCT	GGC	TT			766–750	AS	The present paper
Hg-20	A CT	GTC	ACT	GAT	GTC	GA			428–444	AS	The present paper
Hg-21	ATC	AGC	CGG	TGG	TCA	TC			490–506	S	The present paper
Hg-22	CTG	ATC	CAG	AAC	TGT	GG			1348–1364	S	The present paper

Table 2 Partial cDNA clones used to determine the full nucleotide sequence of HsCctg

cDNA clones were isolated from: ¹oligo-dT-primed cDNA library in lambda gt10, ²random-primed cDNA library in lambda gt10, ³small length cDNA library in pBluescript (UK Human Genome Mapping Resource Centre) and ⁴oligo-dT-primed first-strand cDNA.

Clone name	Source	Insert size	Position in HsCctg
HumAP4	Adult kidney cDNA library ¹	1108 bp	794–1855 bp
HumAP3	Adult kidney cDNA library ¹	1059 bp	794–1901 bp
pBS-NW2	Adult kidney cDNA library ²	839 p	1041–1879 bp
S3/H9	Foetal Brain cDNA library ³	497 bp	1359–1855 bp
TA15	Foetal kidney cDNA ⁴	899 bp	1–899 bp
pCRS16	Foetal kidney cDNA ⁴	819 bp	1–819 bp
Pg-2	Foetal spleen cDNA ⁴	577 bp	190–766 bp

database searching, Bestfit for comparisons of two DNA or amino acid sequences and Clustal for alignment of more than two sequences. Percentage similarities and percentage identities stated in the present paper were determined using Bestfit. The term 'percentage identity' is used to imply that two sequences, when compared, share x % identical residues. The term 'percentage similarity' takes into account the presence of both identical and conserved amino acid changes.

Northern- and Southern-blot analysis

For Northern-blot analysis, approx. 10 μ g of total RNA was loaded on a 5.5 %-formaldehyde/1 %-agarose gel and electrophoresed in Mops buffer [21] at 70 V for 5 h, followed by capillary blotting in 20 $\times$ SSC on to Hybond nfp nylon membranes (Amersham International). The membranes were UV-fixed and hybridized to random-primed [α -³²P]CTP probes [24] in NORTHYB buffer (50 % formamide, 5 $\times$ SSC, 50 mM sodium phosphate buffer, pH 6.5, 10 $\times$ Denhardt's solution and 350 μ g/ml denatured sonicated salmon sperm DNA) at 42 °C in the presence of 12.5 % (v/v) dextran sulphate. The filters were washed to a final stringency of 0.1 or 0.2 $\times$ SSC/0 $\times$ 1 % SDS at 65 °C before autoradiography at –70 °C for 1–5 days. The integrity and the relative amounts of RNA loaded was determined by stripping the filter after hybridization and reprobing with β -actin as described previously [20].

For Southern-blot analysis, genomic DNA was prepared [21] using tissues from various species (human, rat, mouse, sheep and frog). Approx. 10 μ g of DNA was digested with *Eco*RI, electrophoresed on a 0.8 %-agarose gel, transferred to Hybond nfp and hybridized to random-primed [α -³²P]CTP probes [24] at 65 °C in HYBN buffer. Filters were washed as described for the library filters and Northern-blot above.

Amplification of HsCctg using PCR

Two sources of template were used for amplification of HsCctg mRNA, namely cDNA libraries for all adult tissues and first-strand cDNA synthesized from total RNA for all foetal tissues. For adult libraries approx. 10⁶ plaque-forming units were used for each PCR. For foetal tissues, first-strand cDNA was used for PCR as described above under 'Amplification and cloning of 5' region of HsCctg from human kidney'. Reaction mixtures containing appropriate templates, primers Hg-1 (100 ng) and Hg-2 (100 ng), 1 $\times$ reaction buffer and 3 mM MgCl₂ in a final volume of 25 μ l were used to amplify a 638 bp HsCctg fragment. Amplification was carried out as described previously [25], using the following conditions: denaturing, 94 °C, 1 min; annealing, 57 °C, 1.5 min; and extension, 72 °C, 2 min, for a total of 30

cycles. In addition, negative water controls were included in all PCR experiments to ensure that the products were not due to contamination.

In order to confirm that the correct PCR products were being amplified, the products were digested with the restriction enzyme *Pst*I, which should cleave the PCR product yielding two bands of 167 and 471 bp.

