Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Structure and evolution of the spliceosomal peptidyl-prolyl *cis–trans* isomerase Cwc27

Alexander Ulrich and Markus C. Wahl*

Laboratory of Structural Biochemistry, Freie Universität Berlin, Takustrasse 6, 14195 Berlin, Germany

Correspondence e-mail: mwahl@zedat.fu-berlin.de

Received 9 September 2014
Accepted 1 October 2014

PDB references: human Cwc27$^{6-178}$, 4r3e; *Chaetomium thermophilum* Cwc27$^{1-201}$, 4r3f

Cwc27 is a spliceosomal cyclophilin-type peptidyl-prolyl *cis–trans* isomerase (PPIase). Here, the crystal structure of a relatively protease-resistant N-terminal fragment of human Cwc27 containing the PPIase domain was determined at 2.0 Å resolution. The fragment exhibits a C-terminal appendix and resides in a reduced state compared with the previous oxidized structure of a similar fragment. By combining multiple sequence alignments spanning the eukaryotic tree of life and secondary-structure prediction, Cwc27 proteins across the entire eukaryotic kingdom were identified. This analysis revealed the specific loss of a crucial active-site residue in higher eukaryotic Cwc27 proteins, suggesting that the protein evolved from a prolyl isomerase to a pure proline binder. Noting a fungus-specific insertion in the PPIase domain, the 1.3 Å resolution crystal structure of the PPIase domain of Cwc27 from *Chaetomium thermophilum* was also determined. Although structurally highly similar in the core domain, the *C. thermophilum* protein displayed a higher thermal stability than its human counterpart, presumably owing to the combined effect of several amino-acid exchanges that reduce the number of long side chains with strained conformations and create new intramolecular interactions, in particular increased hydrogen-bond networks.

1. Introduction

In peptides, rotation around the planar peptide bond is energetically restricted owing to its partial double-bond character, allowing the peptide backbone to adopt either a *trans* (backbone torsion angle $\omega = 180^\circ$) or a *cis* ($\omega = 0^\circ$) conformation. Under physiological conditions, peptides virtually exclusively adopt *trans* conformations owing to steric hindrance of the *cis* form. The main exceptions are Xaa–Pro peptide bonds (where Xaa is any amino acid). Proline, with its unique backbone-linked five-membered ring, increases the steric conflicts of the *trans* isomer, which lowers the difference in free energy between the *cis* and *trans* isomers and thus raises the *cis* content to 5.7–6.5% (Stewart *et al.*, 1990; MacArthur & Thornton, 1991; Pal & Chakrabarti, 1999).

In the cell, Xaa–Pro peptide-bond isomerization is catalyzed by peptidyl-prolyl *cis–trans* isomerases (PPIases; EC 5.2.1.8; Fischer *et al.*, 1984) that act by stabilizing the *syn* transition state. Currently, the exact catalytic mechanism of PPIases is unclear, although several alternative mechanisms have been proposed (Fanghänel & Fischer, 2004). PPIases usually work as molecular chaperones (Gething & Sambrook, 1992; Göthel & Marahiel, 1999) by catalyzing a rate-limiting folding step that involves a Xaa–Pro *cis–trans* isomerization

(foldase; Lang *et al.*, 1987; Lin *et al.*, 1988; Davis *et al.*, 1989; Kiefhaber *et al.*, 1990; Fransson *et al.*, 1992; Tan *et al.*, 1997; Yang *et al.*, 1997) or by stabilizing a folding intermediate and thereby protecting it from aggregation (holdase; Freskgård *et al.*, 1992; Lilie *et al.*, 1993; Baker *et al.*, 1994; Bose *et al.*, 1996; Freeman *et al.*, 1996; Baneyx, 2004; Ou *et al.*, 2008). Moreover, PPIases can elicit conformational switches in already folded proteins (Yaffe *et al.*, 1997; Zhou *et al.*, 1999; Brazin *et al.*, 2002; Andreotti, 2003; Sarkar *et al.*, 2007). In contrast to the classical view of PPIases as rather promiscuous enzymes, there is evidence for sequence-specific proline isomerases, such as the phosphorylation-dependent Pin1, which binds to pSer/pThr–Pro motifs (Ranganathan *et al.*, 1997; Yaffe *et al.*, 1997; Zhou *et al.*, 1999). Besides the pure enzyme–client relationship, PPIases are capable of mediating long-term interactions *via* their catalytic centre (Dorfman *et al.*, 1997). In other cases, PPIases participate in protein–protein interactions *via* their PPIase domain that do not involve the catalytic centre (Reidt *et al.*, 2003; Xu *et al.*, 2006; Stegmann *et al.*, 2010; Wang *et al.*, 2010) or *via* a non-PPIase domain (Wang, Han *et al.*, 2008). Some PPIases serve as scaffolding proteins in supramolecular complexes (Goel *et al.*, 2001) and are essential for the assembly of large multimeric protein complexes such as photosystem II (Fu *et al.*, 2007). Using these functions, PPIases are involved in many cellular processes including regulation of mitosis (Ping Lu *et al.*, 1996), gene-expression steps (Nelson *et al.*, 2006), including transcription (Morris *et al.*, 1999; Gullerova *et al.*, 2007) and pre-mRNA splicing (Teigelkamp *et al.*, 1998; Horowitz *et al.*, 2002), post-transcriptional gene regulation (Smith *et al.*, 2009; Iki *et al.*, 2012) and cell signalling (Kimmins & MacRae, 2000; Brazin *et al.*, 2002; Yurchenko *et al.*, 2002; Lummis *et al.*, 2005; Sarkar *et al.*, 2007; Schlegel *et al.*, 2009). Furthermore, they play an important role in muscle differentiation (Hong *et al.*, 2002), stress response (Hong *et al.*, 2002; Kumari *et al.*, 2012), mitochondrial apoptosis (Lin & Lechleiter, 2002; Leung *et al.*, 2008) and the immune response (Anderson *et al.*, 1993), and also in phage (Eckert *et al.*, 2005) and viral infections as in the case of HIV (Dorfman *et al.*, 1997; Zhao *et al.*, 1997) and *Hepatitis C virus* (HCV; Watashi *et al.*, 2005; Chatterji *et al.*, 2009).

Cyclophilins (Wang & Heitman, 2005) were the first described PPIase family; they were initially isolated from bovine thymocytes and recognized by their ability to bind the immunosuppressive drug cyclosporine A (CsA; Handschumacher *et al.*, 1984). Later, their PPIase activity was discovered (Fischer *et al.*, 1989; Takahashi *et al.*, 1989). Cyclophilins, which were found to be ubiquitously present in all cellular compartments, including mitochondria, endoplasmic reticulum, Golgi, nucleus and cytoplasm (Danielson *et al.*, 1988; Haendler *et al.*, 1989; Schneuwly *et al.*, 1989; Shieh *et al.*, 1989; Liu & Walsh, 1990; Hayano *et al.*, 1991; Matouschek *et al.*, 1995; Rassow *et al.*, 1995; Dartigalongue & Raina, 1998; Wang & Heitman, 2005; Gullerova *et al.*, 2006), are highly conserved in evolution and have been identified in bacteria, archaea and eukaryota (including fungi, plants, insects and mammals) and share a high degree of sequence similarity (Schönbrunner *et al.*, 1991; Galat, 1999; Pemberton, 2006).

In humans, there are 17 cyclophilins (Davis *et al.*, 2010): PPIA, PPIB, PPIC, PPID, PPIE, PPIF, PPIG, PPIH, PPIL1, PPIL2, PPIL3, PPIL4, PPIL6, PPWD1, RANBP2, *hs*Cwc27 and NKTR. Human cyclophilins can be subdivided into single-domain and multi-domain cyclophilins. Single-domain cyclophilins contain just a PPIase domain, whereas multi-domain cyclophilins contain additional domains/regions such as RNA-recognition motifs (RRMs; in PPIE and PPIL4), arginine–serine-rich (RS) domains (in PPIG and NKTR), U-box motifs (in PPIL2), WD40 domains (in PPWD1) and tetratricopeptide-repeat (TPR) motifs (in PPID and RANBP2) (Davis *et al.*, 2010). High-resolution X-ray or NMR structures of the PPIase domains of 13 human cyclophilins have been determined [PPIA, PDB entry 2cpl (Ke, 1992); PPIB, PDB entry 1cyn (Mikol *et al.*, 1994); PPIF, PDB entry 2bit (Schlatter *et al.*, 2005); PPIH, PDB entry 1qoi (Reidt *et al.*, 2000); PPIL1, PDB entry 1xwn (Xu *et al.*, 2006); PPIL3, PDB entry 2ok3 (Huang *et al.*, 2005); PPIC, PDB entry 2esl; PPIE, PDB entry 2r99; PPIG, PDB entry 2gw2; PPIL2, PDB entry 1zkc; NKTR, PDB entry 2he9; Cwc27, PDB entry 2hq6 (Davis *et al.*, 2010); PPWD1, 2a2n (Davis *et al.*, 2008)]. All cyclophilins share a common architecture consisting of an eight-stranded antiparallel β-sheet and two α-helices that pack against the sheet, with an overall root-mean-square deviation (r.m.s.d.) of below 2 Å. Although conserved in fold, three of 15 tested cyclophilins showed no PPIase activity and did not bind CsA (Davis *et al.*, 2010). The active site is highly conserved and includes the catalytic Arg55 (in PPIA as a reference sequence) and additional residues including Phe60, Met61, Gln63, Ala101, Asn102, Phe113, Trp121, Leu122 and His126 (Ke *et al.*, 1994; Zhao *et al.*, 1997; Howard *et al.*, 2003). The strongest correlation between PPIase activity and natural occurring variation of active-site residues was found at position 121, where deviation from Trp to His or to Glu abolishes PPIase activity (Bossard *et al.*, 1991; Davis *et al.*, 2010).