RESULTS AND DISCUSSION

Cloning of HsCctg cDNA

Cloning of HumAP4, the 3' region of HsCctg

HumAP4 was cloned from a human kidney cDNA library on the basis of sequence similarity to the PCR product obtained using primers designed against the 3' cytoplasmic domain of the Na⁺/H⁺ antiporter gene (NHA1) [22]. Using primers NHA1-5 and NHA1-7, two main PCR products, of 580 bp and 620 bp, described hereafter as NHA5-7, were obtained from cultured human skin fibroblasts and were used to probe a human kidney cDNA library. From 22 hybridizing clones, 16 clones (AP1–AP16) were further analysed. Digestion with *Eco*RI revealed inserts of approx. 0.3 kb (AP1), 1.0 kb (AP8), 1.1 kb (AP3, AP7, AP11–16) and 1.2 kb (AP4, AP6). All of the clones cross-hybridized strongly with AP4, and partial sequencing and subsequent sequence analysis revealed that the clones represented the same gene. In Southern blots the *Eco*RI inserts from the AP clones hybridized strongly to NHA5-7, but not to NHA1 (results not shown). The *Eco*RI insert from AP4 was subcloned into pBluescript SKII+ and designated 'HumAP4'. Partial DNA sequencing and similarity searching of databases showed that the predicted HumAP4 peptide sequence shared significant sequence similarity with TCP-1/CCT. This was the first indication that we had cloned a gene which was distinct from the NHA1 gene. The reasons for the presence of chaperonin-related sequences in the NHA5-7 PCR product remain unclear.

Cloning of the 5' region of HsCctg

Although HumAP4 contained an insert of 1108 bp, Northern-blot analysis revealed a single mRNA transcript of approx. 20 kb which was abundantly expressed in human foetal kidney. These data indicated that only partial cDNA clones had been isolated. Partial DNA sequencing revealed the presence of a poly(A) tail at the 3' region of HumAP4, as well as an internal *Eco*RI site at the 5' region, confirming HumAP4 to be partial in length.

Kubota et al. [9] used the partial DNA sequence of HumAP4 to isolate the murine *Cctg* (*MmCctg*) cDNA by PCR ampli-

Figure 1 Amplification of the 5' region of HsCctg and *MmCctg*

Amplified products obtained by RT-PCR from human foetal kidney (lanes 2 and 3) and mouse testis (lanes 4 and 5) first-strand cDNA. The degenerate primer Hg-15 was used in combination with primers Hg-14 and Hg-8 to amplify the 819 bp and 899 bp fragments respectively. Lane 1 shows the 1 kb DNA ladder.

fication and subsequent cDNA library screening. We in turn used the predicted peptide sequence of *MmCCTγ* to design degenerate primers to amplify the missing 5' region of HsCctg. Alignment of the predicted amino acid sequence of HumAP4 with the peptide sequence of *MmCCTγ* using Bestfit showed that the two sequences shared 98 % similarity. We assumed that the 5' regions of *MmCCTγ* and HsCCTγ would also share an equal degree of similarity. Therefore we reverse-translated the amino acid sequence of *MmCCTγ* in order to design degenerate primers corresponding to the 5' region. MGHRPV was the region chosen, which corresponded to amino acid residues 2–7 of the predicted *MmCCTγ* sequence. In order to allow for all the possible nucleotide combinations coding for the arginine residue (R) at position 5, two degenerate primers, Hg-15 and Hg-16 (Table 1) were designed. Depending on the actual DNA sequence of

HsCctg, we expected only one of these degenerate primers to yield a product when used for PCR.

Hg-15 and Hg-16 were used in conjunction with two staggered antisense primers, Hg-8 and Hg-14, designed against the 5' region of HumAP4 to amplify the missing 5' region of HsCctg. cDNA prepared from human foetal kidney total RNA was used as a template for the amplification, while cDNA prepared from mouse testis total RNA was used as a control for the primers.

No PCR products were detected using primer Hg-16 for either template. The foetal kidney template yielded PCR products of 899 bp and 819 bp when Hg-15 was used with Hg-8 and Hg-14 respectively (Figure 1). Since primers Hg-8 and Hg-14 are staggered by 80 bp and the PCR products differed by approx. 80 bp, this indicated that we had probably amplified the 5' region of HsCctg (Figure 1). In addition, the same primer combinations yielded identically sized products from mouse testis, also suggesting that we had amplified the 5' region of *MmCctg* from mouse testis. The 899 bp and the 819 bp PCR products from mouse testis were cloned and fully sequenced in both directions, confirming that we had amplified and cloned the 5' region of *MmCctg* (results not shown).

The 899 bp and the 819 bp PCR products from human foetal kidney resulting from three independent PCR amplifications were also cloned and used for the determination of the full DNA sequence of HsCctg.

DNA sequencing of HsCctg

The entire HsCctg cDNA sequence was determined for both strands. The strategy used in obtaining the full DNA sequence of HsCctg is illustrated in Figure 2. Briefly, the 1108 bp at the 3' region was sequenced using cDNA clones isolated from kidney cDNA libraries (Table 2). The remaining 5' sequence was determined using at least three independently cloned PCR products which had been cloned as described above.

The DNA sequence of HsCctg (Figure 3) comprises 1901 bp and encodes a predicted 60 kDa peptide consisting of 544 amino acids. A 36 bp poly(A) tail is present at the 3' region 14 bp

Figure 2 Sequencing strategy of HsCctg

TA15 and HumAP4 are represented by the unfilled boxes, and vector sequences are represented by the filled boxes. Sense and antisense primers are shown above and below the DNA respectively. TA15 and HumAP4 overlap by 106 bp, and this region is shown between the dotted lines. The primer used and the region sequenced are shown by arrows. Each arrow represents a minimum of three independent sequencing reactions. Several other independent clones, e.g. pBS-NW2, HumAP3, pCRS16 and Pg-2, were also used to confirm the sequence of HsCctg (not shown).

ATG GGG CAT CGG CCG GTG CTC GTG CTC AGC CAG AAC ACA AAG CGT GAA TCC GGA AGA AAA GTT CAA TCT GGA	72
M G H R P V L V L S Q N T K R E S G R K V Q S G	
AAC ATC AAT GCT GCC AAG ACT ATT GCA GAT ATC ATC CGA ACA TGT TTG GGA CCC AAG TCC ATG ATG AAG ATG	144
N I N A A K T I A D I I R T C L G P K S M M K M	
CTT TTG GAC CCA ATG GGA GGC ATT GTG ATG ACC AAT GAT GGC AAT GCC ATT CTT CGA GAG ATT CAA GTC CAG	216
L L D P M G G I V M T N D G N A I L R E I Q V Q	
CAT CCA GCG GCC AAG TCC ATG ATC GAA ATT AGC CGG ACC CAG GAT GAA GAA GTT GGA GAT GGG ACC ACA TCA	288
H P A A K S M I E I S R T Q D E E V G D G T T S	
GTA ATT ATT CTT GCA GGG GAA ATG CTG TCT GTA GCT GAG CAC TTC CTG GAG CAG CAG ATG CAC CCA ACA GTG	360
V I I L A G E M L S V A E H F L E Q Q M H P T V	
GTG ATC AGT GCT TAC CGC AAG GCA TTG GAT GAT ATG ATC AGC ACC CTA AAG AAA ATA AGT ATC CCA GTC GAC	432
V I S A Y R K A L D D M I S T L K K I S I P V D	
ATC AGT GAC AGT GAT ATG ATG CTG AAC ATC ATC AAC AGC TCT ATT ACT ACC AAA GCC ATC AGC CGG TGG TCA	504
I S D S D M M L N I I N S S I T T K A I S R W S	
TCT TTG GCT TGC AAC ATT GCC CTG GAT GCT GTC AAG ATG GTA CAG TTT GAG GAG AAT GGT CGG AAA GAG ATT	576
S L A C N I A L D A V K M V Q F E E N G R K E I	
GAC ATA AAA AAA TAT GCA AGA GTG GAA AAG ATA CCT GGA GGC ATC ATT GAA GAC TCC TGT GTC TTG CGT GGA	648
D I K K Y A R V K D V T E K I P G G I I E D S C	
GTC ATG ATT AAC AAG GAT GTG ACC CAT CCA CGT ATG CGG CGC TAT ATC AAG AAC CCT CGC ATT GTG CTG CTG	720
V L R G V M I N H P R M R R Y I K N P R I V L L	
GAT TCT TCT CTG GAA TAC AAG AAA GGA GGA AGC CAG ACT GAC ATT GAG ATT ACA CGA GAG GAG GAC TTC ACC	792
D S S L E Y K K G G S Q T D I E I T R E E D F T	
CGA ATT CTC CAG ATG GAG GAA GAG TAC ATC CAG CAG CTC TGT GAG GAC ATT ATC CAA CTG AAG CCC GAT GTG	864
R I L Q M E E E Y I Q Q L C E D I I Q L K P D V	
GTC ATC ACT GAA AAG GGC ATC TCA GAT TTA GCT CAG CAC TAC CTT ATG CGG GCC AAT ATC ACA GCC ATC CGC	936
V I T E K G I S D L A Q H Y L M R A N I T A I R	
AGA GTC CGG AAG ACA GAC AAT AAT CGC ATT GCT AGA GCC TGT GGG GCC CGG ATA GTC AGC CGA CCA GAG GAA	1008
R V R K T D N N R I A R A C G A R I V S R P E E	
CTG AGA GAA GAT GAT GTT GGA ACA GGA GCA GGC CTG TTG GAA ATC AAG AAA ATT GGA GAT GAA TAC TTT ACT	1080
L R E D D V G T G A G L L E I K K I G D E Y F T	
TTC ATC ACT GAC TGC AAA GAC CCC AAG GCC TGC ACC ATT CTC CTC CGG GGG GCT AGC AAA GAG ATT CTC TCG	1152
F I T D C K D P K A C T I L L R G A S K E I L S	
GAA GTA GAA CGC AAC CTC CAG GAT GCC ATG CAA GTG TGT CGC AAT GTT CTC CTG GAC CCT CAG CTG GTG CCA	1224
E V E R N L Q D A M Q V C R N V L L D P Q L V P	
GGG GGT GGG GCC TCC GAG ATG GCT GTC GCC CAT GCC TTG ACA GAA AAA TCC AAG GCC ATG ACT GGT GTG GAA	1296
G G G A S E M A V A H A L T E K S K A M T G V E	
CAA TGG CCA TAC AGG GCT GTT GCC CAG GCC CTA GAG GTC ATT CCT CGT ACC CTG ATC CAG AAC TGT GGG GCC	1368
Q W P Y R A V A Q A L E V I P R T L I Q N C G A	
AGC ACC ATC CGT CTA CTT ACC TCC CTT CGG GCC AAG CAC ACC CAG GAG AAC TGT GAG ACC TGG GGT GTA AAT	1440
S T I R L L T S L R A K H T Q E N C E T W G V N	
GGT GAG ACG GGT ACT TTG GTG GAC ATG AAG GAA CTG GGC ATA TGG GAG CCA TTG GCT GTG AAG CTG CAG ACT	1512
G E T G T L V D M K E L G I W E P L A V K L Q T	
TAT AAG ACA GCA GTG GAG ACG GCA GTT CTG CTA CTG CGA ATT GAT GAC ATC GTT TCA GGC CAC AAA AAG AAA	1584
Y K T A V E T A V L L L R I D D I V S G H K K K	
GGC GAT GAC CAG AGC CGG CAA GGC GGG GCT CCT GAT GCT GGC CAG GAG TGA gtgctaggcaaggctacttcaatgcac	1662
G D D Q S R Q G G A P D A G Q E *	
agaaccagcagagtctcccttttctgagccagagtccaggaaacactgtggacgtctttgttcagaagggatcaggttggggggcagcccca	1757
gtccctttctgtcccagtcagttttccaaaagacactgacatgtaattctctctattgtaagggttccatttagtttgcctccgatgattaaa	1852
tctaagtcatttgaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa	1901