Eight of the 17 human cyclophilins reside in the nucleus and are associated with the spliceosome, a multi-megadalton, highly dynamic protein–RNA machinery responsible for pre-mRNA splicing (Wahl *et al.*, 2009), *i.e.* PPIE (Mi *et al.*, 1996; Rappsilber *et al.*, 2002; Zhou *et al.*, 2002; Chen *et al.*, 2007), PPIG (Bourquin *et al.*, 1997), PPIH (Horowitz *et al.*, 2002; Chen *et al.*, 2007), PPIL1 (Rappsilber *et al.*, 2002; Zhou *et al.*, 2002; Chen *et al.*, 2007), PPIL2 (Rappsilber *et al.*, 2002; Zhou *et al.*, 2002; Chen *et al.*, 2007), PPIL3b (Rappsilber *et al.*, 2002; Zhou *et al.*, 2002; Chen *et al.*, 2007), PPWD1 (Rappsilber *et al.*, 2002; Zhou *et al.*, 2002; Chen *et al.*, 2007) and *hs*Cwc27 (Rappsilber *et al.*, 2002; Chen *et al.*, 2007). The spliceosome has to assemble and disassemble for each splicing reaction, requiring a multitude of fast and precise structural rearrangements. Cyclophilins could serve in such tasks by making use of their manifold potential: they might act as molecular switches involving the *cis–trans* isomerase activity, as protein chaperones, in particular as foldases and holdases, as specific proline binders or as scaffolding units in spliceosomal complexes. Some PPIases associate with the spliceosome in a way that leaves their PPIase active site unobstructed, as in the case of PPIH, which associates with the U4/U6 snRNP protein

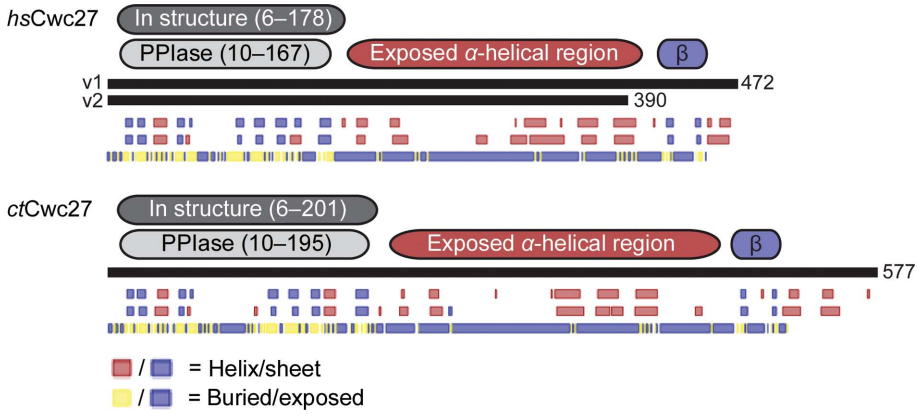

Figure 1
Domain organization of Cwc27. Secondary-structure prediction and domain organization of *H. sapiens* and *C. thermophilum* Cwc27. From top to bottom: fraction present in the X-ray structure, domain organization (β, β-strands 9 and 10), isoforms of full-length proteins (black bars; v1, splice variant 1; v2, splice variant 2), secondary-structure prediction with *REPROFSec* (blue, β-strand; red, helix), secondary-structure prediction with *PROFsec* (blue, β-strand; red, helix) and solvent-accessibility prediction with *PROFAcc* (blue, solvent-exposed; yellow, buried) from *PredictProtein* (Rost *et al.*, 2004).

Prp4 through a region that is located opposite to its active site (Reidt *et al.*, 2003). Functionally, cyclophilins might play a role in alternative splicing in higher eukaryotes (Pemberton, 2006; Wahl *et al.*, 2009), since several members are not present in yeast, or might be involved in coupling splicing to the transcription machinery (Mesa *et al.*, 2008).

The human peptidyl-prolyl *cis–trans* isomerase Cwc27 homologue (*hs*Cwc27), also called antigen NY-CO-10 (Scanlan *et al.*, 1998) or serological defined colon cancer antigen 10 (SDCCAG10), belongs to the multi-domain cyclophilins. In addition to the N-terminal PPIase domain, it contains an elongated, solvent-exposed C-terminus of unknown function. Two isoforms of *hs*Cwc27 exist: both contain an identical N-terminal cyclophilin-type PPIase domain (residues 11–166) but differ in the length of the C-terminal region (residues 167–472 in Q6UX04-1 and residues 167–390 in Q6UX04-2; Fig. 1). *hs*Cwc27 is recruited to the spliceosomal B^{act} complex (Agafonov *et al.*, 2011), a stage where the spliceosome is not yet fully catalytically active (Wahl *et al.*, 2009). Using an NMR-based assay, Davis and coworkers found *hs*Cwc27 to lack *cis–trans* isomerase activity (Davis *et al.*, 2010). The lack of activity was associated with a change at position 121 from tryptophan in the canonical PPIA sequence to glutamate in *hs*Cwc27. Nevertheless, *hs*Cwc27 was able to bind to a proline-containing peptide. In the same study, a crystal structure of the PPIase domain of *hs*Cwc27 (residues 8–172; *hs*Cwc27$^{8–172}$) was found to contain a disulfide bond between Cys44 and Cys164.

Here, we present the crystal structure of an N-terminal fragment of *hs*Cwc27 (residues 6–178; *hs*Cwc27$^{6–178}$) obtained by limited proteolysis, which includes the PPIase domain in its reduced state. Furthermore, we determined the structure of the PPIase domain of a lower eukaryotic orthologue of *hs*Cwc27, *Chaetomium thermophilum* Cwc27 (*ct*Cwc27), the first reported PPIase structure from this species. A comparison of the two structures suggested a molecular basis for the differences in the thermostability of the two proteins. Sequence analysis of Cwc27 orthologues throughout the eukaryotic spectrum suggested that Cwc27 has evolved from a *bona fide* PPIase to a pure proline binder.

2. Materials and methods

2.1. Molecular cloning

A plasmid containing the cDNA of full-length *hs*Cwc27 isoform 2 (UniProt ID Q6UX04-02) was ordered from Biocat. A plasmid containing a synthetic open reading frame (ORF) for full-length *ct*Cwc27 was ordered from Thermo Fisher Scientific. Full-length ORFs were cloned into the pETM11 vector to direct the production of proteins with N-terminal TEV protease-cleavable His_6 tags using EMP cloning as described in Ulrich *et al.* (2012). Truncations were introduced by inverse PCR as described in Ulrich *et al.* (2012).

2.2. Protein production and purification

Proteins were produced in *Escherichia coli* Rosetta 2 (DE3) or *E. coli* BL21 (DE3) RIL cells in auto-inducing ZY medium (Studier, 2005) for 24 h at 18°C. The following steps were performed at 4°C. The cells were resuspended in solubilization buffer (50 m*M* sodium phosphate pH 8.0, 500 m*M* NaCl, 30 m*M* imidazole, 5 m*M* β-mercaptoethanol) and lyzed using a Sonopuls Ultrasonic Homogenizer HD 3100 (Bandelin). The proteins were bound to Ni–NTA resin (GE Healthcare), washed, and eluted with elution buffer (250 m*M* imidazole pH 8.0, 150 m*M* NaCl, 5 m*M* β-mercaptoethanol). The tags were cleaved with 1:50 TEV protease during overnight dialysis against 10 m*M* sodium phosphate pH 8.0, 150 m*M* NaCl, 30 m*M* imidazole, 5 m*M* β-mercaptoethanol and cleaved samples were again passed over Ni–NTA resin. The flow-through was collected, concentrated and subjected to size-exclusion chromatography in SEC buffer (10 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl, 0.1 m*M* EDTA, 1 m*M* DTT). The peak fractions were pooled, concentrated and flash-frozen in liquid nitrogen.

2.3. Thermal stability assay

Samples contained protein at a concentration of 0.05, 0.075 or 0.1 mg ml^{-1} in 20 m*M* sodium phosphate pH 8.0, 300 m*M* NaCl and 5× SYPRO Orange (Life Technologies). Thermal stability assays were performed in parallel using an RT-PCR machine (Mx3005P QPCR system, Agilent Technologies) by monitoring the fluorescence emission at 610 nm while ramping through a temperature gradient from 25 to 95°C in 71 steps. Melting-temperature (T_m) values were calculated from the raw data by curve fitting with the Boltzmann equation using *Origin* (OriginLab). The average T_m was taken as the arithmetic mean of 21 measurements (the unfolding of each protein monitored seven times at each of the three concentrations) with stated errors representing the standard deviations.

2.4. Phylogenetic analysis

A protein *BLAST* search (http://blast.ncbi.nlm.nih.gov) with *hs*Cwc27 as the query sequence was conducted against the high-quality sequenced genomes of 30 model organisms (Fig. 2) which cover the entire eukaryotic tree of life (Ciccarelli *et al.*, 2006; Fritz-Laylin *et al.*, 2010). *BLAST* hits were considered to be significant if the *BLAST* scores were higher than 50, the *E* values were lower than 1×10^{-10} and the query coverage was higher than 40%. These criteria were selected to avoid false positives, which might arise based on the high sequence identity of the PPIase domain.

2.5. Secondary-structure prediction

Secondary-structure predictions were performed with *PredictProtein* (http://www.predictprotein.org; Rost *et al.*, 2004). Helices, β-strands and loops were predicted by *REPROFSec* and *PROFsec*, and solvent accessibility was predicted by *PROFAcc*.

2.6. Crystallization

Full-length *hs*Cwc27 splice variant 2 (residues 1–390) at a protein concentration of 55 mg ml^{-1} was used in sitting-drop vapour-diffusion experiments with commercial screens at 18°C and with 0.1 µl protein solution plus 0.1 µl reservoir solution per drop. Three-dimensional needle crystals were obtained after one month in 0.1 *M* magnesium formate, 15%(*w*/*v*) PEG 3350. This solution had a pH of 6.8 and did not contain a buffer component.

The PPIase domain of *ct*Cwc27 (residues 1–201) at a protein concentration of 43.2 mg ml^{-1} was used in sitting-drop vapour-diffusion experiments with commercial screens at 18°C and 0.1 µl protein solution plus 0.1 µl reservoir solution per drop. Initial crystals were obtained after 2 d in a third of all conditions. The best hits were obtained in 0.1 *M* sodium acetate pH 4.6, 8%(*w*/*v*) PEG 4000. Larger crystals were produced in 24-well sitting-drop plates under the same conditions with 1 µl protein solution plus 1 µl reservoir solution per drop.

2.7. Data collection and structure determination

Prior to data collection, the crystals were mounted in a loop, soaked in cryoprotecting solution [0.1 *M* magnesium formate, 15%(*w*/*v*) PEG 3350, 25%(*v*/*v*) glycerol for *hs*Cwc27; 0.1 *M* sodium acetate pH 4.6, 8%(*w*/*v*) PEG 4000, 30%(*v*/*v*) ethylene glycol for *ct*Cwc27] and flash-cooled in liquid nitrogen. Diffraction data were collected on beamline BL14.2 of the BESSY II storage ring, Berlin, Germany at 100 K. For *hs*Cwc27, a crystal-to-detector distance of 240 mm was used to collect 130 images of 1° per image with an exposure time of 5 s at a wavelength of 0.91841 Å. For *ct*Cwc27, a crystal-to-detector distance of 130 mm was used to collect 300 images of 1° per image with an exposure time of 5 s at a wavelength of 0.91841 Å. Both structures were solved by molecular replacement with *Phaser* (McCoy, 2007) using the structure coordinates of *hs*Cwc27$^{8–172}$ (PDB entry 2hq6; Davis *et al.*, 2010) as a search model. The structures were corrected and completed by manual model building with *Coot* (Emsley & Cowtan, 2004). Refinement of the *hs*Cwc27 structure was performed with *REFMAC*5 (Murshudov *et al.*, 2011). *B* factors were refined isotropically. The *ct*Cwc27 structure was refined with *phenix.refine* (Afonine *et al.*, 2012). *B* factors were refined anisotropically with the exception of H atoms.