Figure 3 Nucleotide and deduced amino acid sequence of HsCctg cDNA

Nucleotides are numbered on the right of the sequence, and the amino acids are represented below. The coding region is shown in upper-case letters with the 3' untranslated region in lower-case letters. The putative polyadenylation signal is underlined, and the stop codon is shown at position 1635 bp.

	266		285
HsCCT γ	I L Q M E E E Y I Q Q L C E D I I Q L K		
P5/T36	N * * * * * Q * * * * * X X * * *		
	399		423
HsCCT γ	N V L L D P Q L V P G G G A S E M A V A H A L T E		
P5/T32	* * * * * * * * * * * * * * * * * * *		

Figure 4 Amino acid alignment showing the high degree of sequence similarity between bovine P5 tryptic peptides and HsCCT γ

Amino acid residues denoted by X in the tryptic peptides were not determined by Frydman et al. [8]. Amino acids identical to HsCCT γ are indicated by an asterisk.

downstream from the polyadenylation signal ATTAAA. Because of the cloning strategy employed, no sequence is currently available for the 5' untranslated region.

HsCCT γ belongs to the TF55/TCP-1 family of chaperonins

Database searches at the nucleotide and peptide levels were carried out using the University of Wisconsin Genetic Computing Group packages [26] to try and assign a possible function to HsCCT γ . Sequences showing significant sequence similarity were then compared with HsCctg using Bestfit and the results expressed as percentage identity and percentage similarity. Comparisons of HsCCT γ with known human chaperonins showed 33% identity with, and 55% similarity to, TCP-1/CCT α and 29% identity with, and 54% similarity to, TCP-20/CCT ζ . Comparisons with TF55, the major heat-shock protein of the thermophilic archaebacterium *Sulfolobus shibatae* revealed 36% identity and 59% similarity. This extent of sequence similarity strongly supports the inclusion of HsCCT γ in the TF55/TCP-1 family of chaperonins.

HsCctg encodes the γ -subunit of the CCT complex

Frydman et al. [8] purified and sequenced partial peptides corresponding to bovine TRiC subunits described as P1, P3 and P5 and showed them to be structurally related to TCP-1/CCT α . Comparison of the peptide sequence of HsCCT γ with the partial peptide sequences of bovine P5/T36 and P5/T32 revealed that HsCCT γ shared a high degree of similarity with these peptides (Figure 4). P5 was shown to be the equivalent of the CCT γ -subunit [9] and, as a result, we proposed that HsCCT γ was the human γ equivalent.

Kubota et al. [9] designed primers against HumAP4 and used them to generate a mouse cDNA PCR product which in turn was used to isolate the *MmCctg* cDNA by library screening. Alignment of the predicted peptide sequences of *MmCCT γ* and HsCCT γ reveals that they share a high degree of sequence similarity (97% identity, 98% similarity), providing further evidence that HsCCT γ represents the γ -subunit of the CCT complex. Two further γ -subunits have been reported in *Saccharomyces cerevisiae* (BIN2) [13] and *T. pyriformis* (*TpCCT γ*) [14], both of which also share significant sequence similarity with HsCCT γ (59% identity, 75% similarity and 57% identity, 76% similarity respectively). More recently we have cloned and characterized *XlCctg*, the *Xenopus laevis* Cctg (N. A. Walkley, R. A. Page and A. N. Malik, unpublished work). The predicted *XlCCT γ* shares 87% identity, 94% similarity with HsCCT γ . Thus all the γ -subunits so far described share significant sequence similarity.

Figure 5 shows multiple alignments of HsCCT γ with *MmCCT γ* , BIN2, and *TpCCT γ* using the Clustal method

(Genetics Computer Group). We identified eight stretches of amino acids which showed more than six consecutive identical residues at positions 50–58, 67–72, 76–82, 210–216, 253–258, 298–303, 305–312 and 364–369. Four stretches of 14 or more consecutive residues containing both identical and conserved residues are found at positions 41–72, 94–109, 297–314, and 320–333. This extent of similarity between the γ -subunits from different species appears to imply evolutionary conservation, as has been previously proposed for other chaperonin genes [27].

HsCCT γ retains the major domains identified in the CCT subunits, GroEL and other chaperonins. The GDGTT(S/T) domain is present in all four γ -subunits aligned in Figure 5 as well as in GroEL, mitochondrial Hsp60 and ribulose biphosphate carboxylase (Rubisco)-subunit-binding protein. This domain has been shown to have sequence similarity with the binding domain of the β -phosphate of ATP found in cyclic AMP-dependent kinase family [11] and has been proposed to be involved in ATP binding and hydrolysis [9].

The V(A/P)GGG domain is also highly conserved in the human, mouse and *T. pyriformis* γ -subunits, but shows a change from valine (V) to serine (S) in yeast. This domain has also been identified in other chaperonins, including TF55, GroEL, mitochondrial Hsp60 and Rubisco subunit binding protein.

The TNDGATIL domain, which has been found to be highly conserved in the majority of CCT subunits described so far [9], as well as in TF55 [2], also displays strong conservation in the γ -subunits occurring as TNDGNAIL in human, mouse and *T. pyriformis*, while an asparagine (N) to histidine (H) substitution at position 5 is observed in yeast, making the domain TNDGHAIL.

Cctg mRNA is expressed at higher levels in mouse testis

The higher expression of *Tcp-1/Cctg* mRNA in mouse testis relative to other tissues has been previously described [28,29], although the reasons for this are unclear. Since the *Tcp-1/Cctg* mRNA representing the *TCP-1/CCT α* subunit of the CCT complex is more highly expressed in testis, it is possible that other mRNAs representing the various members of this hetero-oligomeric CCT complex also display higher levels of steady-state mRNA in the testis. However, to date there are no reports on the relative abundances of CCT subunit mRNAs in different tissues.