2.8. Accession numbers

Structure coordinates and diffraction data for *hs*Cwc27$^{6–178}$ and *ct*Cwc27$^{6–201}$ have been deposited in the Protein Data Bank (http://www.pdb.org) as PDB entries 4r3e and 4r3f, respectively.

3. Results and discussion

3.1. Structure of a relatively protease-resistant fragment of *hs*Cwc27

As of 20 May 2014, close to 300 structures of PPIase domains had been deposited in the Protein Data Bank (http://www.pdb.org), but although many PPIases contain additional domains next to their PPIase domain, no multi-domain PPIase structure has been determined to date. In *hs*Cwc27, an elongated C-terminus that is predicted to be solvent-exposed and partially α-helical is appended to the N-terminal cyclophilin-type PPIase domain (Fig. 1). We cloned and purified full-length *hs*Cwc27 (isoform 2) but noticed signs of protein degradation during purification and upon incubation at room temperature (Fig. 3*a*). Crystals only appeared about one month after crystallization setup and analysis of the crystals by SDS–PAGE showed a single band of ∼17 kDa (full-length protein 53.8 kDa; Fig. 3*b*). Thus, a stable *hs*Cwc27 fragment formed in the crystallization drops, presumably produced by residual proteases in the preparation, and eventually crystallized. Diffraction analysis revealed that the crystals belonged to the same space group and exhibited similar unit-cell parameters as a published PPIase-domain structure of *hs*Cwc27 (PDB entry 2hq6; Davis *et al.*, 2010; Table 1). We solved the 2.0 Å resolution structure of the *hs*Cwc27 fragment contained in our crystals by molecular replacement with the coordinates

	#	Organism	TaxID	UniProt ID[a]	BLAST score[b]	Query coverage[c]	Seq. identity[d]	E-value[e]	Reviewed[f]
Eukaryotes / Unikonts / Animals	1	Homo sapiens	9606	Q6UX04	967	100%	100%	0.E+00	Yes
	2	Ornithorhynchus anatinus	9258	XP_007663633.1	199	23%	83%	5.E-60	No
	3	Gallus gallus	9031	E1C0I5	667	100%	77%	0.E+00	No
	4	Takifugu rubripes	31033	XP_003965379.1	538	97%	67%	0.E+00	No
	5	Danio rerio	7955	Q7ZW86	600	100%	71%	0.E+00	Yes
	6	Branchiostoma floridae	7739	C3ZRN2	466	99%	56%	4.E-160	No
	7	Ciona intestinalis	7719	XP_002125951.2	419	100%	50%	1.E-141	No
	8	Strongylocentrotus purpuratus	7668	W4XQF1	417	100%	53%	1.E-140	No
	9	Drosophila melanogaster	7227	Q9VTN7	380	97%	46%	8.E-126	No
	10	Caenorhabditis elegans	6239	Q9XXI7	313	97%	42%	1.E-100	No
	11	Nematostella vectensis	45351	A7RQN6	449	99%	54%	9.E-154	No
	12	Trichoplax adhaerens	10228	B3S110	421	96%	52%	2.E-143	No
	13	*Amphimedon queenslandica*	*400682*	*XP_003390321.1*	*236*	*53%*	*56%*	*2.E-74*	*No*
Unikonts / C	14	Monosiga brevicollis	81824	A9UTI0	281	44%	63%	5.E-92	No
Unikonts / Fungi	15	Pseudozyma antarctica	84753	M9M3H9	238	52%	49%	4.E-73	No
	16	Schizosaccharomyces pombe	4896	O42941	255	44%	56%	3.E-79	No
	17	Saccharomyces cerevisiae	4932	Q02770	75.9	55%	28%	2.E-14	Yes
	18	Neurospora crassa	5141	Q7SBX8	236	90%	36%	1.E-70	Yes
	19	Chaetomium thermophilum	209285	G0RY38	227	90%	35%	2.E-67	No
Unikonts / Am	20	Dictyostelium discoideum	44689	Q55G43	230	44%	51%	5.E-69	No
	21	Entamoeba histolytica	5759	C4M525	110	26%	42%	3.E-28	No
	22	Physarum polycephalum	5791	–	23.5	5%	41%	1.5	–
Unikonts / Pl	23	Arabidopsis thaliana	3702	Q6Q152	308	95%	41%	3.E-98	No
	24	Physcomitrella patens	3218	A9T5X8	286	95%	41%	7.E-90	No
SAR / Pl	25	Chlamydomonas reinhardtii	3055	A8J9P9	285	44%	62%	3.E-88	No
	26	Paramecium tetraurelia	5888	A0BH25	242	50%	46%	3.E-73	No
SAR / A	*27*	*Thalassiosira pseudonana*	*35128*	*B8BQW7*	*137*	*31%*	*47%*	*2.E-38*	*No*
SAR / St	28	Phytophthora ramorum	164328	–	22.7	11%	28%	4.E+00	–
Excavates / Di	29	*Naegleria gruberi*	*5762*	*D2VH68*	*197*	*43%*	*48%*	*1.E-59*	*No*
	30	Leishmania major	5664	E9AFV2	111	23%	44%	2.E-27	No
Excavates / F	31	Giardia lamblia	5741	V6TZ19	98.6	33%	38%	2.E-23	No

F = Fornicata
Di = Discicristates
St = Stramenopiles
A = Alveolates
Pl = Plantae
Am = Amoebozoa
C = Choanoflagellates
Exc = Excavates
SAR = Stramenopiles, Alveolates, Rhizaria

[a] If the UniProt ID is not available, the NCBI reference sequence ID is indicated.
[b] BLAST scores are calculated with the BLOSUM62 scoring matrix.
[c] Percentage of query sequence covered by alignment with the database sequence.
[d] Percentage of identical residues of query sequence and BLAST hit within covered region.
[e] E-value: number of hits expected by chance when searching a database of particular size.
[f] Protein sequence entry reviewed by UniProtKB/Swiss-Prot.

Figure 2
Cwc27 tree-of-life analysis: results of *BLAST* searches with *hs*Cwc27 against the indicated species.

of the published *hs*Cwc27 PPIase-domain structure and refined it to an R_{work} of 17.2% and an R_{free} of 21.7% (Table 1). Residues 6–178 of *hs*Cwc27 (*hs*Cwc27$^{6–178}$) could be traced in the electron density, closely corresponding to the region contained in the previously published structure (residues 8–172). The common regions of both structures are very similar, with an r.m.s.d. of 0.5 Å for 167 common C^{α} atoms; Supplementary Fig. S1[1]). Our structure contains one Ramachandran outlier, Asp150, located in a highly flexible loop between α-helix 2 and β-strand 8, and two rotamer outliers, Ile6 and Gln7, representing the two N-terminal residues.

Besides the canonical cyclophilin core with eight antiparallel β-strands, two α-helices and a short α-helical turn, *hs*Cwc27$^{6–178}$ contains a short β-segment (Fig. 4*a*). Residues Arg56, Phe61, Ile62, Gln64, Ala102, Asn103, Phe114, Glu122, Leu123 and His127 in β-strands 3, 4 and 6, as well as in the extended loop connecting β-strands 6 and 7, form the proline-binding pocket (Fig. 4*b*). This pocket is occupied by a glycerol molecule, the hydrophobic hydrocarbon scaffold of which faces the hydrophobic part of the pocket, while its three hydroxyl groups are involved in an extensive hydrogen-bond network involving Arg56, Gln64, Asn103, Glu122, His127 and Glu153 (Fig. 4*b* and Supplementary Fig. S2). In *hs*Cwc27, a glutamate at position 122 replaces the canonical tryptophan, tyrosine or histidine of active PPIase (Davis *et al.*, 2010) but is not involved in binding the glycerol.

Although globally highly similar, the *hs*Cwc27$^{6–178}$ structure and the previously determined *hs*Cwc27$^{8–172}$ structure exhibit two subtle differences. Firstly, our structure includes a C-terminal stretch (residues 173–178) that is not contained in *hs*Cwc27$^{8–172}$ and which does not belong to the canonical

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: DW5118).

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data set	*hs*Cwc27$^{6–178}$	*ct*Cwc27$^{6–201}$
Data collection		
Wavelength (Å)	0.918	0.918
Temperature (K)	100	100
Space group	$P6_5$	$P2_12_12_1$
Unit-cell parameters (Å, °)	$a = b = 85.9$, $c = 55.0$, $\alpha = \beta = 90.0$, $\gamma = 120.0$	$a = 58.6$, $b = 58.7$, $c = 59.3$, $\alpha = \beta = \gamma = 90.0$
Matthews coefficient (Å^3 Da^{-1})	3.01	2.18
Solvent content (%)	58.8	43.1
Resolution range (Å)	44.27–2.00 (2.05–2.00)	33.98–1.30 (1.33–1.30)
Reflections		
Total	120743 (5456)	297230 (17551)
Unique	30399 (2114)	49760 (3368)
Completeness (%)	99.3 (93.5)	97.7 (90.7)
Multiplicity	4.0 (2.6)	6.0 (5.2)
R_{meas}† (%)	14.3 (84.6)	3.7 (25.9)
$\langle I/\sigma(I)\rangle$	9.7 (1.6)	29.7 (6.6)
$CC_{1/2}$	99.4 (48.9)	99.9 (61.8)
Wilson *B* factor (Å^2)	30.5	19.1
Refinement		
Unique reflections	14902	49755
R_{work}‡ (%)	17.2 (24.2)	12.1 (13.6)
R_{free}§ (%)	21.7 (28.3)	14.3 (18.9)
No. of atoms		
Total	1588	3586
Protein	1371	3171
Water	199	296
Glycerol	18	—
Ethylene glycol	—	90
Polyethylene glycol	—	22
Acetate	—	7
B factors (Å^2)		
Total	26.4	20.1
Protein	24.7	17.7
Water	36.6	33.7
Glycerol	43.0	—
Ethylene glycol	—	48.1
Polyethylene glycol	—	55.6
Acetate	—	59.9
R.m.s.d.¶		
Bond lengths (Å)	0.006	0.008
Bond angles (°)	1.084	1.246
Ramachandran plot		
Favoured (%)	98.3	98.0
Allowed (%)	1.2	2.0
Outliers (%)	0.6	0.0
Rotamer outliers (%)	1.4	0.5

† $R_{meas} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean intensity of symmetry-equivalent reflections and $N(hkl)$ is the multiplicity. ‡ $R_{work} = \sum_{hkl}||F_{obs}| - |F_{calc}||/\sum_{hkl}|F_{obs}|$ (working set, no σ cutoff applied). § R_{free} is the same as R_{work} but calculated on 5% of the data that were excluded from refinement. ¶ Root-mean-square deviation from target geometry.