Examination by Northern analysis of Cctg mRNA from mouse testis, brain and kidney was carried out (Figure 6). A 2.0 kb *MmCctg* mRNA transcript was detected in all three tissues and showed significantly higher expression in testis when compared with brain and kidney. To correct for unequal loading, the blots were reprobbed with β -actin and shown to be approximately equivalent. These data show that, like Cctg/Tcp-1 [29], Cctg also displays higher mRNA levels in mouse testis when compared with mouse brain and kidney. This higher level of *MmCctg* mRNA may reflect changes in transcriptional regulation of the gene or may be a consequence of the stability of the mRNA in different tissues.

Expression of HsCctg in human tissues

The expression of HsCctg mRNA was examined in a range of human adult and foetal tissues using PCR. For adult tissues, cDNA libraries from heart, kidney, retina and thyroid were used as templates. For foetal tissues, cDNA was synthesized using total RNA from spleen, brain, liver, heart and kidney and used as template for PCR. Using primers Hg-1 and Hg-2 (Table 1) designed against the HsCctg cDNA sequence, we amplified PCR products of the expected 638 bp in all tissues examined except in

HsCCT γ	120	121	122
MmCCT γ	120	121	122
Bin2	120	121	122
TpCCT γ	120	121	122
HsCCT γ	240	241	242
MmCCT γ	240	241	242
Bin2	240	241	242
TpCCT γ	240	241	242
HsCCT γ	360	361	362
MmCCT γ	360	361	362
Bin2	360	361	362
TpCCT γ	360	361	362
HsCCT γ	480	481	482
MmCCT γ	480	481	482
Bin2	480	481	482
TpCCT γ	480	481	482
HsCCT γ	600	601	602
MmCCT γ	600	601	602
Bin2	600	601	602
TpCCT γ	600	601	602

Figure 5 Multiple alignment of HsCCT with other CCT γ subunits

Multiple alignment using the Clustal method [26] of the CCT γ subunits from human (HsCCT γ , X74801), mouse (*Mm*CCT γ , Z31556), yeast (*Bla2*, U09480) and *Tetrahymena pyriformis* (*7p*CCT γ , Z34885). Residues conserved in all of the subunits are indicated by asterisks, and conserved amino acid changes are indicated by a dot. Spaces were introduced to maximize the alignments and are indicated by a hyphen. The solid lines above the alignments highlight the TNDGNAIL, GDGTTs and VPGGG domains which are conserved in other members of the chaperonin gene family. The bars below the alignments represent 14 or more consecutive residues which are conserved in all four CCT γ subunits.

Figure 6 Expression of *Ctctg* mRNA in mouse tissues

Total RNA (10 µg) was isolated from mouse testis, brain and kidney and subjected to Northern-blot analysis. The blot was probed with HumAP4 and washed stringently. A *Ctctg* mRNA transcript of approx. 2.0 kb was detected in mouse testis (MT), mouse brain (MB) and mouse kidney (MK) as shown. Two different exposures of the same blot [4 h (a) and 18 h (b)] are illustrated. As a control for RNA loading, β -actin was subsequently hybridized to the same filter, and 4 h and 18 h exposures are again illustrated.

the case of adult heart, where the PCR product appeared larger than expected in a number of independent experiments (Figure 7a). The reason for the larger PCR product from adult heart is unclear. No products were observed in the negative controls where water was substituted for the template, showing that the PCR products did not result from cross-contamination.

Restriction-enzyme analysis was carried out to validate the specificity of the 638 bp PCR products. The *HsCtctg* cDNA sequence contains a single *Pst*I restriction site at position 1508 bp which is situated within the amplified fragments (region 1037–1675 bp) obtained using primers Hg-1 and Hg-2. *Pst*I digestion should result in cleavage of the 638 bp fragment into two fragments of 471 bp and 167 bp. As expected, all the PCR products examined except for adult heart yielded two fragments of 471 bp and 167 bp (Figure 7b), thus confirming the identity of the amplified cDNA fragments. In the case of adult heart, a PCR fragment larger than 638 bp had been obtained, and digestion of the PCR product with *Pst*I resulted in two fragments, one of the expected 167 bp and another that appeared slightly larger than

Figure 7 (a) Amplification of *HsCtctg* from human tissues and (b) restriction-enzyme analysis of the 3' region of *HsCtctg* amplified from a range of human tissues

(a) Amplified PCR products from a number of human adult tissues (left-hand side, lanes 2–6) and foetal tissues (right-hand side, lanes 7–11) are shown. Primers Hg-1 and Hg-2 were used to amplify PCR products from cDNA libraries from adult heart (lane 2), adult kidney (lanes 3 and 4), adult retina (lane 5), and adult thyroid (lane 6), and cDNA from foetal spleen (lane 7), foetal brain (lane 8), foetal liver (lane 9), foetal heart (lane 10) and foetal kidney (lane 11). PCR products of 638 bp were observed from all tissues except adult heart, which yielded a slightly larger product. A 1 kb DNA ladder is shown in lane 1, and the water control is shown in lane 12. (b) Restriction fragments of 471 and 167 bp observed upon *Pst*I restriction analysis of the 638 bp PCR products described in (a) from adult heart (lane 2), adult kidney (lanes 3 and 4), adult retina (lane 5), adult thyroid (lane 6), foetal spleen (lane 7), foetal brain (lane 8), foetal liver (lane 9), foetal heart (lane 10) and foetal kidney (lane 11). A 1 kb DNA ladder is shown in lane 1.

471 bp (Figure 7b). Whether the larger PCR fragment from adult heart has resulted from an artefact in the cDNA library, or whether it represents a larger transcript in adult heart, remains to be clarified.

In summary we have detected *Ctctg* mRNA, either by Northern blotting or amplification using PCR in (a) human foetal tissues (kidney, brain, heart, spleen, liver and lung), (b) human adult tissues (kidney, brain, heart, thyroid, retina and cultured skin fibroblasts) and (c) mouse tissues (kidney, brain, testis and liver). These data indicate that, like *Tcp-1/Ccta* mRNA, *Ctctg* mRNA is widely expressed both in developing and adult tissues.

***Ctctg* is a highly conserved gene.**

Tcp-1/Ccta genes have been identified in a wide variety of species, including human [17], mouse [29], fruitfly [30], pea [6] and yeast [31]. To establish whether *Ctctg* is also highly conserved, we examined genomic DNAs from human, rat, mouse, sheep and frog. The genomic DNA was digested with *Eco*RI and hybridized

Figure 8 Detection of Cctg homologues in other species

A Southern blot showing the presence of the Cctg gene in human (lane 1), rat (lane 2), mouse (lane 3), sheep (lane 4) and frog (lane 5). Approx. 10 μ g of genomic DNA was digested with *Eco*RI and hybridized with HumAP4. The membrane was washed to a stringency of $0.2 \times \text{SSC}/0.1\%$ SDS at 65°C for 1 h before being subjected to autoradiography at -70°C for 10 days.

with a mixture of equally labelled TA15 (corresponding to the 5' region of HsCctg) and HumAP4 (corresponding to the 3' region of HsCctg). Hybridizing bands were observed in all DNAs examined (Figure 8). The intensity of the hybridizing bands was lowest in frog DNA, followed by sheep, mouse and rat, with the human DNA displaying the strongest signals. This is probably a consequence of the sequence disparity among the different species, as all DNAs were equally loaded. In some cases, for example in human, mouse and rat, both strongly and weakly hybridizing bands could be seen within the same lane. It is not clear whether this is due to cross-hybridization with related genes. Cctg was also found to be conserved in hamster, stone loach and pike (results not shown).

In conclusion we have described the molecular cloning, DNA sequencing and characterization of an abundantly expressed and highly conserved human gene, HsCctg. Like Tcp-1, Cctg mRNA is expressed at higher levels in mouse testis when compared with other tissues. HsCctg mRNA is expressed in all tissues so far examined. HsCCT γ shares significant sequence similarity with other members of the CCT family of chaperonins, as well sharing conserved domains with more distantly related chaperonins such as GroEL and TF55. The predicted human CCT γ amino acid sequence displays high sequence similarity with γ -subunits from other species displaying long stretches of identical and/or conserved amino acid residues.

We are grateful to, and acknowledge the assistance of, the following: Dr. Keith Willison (Institute of Cancer Research, London) for helpful advice and discussion during the course of this work. Dr. Roger Patient (Kings' College London) and Professor GianCarlo Viberti (Guy's Hospital, London) for advice and discussions in the early stages of this work. Dr. Rachel Page (School of Molecular and Medical

Biosciences, University of Wales College of Cardiff) for attempted cloning and sequencing of NHA5-7. Rachel Gallagher and Dr. Peter Kille for providing genomic DNA preparations. Quratulain Zaidi for foetal spleen RNA and Ceri Morris for the collection of rat tissues. Dr. Leslie Wong (MRC Tissue Bank, Royal Marsden Hospital, London) for the supply of foetal tissues. Dr. Sahar Almahdawi (Medical Genetics, St. Mary's Hospital, London) for donating human foetal and adult heart libraries, and Dr. Andy Roberts (Cambridge Antibody Technology, Cambridge, U.K.) for human thyroid and human retina cDNA libraries. We thank Guy Pitt for photography. We also thank the British Heart Foundation and The Royal Society for grants which supported part of this work. N.A.W. is supported by a University of Wales Studentship.