PPIase domain (Fig. 4*a*). It lies on top of helix α1 and forms interactions with the loop connecting β-strands 4 and 5. More precisely, Asp173 forms a salt bridge with Arg38, Ile174 is engaged in hydrophobic interactions with Ile41 and Phe171, the backbone O and N atoms of Ile175 make hydrogen bonds to Gln42 both directly and *via* a water molecule, and Arg177 forms hydrogen bonds directly to Glu46 and *via* water to Glu76 and Ser77 (Fig. 4*c*). These extensive interactions explain the relative protease resistance of the C-terminal appendix and suggest that it structurally stabilizes the PPIase core. Secondly, our structure is in its fully reduced state and does not contain a disulfide bond between Cys44 and Cys164 as present in *hs*Cwc27$^{8–172}$ (Fig. 4*d*). *hs*Cwc27 is a nuclear protein and is therefore exposed to a reductive environment which does not support the formation of disulfide bonds. Furthermore, the very similar structures of both molecules show that the disulfide bond is not necessary for fold stability. In addition, the cysteines involved in disulfide-bond formation in *hs*Cwc27$^{8–172}$, Cys44 and Cys164, are not universally conserved in Cwc27 proteins (Fig. 5 and Supplementary Fig. S3). Finally, although Cys44 and Cys164 are conserved in several human cyclophilins (in 12 of 17 in the case of Cys44 and in ten of 17 in the case of Cys164), none of the 13 known cyclophilin structures contains a disulfide bond between these residues. Therefore, the physiological relevance of the disulfide bond observed in the previous *hs*Cwc27$^{8–172}$ structure is questionable.

As the regions of *hs*Cwc27 C-terminal of residue 178 are highly susceptible to degradation (Fig. 3*a*), they are most either poorly structured or attached *via* a flexible linker to the PPIase domain. The former possibility is supported by the predicted solvent-exposed nature of the C-terminal part with isolated secondary-structure units (Fig. 1). As in many other intrinsically unstructured protein regions, the C-terminal portion of *hs*Cwc27 may be involved in protein–protein interactions, for example with the spliceosomal C complex protein FRA10AC1 (Hegele *et al.*, 2012).

3.2. Identification of Cwc27 proteins across the eukaryotic kingdom

While the cyclophilin protein family comprises 17 members in humans, only eight cyclophilins are found in yeast (Pemberton, 2006), among which is a Cwc27 protein (Fabrizio *et al.*, 2009). Notably, eight of the human cyclophilins are associated with the spliceosome (Wahl *et al.*, 2009), while the *Saccharomyces cerevisiae* Cwc27 protein (*sc*Cwc27) is the single spliceosomal cyclophilin in this species (Fabrizio *et al.*, 2009). As alternative splicing is pervasive in humans (Wang, Sandberg *et al.*, 2008) but has a significantly lower extent in yeast (Kempken, 2013), the above situation suggests that human spliceosomal cyclophilins may be involved in alternative splicing decisions and that the functions of spliceosomal cyclophilins may have changed during the course of evolution. To further investigate the latter possibility, we performed a comprehensive phylogenetic search for Cwc27 proteins, combining multiple sequence alignments, tree-of-life analysis and secondary-structure prediction. A major difficulty in specifically identifying Cwc27 proteins is the high sequence conservation of all cyclophilin-type PPIase domains (>60% in all binary combinations) and the relatively low conservation of the regions C-terminal of the PPIase domains. In identifying true Cwc27 proteins, we raised the cutoff for the query coverage to 40% in addition to the criteria *E*-value $< 1 \times 10^{-10}$ and *BLAST* score > 50. Furthermore, we noticed that 11 of the 17 human cyclophilins can easily be distinguished because their sequences terminate 13 or fewer residues C-terminal of

their PPIase domains (PPIA, PPIB, PPIC, PPIE, PPIF, PPIH, PPIL1, PPIL3, PPIL6, PPWD1 and RBP2), in sharp contrast to the 305-residue C-terminus of *hs*Cwc27. Conversely, NKTR proteins exhibit an extremely long C-terminus (1286 residues in human). Furthermore, PPIL2 proteins can easily be distinguished because of their ∼250-residue N-terminal extension, which is not present in Cwc27 proteins, which start with the PPIase domain. To additionally distinguish true Cwc27 proteins from other cyclophilins, we performed secondary-structure predictions for all hit sequences (Supplementary Fig. S4), as the N-terminus of *hs*Cwc27 exhibits a unique signature of secondary-structure elements that clearly differs from all other human cyclophilins. Finally, some cyclophilins that show a similar position of their PPIase domain and a similar overall length to *hs*Cwc27 differ from Cwc27 proteins in the secondary-structure signatures of their C-termini (Supplementary

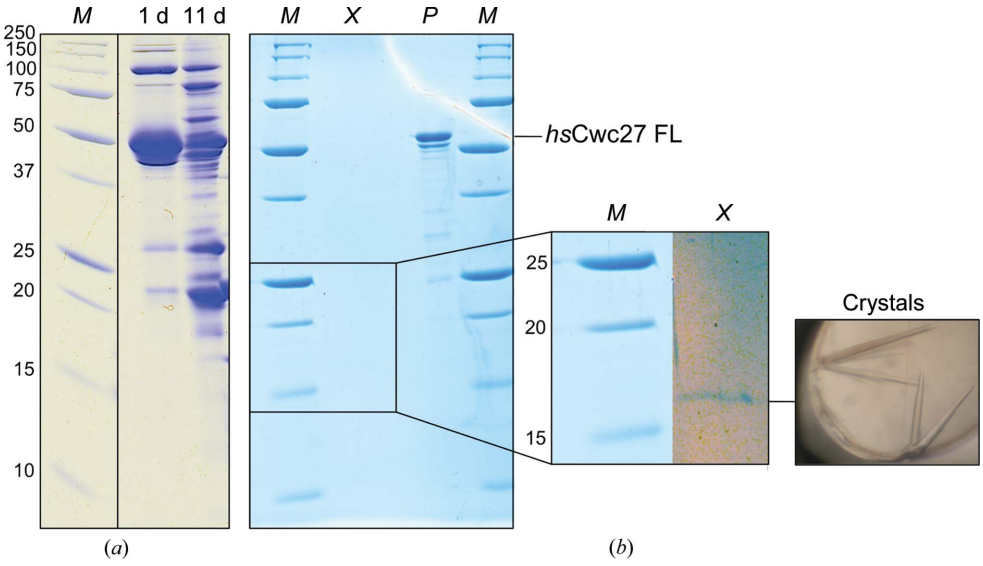

Figure 3
Crystallization of *hs*Cwc27. (*a*) SDS–PAGE of full-length *hs*Cwc27 splice variant 2 after 1 d and after 11 d of incubation at room temperature. (*b*) Left panel, SDS–PAGE of washed *hs*Cwc27 crystals (lane *X*) grown from a setup with full-length protein (lane *P*) as the input material. Middle panel, lane *X* with increased contrast. Right panel, *Hs*Cwc27 crystals obtained approximately one month after crystallization setup. Lane *M* contains molecular-mass marker (labelled in kDa).

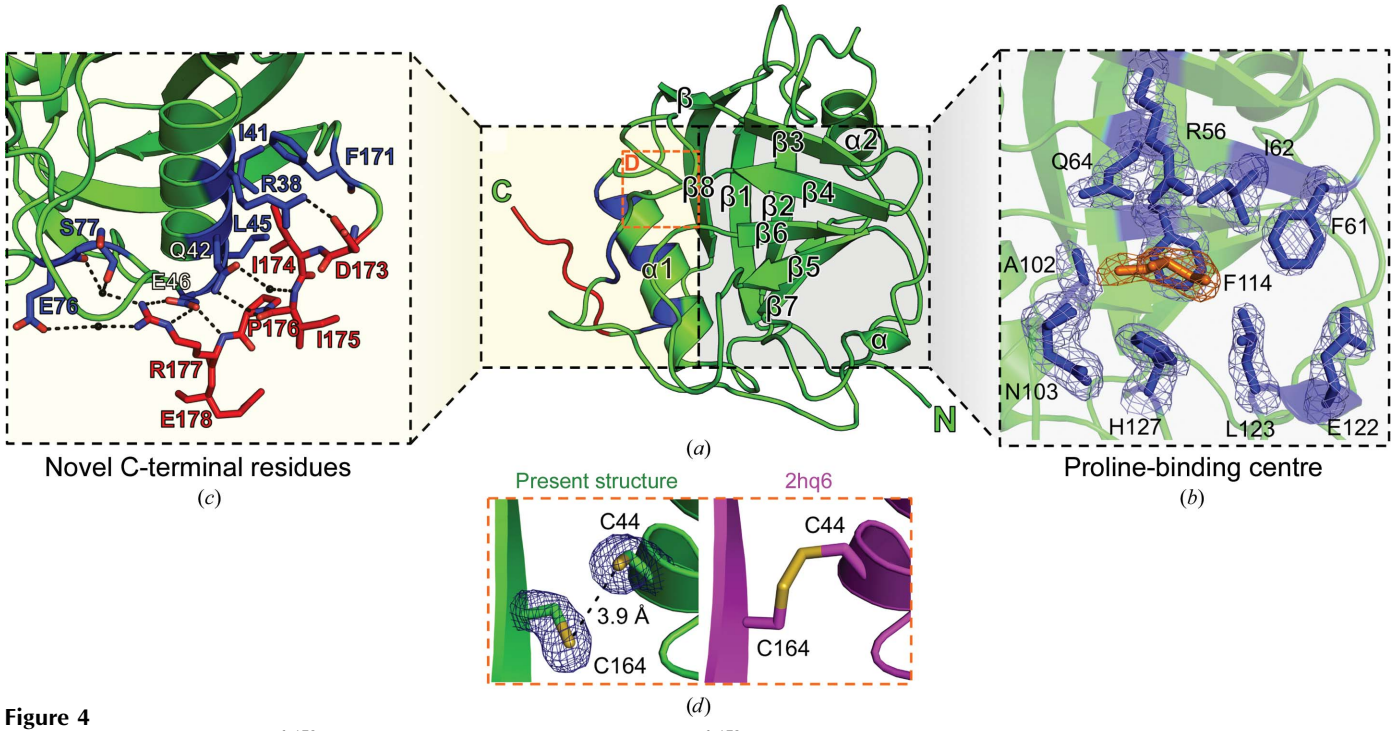

Figure 4
Crystal structure of *hs*Cwc27$^{6–178}$. (*a*) Cartoon representation of *hs*Cwc27$^{6–178}$. Secondary-structure elements are labelled (α, α-helix; β, β-strand). Residues 173–178, which are not present in the previously published structure of the *hs*Cwc27 PPIase domain (PDB entry 2hq6; Davis *et al.*, 2010), are coloured red. Residues of the PPIase core are shown in blue. (*b*) The proline-binding pocket. Residues homologous to the active-centre residues of *hs*PPIA are presented as blue sticks; a glycerol molecule located in the pocket is presented as orange sticks. The electron-density map is an $F_o - F_c$ OMIT map contoured at $\pm 3\sigma$ calculated from a refined model without active-site residues and glycerol. (*c*) Details of the interaction of residues 173–178 (red sticks) with residues of the PPIase core (blue sticks). Relevant water molecules are shown in black; hydrogen bonds and salt bridges are shown as dashed black lines. This view is rotated +67° about the *x* axis; +83° about the *y* axis and +164° about the *z* axis compared with that in (*a*). (*d*) Comparison of *hs*Cwc27 residues Cys44 and Cys164 in their reduced state (present structure) and in their oxidized state (PDB entry 2hq6; Davis *et al.*, 2010). The electron-density map is an $F_o - F_c$ OMIT map contoured at $\pm 3\sigma$ calculated from a refined model without Cys44 and Cys164.

Fig. S4). For example, PPIL4 proteins contain a C-terminal RRM domain with the typical $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology (Maris *et al.*, 2005) and a unique exposed/buried pattern, while PPIG proteins display a presumably weakly structured C-terminal RS-domain and PPID proteins contain three TPR motifs, resulting in an array of buried α-helices. In contrast to these proteins, Cwc27 proteins have a mainly α-helical, solvent-exposed C-terminus, with two predicted short, adjacent β-strands (β9 and β10). All *BLAST* hits which passed our quality criteria showed a Cwc27-like pattern of secondary-structure elements (Supplementary Figs. S3 and S4).

Based on the above criteria, we found that Cwc27 is conserved from animals to fungi, amoebozoa, plants and even alveolates (*Paramecium tetraurelia*; Fig. 2), which separated from the animal, fungi and plant lineages almost two billion years ago (Hedges *et al.*, 2004). One of the hits with the lowest sequence identity to *hs*Cwc27 was *sc*Cwc27 (28%; Q02770), which differed significantly more than Cwc27 representatives from other fungal species (*Pseudozyma antarctica*, 49% sequence identity; *Schizosaccharomyces pombe*, 56%; *Neurospora crassa*, 36%; *C. thermophilum*, 35%; Fig. 2). The sequence deviation manifests especially in a series of charge-inversion mutations of surface residues in the PPIase domain of *sc*Cwc27: *hs*Lys35 *versus* *sc*Glu33, *hs*Arg96 *versus* *sc*Asp87R, *hs*Arg119 *versus* *sc*Asp106, *hs*Asp121 *versus* *sc*Lys108, *hs*Lys126 *versus* *sc*Asp113, *hs*Asp136 *versus* *sc*Lys123 and *hs*Arg143 *versus* *sc*Glu130. Even *hs*Arg56, which corresponds to a key catalytic residue in enzymatically active cyclophilins, is replaced by Glu54 in *sc*Cwc27 (Supplementary Figs. S3 and S5). Some of these changes occurred in a compensatory fashion, *e.g.* the salt bridge *hs*Arg119–*hs*Asp121 in human was converted to an Asp106–Lys108 interaction in *S. cerevisiae*. The large number of sequence changes in *sc*Cwc27 compared with *hs*Cwc27 is most likely not owing to evolutionary distance alone, as Cwc27 proteins in other fungal species and in evolutionarily more distant species (plants and alveolates) show a higher sequence identity (Fig. 2). A more likely reason is a genome-wide duplication event which occurred in *S. cerevisiae* after its separation from the *Kluyveromyces* lineage (Wolfe & Shields, 1997; Seoighe & Wolfe, 1999; Ladrière *et al.*, 2000). An increased number of copies per gene allows a higher mutation rate. Subsequently, because of functional redundancy, extra copies of genes are rapidly lost and the remaining copies comprise a higher number of mutations compared with the original gene.

3.3. Cwc27 developed from an active PPIase to a pure proline binder

Previously, it was noted that a tryptophan at position 122 (*hs*Cwc27 numbering; corresponding to position 121 in the archetypical *hs*PPIA) is optimal for prolyl-isomerase activity and that tyrosine or histidine at this position strongly reduce but do not necessarily exclude this function, whereas a glutamate at this site, as found in Cwc27, abolishes PPIase activity (Davis *et al.*, 2010). Of the 17 human cyclophilins, only *hs*Cwc27 contains a glutamate at position 122. 13 of the 22 Cwc27 proteins identified in our tree-of-life analysis also contain a glutamate at the position homologous to position 122 in human and thus are also expected to be inactive (Fig. 5 and Supplementary Figs. S3). In addition, *Dictyostelium discoideum* Cwc27 contains an asparagine at position 122, which also renders it inactive (Bossard *et al.*, 1991; Liu *et al.*, 1991; Zydowsky *et al.*, 1992). However, several Cwc27 proteins contain a histidine at this position (*Bronchiostoma floridae*, *Stronglyocentrotus purpuratus* and *Saccharomyces cerevisiae*),

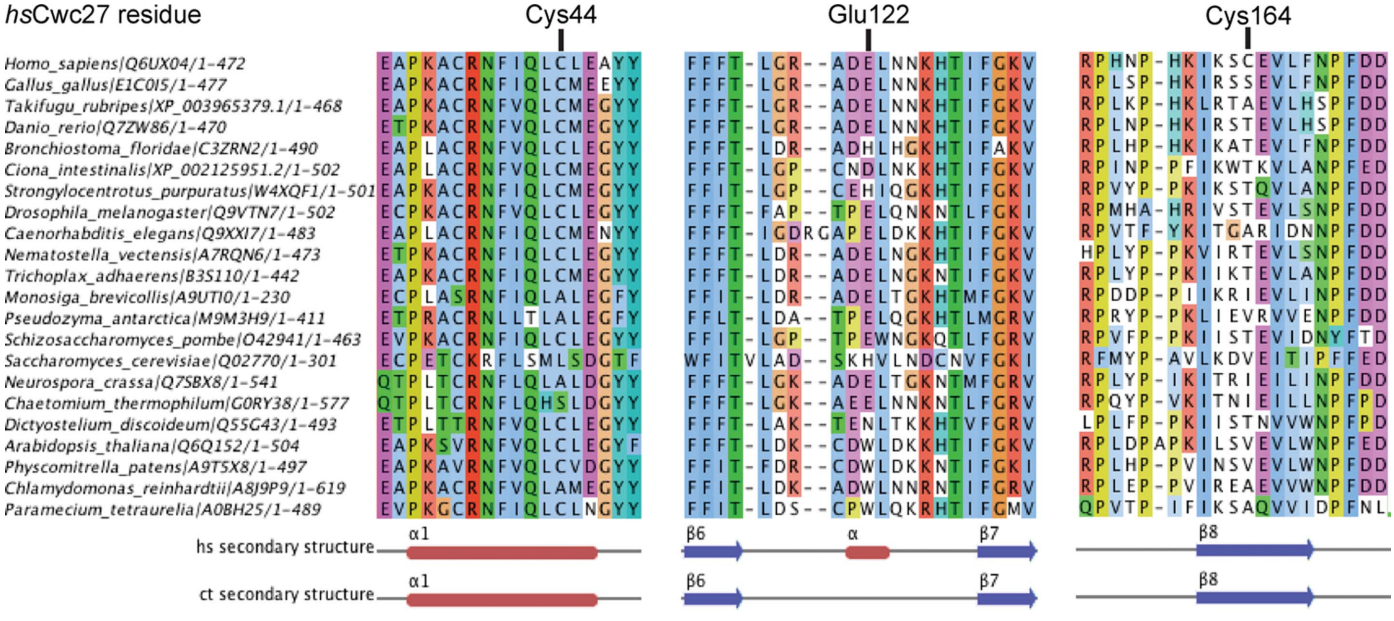

Figure 5
Conservation of *hs*Cys44, *hs*Glu122 and *hs*Cys164. Excerpts from a multiple sequence alignment of 22 Cwc27 proteins identified in our tree-of-life analysis. The alignment was built with *Jalview* (Waterhouse *et al.*, 2009) and calculated with the *MUSCLE* algorithm (Edgar, 2004). Secondary-structure annotations are derived from our structures. The full alignment is presented in Supplementary Fig. S3.

allowing moderate isomerase activity. A tryptophan, required for optimal isomerase activity, is present in the Cwc27 proteins of the four species in our analysis that are most divergent from human: *Arabidopsis thaliana*, *Physcomitrella patens*, *Chlamydomonas reinhardtii* and *Paramecium tetraurelia*. This observation indicates that either these four species readopted a PPIase activity in their Cwc27 proteins or, more likely, that Cwc27 started as an active PPIase (Trp122) but over time significantly reduced (Tyr122 or His122) or lost (Glu122) its catalytic activity. The inactive Cwc27 proteins retained the ability to bind proline-containing peptides (Davis *et al.*, 2010).

Many other enzyme families also encompass members that lack key catalytic residues and have lost their enzymatic activities. However, such pseudo-enzymes are not 'dead' proteins but are capable of fulfilling important cellular functions, in many cases by having retained certain aspects of their original roles as an enzyme, such as certain binding/interaction abilities. For example, in the case of rhomboid proteases, proteolytically inactive paralogues, the so-called iRhoms, have maintained the ancestral ability to bind to type I transmembrane proteins but have lost their ability to cleave their targets (Adrain & Freeman, 2012). Furthermore, some proteins were originally considered to be pseudo-enzymes but have apparently retained enzymatic functions under very specialized conditions. For instance, the neuronal Ca^{2+}/calmodulin-activated serine-threonine kinase CASK lacks residues indispensable for canonical phosphoryl transfer but may embody a noncanonical kinase mechanism (Mukherjee *et al.*, 2008). Higher eukaryotic Cwc27 proteins may have undergone a similar evolution. While they may exert roles in spliceosomes that are based purely on their remaining ability to bind proline-containing peptides, it remains to be seen whether these proteins, in contrast to isomerizing many different target peptides, might have evolved to catalyze the isomerization of very specific sequences, potentially involving Glu122 in the selection process. Further work, especially the identification of the *in vivo* targets of Cwc27, will be necessary to better understand the role of Cwc27 as a proline binder or a specialized enzyme.