REFERENCES

- Horwich, A. L. and Willison K. R. (1993) *Philos. Trans. R. Soc. London* **339**, 313–326
- Trent, J. D., Nimmesgern E., Wall, J. S., Hartl, F.-U. and Horwich, A. L. (1991) *Nature (London)* **354**, 490–493
- Gething, M.-J. and Sambrook J. (1992) *Nature (London)* **355**, 33–45
- Craig, E. A. (1993) *Science* **260**, 1902–1903
- Hemmingsen, S. M., Woolford C., van der Vies, S., Tilly, K., Dennis, D. T., Georgopoulos, C. P., Hendrix, R. W. and Ellis, R. J. (1988) *Nature (London)* **333**, 330–344
- Ellis, R. J. (1990) *Science* **250**, 954–959
- Gupta, R. S. (1990) *Bioch. Int.* **20**, 833–841
- Frydman, J., Nimmesgern E., Erdjument-Bromage, H., Wall, J. S., Tempst, P. and Hartl, F.-U. (1992) *EMBO J.* **11**, 4767–4778
- Kubota, H., Hynes, G., Carne, A., Ashworth, A. and Willison, K. (1994) *Curr. Biol.* **4**, 89–99
- Lewis, V. A., Hynes G. M., Zheng, D., Saibil, H. and Willison, K. (1992) *Nature (London)* **358**, 249–252
- Kubota, H., Hynes, G. and Willison, K. (1995) *Gene* **154**, 231–236
- Malik, A. N., Thompson, A., Viberti, G., Patient, R. and Demaine, A. G. (1990) in *Frontiers in Diabetes Research: Lessons from Animal Diabetes III* (Shafir, E., ed.), pp. 426–431, Smith-Gordon and Co., London
- Chen, X., Sullivan, D. S. and Huffaker, T. C. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9111–9115
- Soares, H., Penque, D., Mouta, C. and Rodrigues-Pousada, C. (1994) *J. Biol. Chem.* **269**, 29299–29307
- Yaffe, M. B., Farr G. W., Miklos, D., Horwich, A. L., Sternlicht, M. L. and Sternlicht, H. (1992) *Nature (London)* **358**, 245–248
- Gao, Y., Thomas J. O., Chow, R. L., Lee, G.-W. and Cowan, N. J. (1992) *Cell* **69**, 1043–1050
- Kirchoff, C. and Willison, K. R. (1990) *Nucleic Acids Res.* **18**, 4247
- Li, W.-Z., Lin, P., Lin, P., Frydman, J., Boal, T. R., Cardillo, T. S., Richard, L. M., Toth, D., Lichtman, M. A., Hartl, F.-U., Sherman, F. and Segel, G. B. (1994) *J. Biol. Chem.* **269**, 18616–18622
- Chomczynski, P. and Sachi, N. (1987) *Anal. Biochem.* **162**, 156–159
- Gallagher, R. A., McClean, P. M. and Malik, A. N. (1994) *Biochim. Biophys. Acta* **1217**, 329–332
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual* (Nolan, C., ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Sardet, C., Franchi A. and Pouyssegur, J. (1989) *Cell* **56**, 271–280
- Huynh, T. V. and Young, R. A. (1986) in *DNA Cloning*, vol. 1: A Practical Approach (Glover D. M., ed.), IRL Press, Oxford
- Feinberg, A. P. and Vogelstein, B. (1983) *Anal. Bioch.* **132**, 61–78
- Malik, A. N., McLean, P. M., Roberts, A., Barnett, P. S., Demaine, A. G., Banga, J. P. and McGregor, A. M. (1990) *Nucleic Acids Res.* **18**, 4031–4032
- Devereux, J. (1989) *The GCG procedure Library*, version 6.0, Genetics Computer Group Inc., Madison
- Gupta, R. S. (1995) *Mol. Microbiol.* **15**, 1–11
- Dudley, K., Potter, J., Lyon, M. and Willison, K. R. (1984) *Nucleic Acids Res.* **12**, 4281–4293
- Kubota, H., Willison, K., Ashworth, A., Nozaki, M., Miyamoto, H., Yamamoto, H., Matsushiro, A. and Morita, T. (1992) *Gene* **120**, 207–215
- Ursic, D. and Ganetzky, B. (1988) *Gene* **68**, 267–274
- Ursic, D. and Culbertson, M. R. (1991) *Mol. Cell Biol.* **11**, 2629–2640