3.4. Structural comparison with a lower eukaryotic Cwc27 PPIase domain

While establishing the evolutionary history of Cwc27 proteins, we noted that certain fungal Cwc27 representatives (from *N. crassa* and *C. thermophilum*) had acquired a specific

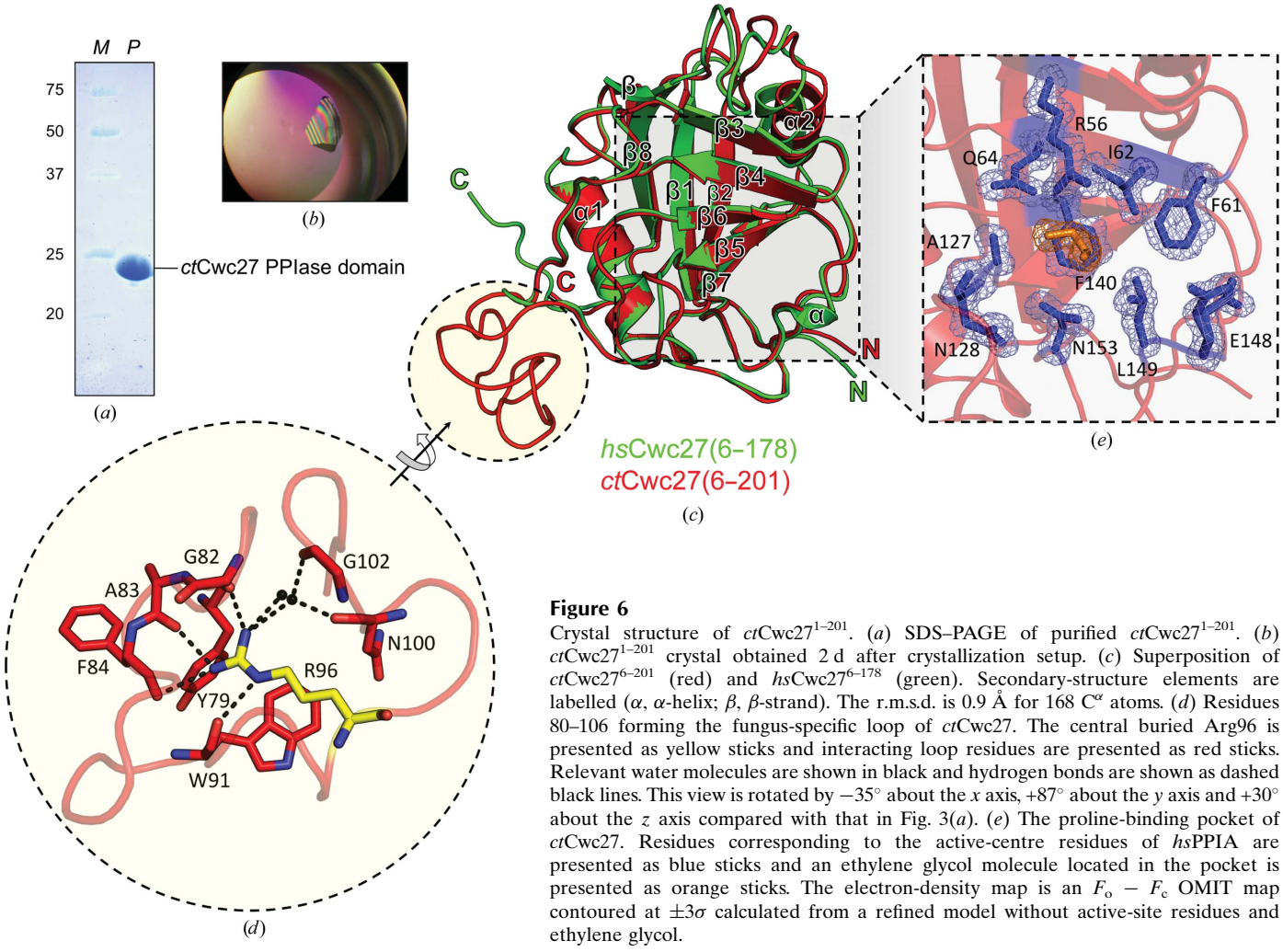

Figure 6
Crystal structure of *ct*Cwc27$^{1–201}$. (*a*) SDS–PAGE of purified *ct*Cwc27$^{1–201}$. (*b*) *ct*Cwc27$^{1–201}$ crystal obtained 2 d after crystallization setup. (*c*) Superposition of *ct*Cwc27$^{6–201}$ (red) and *hs*Cwc27$^{6–178}$ (green). Secondary-structure elements are labelled (α, α-helix; β, β-strand). The r.m.s.d. is 0.9 Å for 168 C^{α} atoms. (*d*) Residues 80–106 forming the fungus-specific loop of *ct*Cwc27. The central buried Arg96 is presented as yellow sticks and interacting loop residues are presented as red sticks. Relevant water molecules are shown in black and hydrogen bonds are shown as dashed black lines. This view is rotated by −35° about the *x* axis, +87° about the *y* axis and +30° about the *z* axis compared with that in Fig. 3(*a*). (*e*) The proline-binding pocket of *ct*Cwc27. Residues corresponding to the active-centre residues of *hs*PPIA are presented as blue sticks and an ethylene glycol molecule located in the pocket is presented as orange sticks. The electron-density map is an $F_o - F_c$ OMIT map contoured at $\pm 3\sigma$ calculated from a refined model without active-site residues and ethylene glycol.

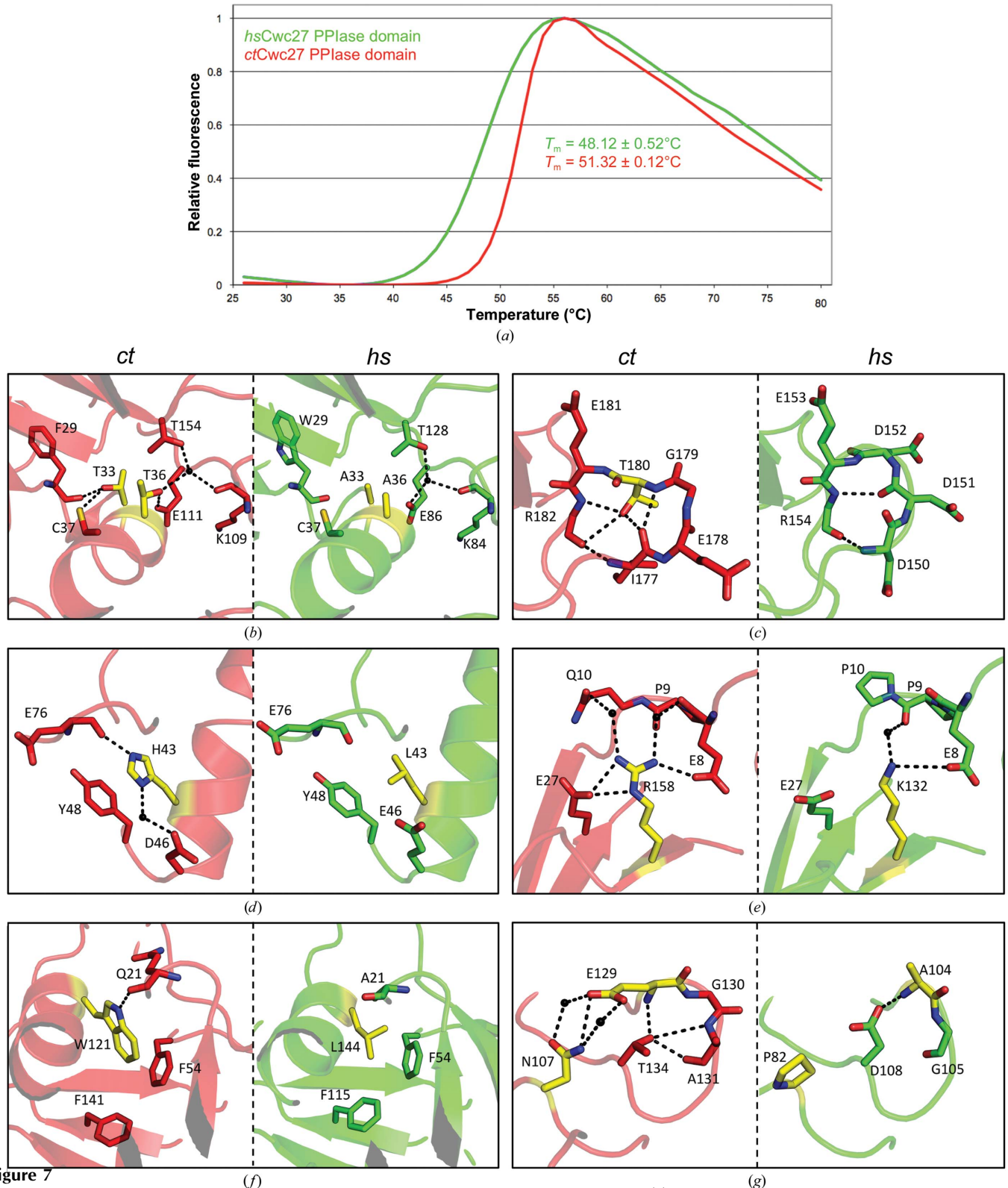

Figure 7
Molecular basis for the increased thermal stability of the *ct*Cwc27 and *hs*Cwc27 PPIase domains. (*a*) Representative thermal unfolding graphs of PPIase domains of *hs*Cwc27$^{6–178}$ (green) and *ct*Cwc27$^{1–201}$ (red). Relative fluorescence is indicated in arbitrary units. The melting temperature (T_m) is the average T_m of 21 independent measurements each; the reported errors represent the standard deviations. For more detailed results, see Supplementary Table S1. (*b*) Comparison of the intramolecular interactions of residues that are presumably relevant for the increased thermal stability of the PPIase domains of *ct*Cwc27$^{1–201}$ (red) compared with the PPIase domains of *hs*Cwc27$^{6–178}$ (green). Substituted residues are presented as yellow sticks; interacting residues are presented as red (*ct*) or green (*hs*) sticks. Relevant water molecules are shown in black and hydrogen bonds are depicted as dashed black lines. Comparisons of the environments of *ct*Thr33 and *ct*Thr36 with *hs*Ala33 and *hs*Ala36 (*b*), of *ct*Thr180 with the homologous region in *hs*Cwc27 (*c*), of *ct*His43 with *hs*Leu43 (*d*), of *ct*Arg158 with *hs*Lys132 (*e*), of *ct*Trp121 with *hs*Leu144 (*f*) and of *ct*Glu129 and *ct*Asn107 with *hs*Ala104 and *hs*Pro82 (*g*) are shown.

~25-residue insertion in their PPIase domains, the sequence of which is conserved (20 of 25 residues; Supplementary Fig. S3). To investigate the structural consequences of this additional element, we produced and purified the PPIase domain of *C. thermophilum* (UniProt ID G0RY38; residues 1–201; *ct*Cwc27$^{1–201}$; Fig. 6*a*). Well diffracting crystals were obtained within 2 d (Fig. 6*b*). The structure was solved using the published structure of *hs*Cwc27$^{8–172}$ as a search model and was refined at 1.3 Å resolution to an R_{work} of 12.1% and an R_{free} of 14.3% (Table 1).

Comparison to the *hs*Cwc27 structures revealed that the fungal-specific element (residues 80–106 in *ct*Cwc27) formed an additional loop between β-strands 4 and 5 which did not alter the structure of the core PPIase domain (r.m.s.d. of 0.9 Å for 168 common C^{α} atoms between *ct*Cwc27 and *hs*Cwc27; Figs. 6*c*, 6*d* and Supplementary Fig. S6). The proline-binding centre of the *ct*Cwc27 PPIase domain contains the conserved residues Arg56, Phe61, Ile62, Gln64, Ala127, Asn128, Phe140, Glu148 and Leu149, as well as the nonconserved Asn153 (Fig. 6*e* and Supplementary Fig. S3). It is occupied in our structure by an ethylene glycol molecule, which is located approximately at the position of the glycerol molecule in the human structure. Again, the hydrophobic hydrocarbon backbone faces the hydrophobic pocket, whereas the hydroxyl groups are participating in an extensive hydrogen-bond network in the proline-binding pocket. One hydroxyl group directly interacts with Gln64 and *via* water molecules with Arg56, Asn128 and Asn128; the second hydroxyl group binds *via* water to Arg56 and to Asn153. In contrast to the situation in the human structure, in the *C. thermophilum* structure we can observe double conformations of Glu148 and Lys152, allowing both residues to participate in the hydrogen-bond network of the proline-binding centre (Fig. 6*e* and Supplementary Fig. S7). The cysteines of *hs*Cwc27 that form a disulfide bond in the oxidized PPIase-domain structure (PDB entry 2hq6) are not conserved in *C. thermophilum* (*hs*Cys44 *versus ct*Ser44 and *hs*Cys164 *versus ct*Ile192). The additional fungal-specific loop is internally stabilized by a buried arginine residue, *ct*Arg96, that engages in hydrogen bonds to *ct*Tyr79, *ct*Ala83, *ct*Phe84, *ct*Trp91 and, *via* water, to *ct*Asn100 and *ct*Gly102, as well as in a π-stacking interaction with *ct*Trp91 (Fig. 6*d*). It engages in extensive interactions with the PPIase core (loop residues *ct*Asp80, *ct*His99 and *ct*Asn100 hydrogen-bonding to *ct*Gly75, *ct*Ser77, *ct*Asn107, *ct*Glu129 and *ct*Thr134; *ct*Trp81 and *ct*Gln95 forming water-bridge interactions with *ct*Asp110 and *ct*Gly111). Based on these extensive interactions with the core domain, the loop could contribute to the fold stability of the protein. Alternatively, or in addition, based on its exposure at the surface of the PPIase domain it could form part of a protein–protein interaction site.

3.5. Structural basis for the increased thermal stability of *ct*Cwc27

The optimal growth temperature of *C. thermophilum* is 45–55°C and the organism can even tolerate temperatures of up to 60°C (La Touche, 1948; Amlacher *et al.*, 2011). In a comparative thermal stability assay, we measured the temperature-dependent unfolding of human (residues 1–173) and *C. thermophilum* (residues 1–201) Cwc27 PPIase-domain constructs. Both proteins were measured at three different concentrations, each with seven replicates, resulting in 21 measurements per protein. We observed T_m values of 48.12 ± 0.52 and 51.32 ± 0.12°C for the human and *C. thermophilum* proteins, respectively (Fig. 7*a* and Supplementary Table S1), resulting in a net stability gain of 3.2 ± 0.64°C for *ct*Cwc27$^{1–201}$. The start of the transition in the *C. thermophilum* protein was shifted upwards by ~3°C (38.14 ± 1.56°C compared with 35.24 ± 0.83 °C in the human protein) and exhibited a steeper slope (more cooperative unfolding) so that it completed at approximately the same temperature as in the human protein (55.71 ± 0.56°C compared with 55.38 ± 1.60°C in the human protein; Supplementary Table S1). Thus, the higher T_m of the *ct*Cwc27 PPIase domain is almost entirely accounted for by the stabilization of the fully folded state, as opposed to the stabilization of intermediate conformations.

In some cases, the higher thermal stability of thermotolerant proteins compared with their mesophilic counterparts can be explained by major structural differences, such as clamping 'thermo helices' (Than *et al.*, 1997; Auerbach *et al.*, 1998) or homodimerization (Dams *et al.*, 2000). A similar effect may be exerted by the fungus-specific loop of *ct*Cwc27, which is engaged in extensive interactions with the PPIase core (see above). However, the loop is presumably not the only reason for the increased thermal stability, as it is also found in the Cwc27 protein of the mesophilic fungus *N. crassa* (Supplementary Fig. S3). Another common strategy for thermal stabilization in proteins is the replacement of residues with left-handed side chains (φ, ψ of approximately 60°, 30°; Nicholson *et al.*, 1989) by glycine (90°, 0°) since the β-carbon, which is not present in glycine, causes steric hindrance with the carbonyl O atom of the same residue (Ishikawa *et al.*, 1993; Macedo-Ribeiro *et al.*, 1996, 2001). Although a substitution to glycine increases the entropy difference between the unfolded and folded state, the released conformational strain leads to an overall gain in thermostability (Ishikawa *et al.*, 1993). The amino-acid composition of the human and *C. thermophilum* PPIase domains is rather similar, but there is an overall reduction in long charged amino acids, *i.e.* Asp, Glu, Lys and Arg (*hs*, 25.1%; *ct*, 20.9%) and an increase in glycines (*hs*, 9.5%; *ct*, 13.2%) in *C. thermophilum* (Supplementary Table S2). At the structural level, we observed the replacement of residues with left-handed side chains by glycines: *hs*Ala47 (φ = 63.1°, ψ = 21.1°) by *ct*Gly47 (φ = 82.2°, ψ = 11.0°) and *hs*Asn51 (φ = 59.2°, ψ = 27.7°) by *ct*Gly51 (φ = 81.9°, ψ = −2.5°). An additional way to achieve higher thermostability is to increase the number of intramolecular interactions (Macedo-Ribeiro *et al.*, 1996; Auerbach *et al.*, 1997). In *ct*Cwc27, amino-acid substitutions that generate additional intramolecular interactions include changes to threonines (*ct*Thr33 *versus hs*Ala33 hydrogen-bonding to *ct*Phe29 and *ct*Cys37; *ct*Thr36 *versus hs*Ala36 hydrogen-bonding to *ct*Glu111 and *via* water to *ct*Lys109 and *ct*Thr154; *ct*Thr180 *versus hs*Asp152 hydrogen-bonding to *ct*Arg182 and *ct*Ile177; Figs. 7*b* and 7*c*), although

the fraction of threonines is only slightly increased (*hs*Cwc27, 6.0%; *ct*Cwc27, 7.1%; Supplementary Table S2). Additional interactions are also owing to histidines (*ct*His43 *versus* *hs*Leu43 hydrogen-bonding to *ct*Tyr48, *via* water with *ct*Asp46 and π-stacking with *ct*Tyr48; Fig. 7*d*), arginines (*ct*Arg158 *versus* *hs*Lys132 interacting with *ct*Glu27 and hydrogen-bonding to *ct*Gln10; Fig. 7*e*), tryptophan (*ct*Trp171 *versus* *hs*Leu144 hydrogen-bonding to *ct*Gln21 and π-stacking with *ct*Phe54; Fig. 7*f*) and glutamate or asparagine (*ct*Glu129 *versus* *hs*Ala104 hydrogen-bonding to *ct*Asn107 *versus* *hs*Pro82; Fig. 7*g*). Thus, as in many other thermally stabilized proteins (Macedo-Ribeiro *et al.*, 1996, 2001; Auerbach *et al.*, 1997), the increased thermal stability of the *ct*Cwc27 PPIase domain compared with that of *hs*Cwc27 is most likely owing to the combined effects of several individual amino-acid exchanges. Several of the contributing amino-acid variations, such as *ct*His43 *versus* *hs*Leu43 or *ct*Trp171 *versus* *hs*Leu144, are unique to *C. thermophilum* among the Cwc27 proteins identified in our study.

AU and MCW conceived and designed the experiments, analyzed and interpreted the data and wrote the manuscript. AU performed the experiments. We thank Thorsten Fogge for help in crystallization screening, Haydar Bulut and Oleg Ganichkin for help with the thermostability assay and Bernhard Loll for help with diffraction data collection (Laboratory of Structural Biochemistry, Freie Universität Berlin, Germany). We acknowledge access to beamline BL14.2 of the BESSY II storage ring, Berlin, Germany *via* the Joint Berlin MX-Laboratory sponsored by the Helmholtz Zentrum Berlin für Materialien und Energie, the Freie Universität Berlin, the Humboldt-Universität zu Berlin, the Max-Delbrück Centrum and the Leibniz-Institut für Molekulare Pharmakologie. This work was supported by grant WA1126/7-1 from the Deutsche Forschungsgemeinschaft (to MCW).

References

Adrain, C. & Freeman, M. (2012). *Nature Rev. Mol. Cell Biol.* **13**, 489–498.

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.

Agafonov, D. E., Deckert, J., Wolf, E., Odenwälder, P., Bessonov, S., Will, C. L., Urlaub, H. & Lührmann, R. (2011). *Mol. Cell. Biol.* **31**, 2667–2682.

Amlacher, S., Sarges, P., Flemming, D., van Noort, V., Kunze, R., Devos, D. P., Arumugam, M., Bork, P. & Hurt, E. (2011). *Cell*, **146**, 277–289.

Anderson, S. K., Gallinger, S., Roder, J., Frey, J., Young, H. A. & Ortaldo, J. R. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 542–546.

Andreotti, A. H. (2003). *Biochemistry*, **42**, 9515–9524.

Auerbach, G., Huber, R., Grättinger, M., Zaiss, K., Schurig, H., Jaenicke, R. & Jacob, U. (1997). *Structure*, **5**, 1475–1483.

Auerbach, G., Ostendorp, R., Prade, L., Korndörfer, I., Dams, T., Huber, R. & Jaenicke, R. (1998). *Structure*, **6**, 769–781.

Baker, E. K., Colley, N. J. & Zuker, C. S. (1994). *EMBO J.* **13**, 4886–4895.

Baneyx, F. (2004). *Microb. Cell Fact.* **3**, 6.

Bose, S., Weikl, T., Bügl, H. & Buchner, J. (1996). *Science*, **274**, 1715–1717.

Bossard, M. J., Koser, P. L., Brandt, M., Bergsma, D. J. & Levy, M. A. (1991). *Biochem. Biophys. Res. Commun.* **176**, 1142–1148.

Bourquin, J., Stagljar, I., Meier, P., Moosmann, P., Silke, J., Baechi, T., Georgiev, O. & Schaffner, W. (1997). *Nucleic Acids Res.* **25**, 2055–2061.

Brazin, K. N., Mallis, R. J., Fulton, D. B. & Andreotti, A. H. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 1899–1904.

Chatterji, U., Bobardt, M., Selvarajah, S., Yang, F., Tang, H., Sakamoto, N., Vuagniaux, G., Parkinson, T. & Gallay, P. (2009). *J. Biol. Chem.* **284**, 16998–17005.

Chen, Y.-I. G., Moore, R. E., Ge, H. Y., Young, M. K., Lee, T. D. & Stevens, S. W. (2007). *Nucleic Acids Res.* **35**, 3928–3944.

Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B. & Bork, P. (2006). *Science*, **311**, 1283–1287.

Dams, T., Auerbach, G., Bader, G., Jacob, U., Ploom, T., Huber, R. & Jaenicke, R. (2000). *J. Mol. Biol.* **297**, 659–672.

Danielson, P. E., Forss-Petter, S., Brow, M. A., Calavetta, L., Douglass, J., Milner, R. J. & Sutcliffe, J. G. (1988). *DNA*, **7**, 261–267.

Dartigalongue, C. & Raina, S. (1998). *EMBO J.* **17**, 3968–3980.

Davis, J. M., Boswell, B. A. & Bächinger, H. P. (1989). *J. Biol. Chem.* **264**, 8956–8962.

Davis, T. L., Walker, J. R., Campagna-Slater, V., Finerty, P. J., Paramanathan, R., Bernstein, G., MacKenzie, F., Tempel, W., Ouyang, H., Lee, W. H., Eisenmesser, E. Z. & Dhe-Paganon, S. (2010). *PLoS Biol.* **8**, e1000439.

Davis, T. L., Walker, J. R., Ouyang, H., MacKenzie, F., Butler-Cole, C., Newman, E. M., Eisenmesser, E. Z. & Dhe-Paganon, S. (2008). *FEBS J.* **275**, 2283–2295.

Dorfman, T., Weimann, A., Borsetti, A., Walsh, C. T. & Göttlinger, H. G. (1997). *J. Virol.* **71**, 7110–7113.

Eckert, B., Martin, A., Balbach, J. & Schmid, F. X. (2005). *Nature Struct. Mol. Biol.* **12**, 619–623.

Edgar, R. C. (2004). *Nucleic Acids Res.* **32**, 1792–1797.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H. & Lührmann, R. (2009). *Mol. Cell*, **36**, 593–608.

Fanghänel, J. & Fischer, G. (2004). *Front. Biosci.* **9**, 3453–3478.

Fischer, G., Bang, H. & Mech, C. (1984). *Biomed. Biochim. Acta*, **43**, 1101–1111.

Fischer, G., Wittmann-Liebold, B., Lang, K., Kiefhaber, T. & Schmid, F. X. (1989). *Nature (London)*, **337**, 476–478.

Fransson, C., Freskgård, P. O., Herbertsson, H., Johansson, A., Jonasson, P., Mårtensson, L. G., Svensson, M., Jonsson, B. H. & Carlsson, U. (1992). *FEBS Lett.* **296**, 90–94.

Freeman, B. C., Toft, D. O. & Morimoto, R. I. (1996). *Science*, **274**, 1718–1720.

Freskgård, P. O., Bergenhem, N., Jonsson, B. H., Svensson, M. & Carlsson, U. (1992). *Science*, **258**, 466–468.

Fritz-Laylin, L. K. *et al.* (2010). *Cell*, **140**, 631–642.

Fu, A., He, Z., Cho, H. S., Lima, A., Buchanan, B. B. & Luan, S. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 15947–15952.

Galat, A. (1999). *Arch. Biochem. Biophys.* **371**, 149–162.

Gething, M.-J. & Sambrook, J. (1992). *Nature (London)*, **355**, 33–45.

Goel, M., Garcia, R., Estacion, M. & Schilling, W. P. (2001). *J. Biol. Chem.* **276**, 38762–38773.

Göthel, S. F. & Marahiel, M. A. (1999). *Cell. Mol. Life Sci.* **55**, 423–436.

Gullerova, M., Barta, A. & Lorkovic, Z. J. (2006). *RNA*, **12**, 631–643.

Gullerova, M., Barta, A. & Lorkovic, Z. J. (2007). *Mol. Cell. Biol.* **27**, 3601–3611.

Haendler, B., Keller, R., Hiestand, P. C., Kocher, H. P., Wegmann, G. & Movva, N. R. (1989). *Gene*, **83**, 39–46.

Handschumacher, R. E., Harding, M. W., Rice, J., Drugge, R. J. & Speicher, D. W. (1984). *Science*, **226**, 544–547.

Hayano, T., Takahashi, N., Kato, S., Maki, N. & Suzuki, M. (1991). *Biochemistry*, **30**, 3041–3048.

Hedges, S. B., Blair, J. E., Venturi, M. L. & Shoe, J. L. (2004). *BMC Evol. Biol.* **4**, 2.

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C. L., Pena, V., Lührmann, R. & Stelzl, U. (2012). *Mol. Cell*, **45**, 567–580.

Hong, F., Lee, J., Song, J.-W., Lee, S. J., Ahn, H., Cho, J. J., Ha, J. & Kim, S. S. (2002). *FASEB J.* **16**, 1633–1635.

Horowitz, D. S., Lee, E. J., Mabon, S. A. & Misteli, T. (2002). *EMBO J.* **21**, 470–480.

Howard, B. R., Vajdos, F. F., Li, S., Sundquist, W. I. & Hill, C. P. (2003). *Nature Struct. Biol.* **10**, 475–481.

Huang, L.-L., Zhao, X.-M., Huang, C.-Q., Yu, L. & Xia, Z.-X. (2005). *Acta Cryst.* D**61**, 316–321.

Iki, T., Yoshikawa, M., Meshi, T. & Ishikawa, M. (2012). *EMBO J.* **31**, 267–278.

Ishikawa, K., Nakamura, H., Morikawa, K., Kimura, S. & Kanaya, S. (1993). *Biochemistry*, **32**, 7136–7142.

Ke, H. (1992). *J. Mol. Biol.* **228**, 539–550.

Ke, H., Mayrose, D., Belshaw, P. J., Alberg, D. G., Schreiber, S. L., Chang, Z. Y., Etzkorn, F. A., Ho, S. & Walsh, C. T. (1994). *Structure*, **2**, 33–44.

Kempken, F. (2013). *Appl. Microbiol. Biotechnol.* **97**, 4235–4241.

Kiefhaber, T., Grunert, H. P., Hahn, U. & Schmid, F. X. (1990). *Biochemistry*, **29**, 6475–6480.

Kimmins, S. & MacRae, T. H. (2000). *Cell Stress Chaperones*, **5**, 76–86.

Kumari, S., Roy, S., Singh, P., Singla-Pareek, S. L. & Pareek, A. (2012). *Plant Signal. Behav.* **8**, e22734.

Ladrière, J.-M., Georis, I., Guérineau, M. & Vandenhaute, J. (2000). *Gene*, **255**, 83–91.

Lang, K., Schmid, F. X. & Fischer, G. (1987). *Nature (London)*, **329**, 268–270.

La Touche, C. J. (1948). *Nature (London)*, **161**, 320.

Leung, A. W. C., Varanyuwatana, P. & Halestrap, A. P. (2008). *J. Biol. Chem.* **283**, 26312–26323.

Lilie, H., Lang, K., Rudolph, R. & Buchner, J. (1993). *Protein Sci.* **2**, 1490–1496.

Lin, D.-T. & Lechleiter, J. D. (2002). *J. Biol. Chem.* **277**, 31134–31141.

Lin, L.-N., Hasumi, H. & Brandts, J. F. (1988). *Biochim. Biophys. Acta*, **956**, 256–266.

Liu, J., Chen, C. M. & Walsh, C. T. (1991). *Biochemistry*, **30**, 2306–2310.

Liu, J. & Walsh, C. T. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 4028–4032.

Lu, K. P., Hanes, S. D. & Hunter, T. (1996). *Nature (London)*, **380**, 544–547.

Lummis, S. C. R., Beene, D. L., Lee, L. W., Lester, H. A., Broadhurst, R. W. & Dougherty, D. A. (2005). *Nature (London)*, **438**, 248–252.

MacArthur, M. W. & Thornton, J. M. (1991). *J. Mol. Biol.* **218**, 397–412.

Macedo-Ribeiro, S., Darimont, B., Sterner, R. & Huber, R. (1996). *Structure*, **4**, 1291–1301.

Macedo-Ribeiro, S., Martins, B. M., Pereira, P. J., Buse, G., Huber, R. & Soulimane, T. (2001). *J. Biol. Inorg. Chem.* **6**, 663–674.

Maris, C., Dominguez, C. & Allain, F. H. T. (2005). *FEBS J.* **272**, 2118–2131.

Matouschek, A., Rospert, S., Schmid, K., Glick, B. S. & Schatz, G. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 6319–6323.

McCoy, A. J. (2007). *Acta Cryst.* D**63**, 32–41.

Mesa, A., Somarelli, J. A. & Herrera, R. J. (2008). *FEBS Lett.* **582**, 2345–2351.

Mi, H., Kops, O., Zimmermann, E., Jäschke, A. & Tropschug, M. (1996). *FEBS Lett.* **398**, 201–205.

Mikol, V., Kallen, J. & Walkinshaw, M. D. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 5183–5186.

Morris, D. P., Phatnani, H. P. & Greenleaf, A. L. (1999). *J. Biol. Chem.* **274**, 31583–31587.

Mukherjee, K., Sharma, M., Urlaub, H., Bourenkov, G. P., Jahn, R., Südhof, T. C. & Wahl, M. C. (2008). *Cell*, **133**, 328–339.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Nelson, C. J., Santos-Rosa, H. & Kouzarides, T. (2006). *Cell*, **126**, 905–916.

Nicholson, H., Söderlind, E., Tronrud, D. E. & Matthews, B. W. (1989). *J. Mol. Biol.* **210**, 181–193.

Ou, W.-B., Luo, W., Park, Y.-D. & Zhou, H.-M. (2008). *Protein Sci.* **10**, 2346–2353.

Pal, D. & Chakrabarti, P. (1999). *J. Mol. Biol.* **294**, 271–288.

Pemberton, T. J. (2006). *BMC Genomics*, **7**, 244.

Ranganathan, R., Lu, K. P., Hunter, T. & Noel, J. P. (1997). *Cell*, **89**, 875–886.

Rappsilber, J., Ryder, U., Lamond, A. I. & Mann, M. (2002). *Genome Res.* **12**, 1231–1245.

Rassow, J., Mohrs, K., Koidl, S., Barthelmess, I. B., Pfanner, N. & Tropschug, M. (1995). *Mol. Cell. Biol.* **15**, 2654–2662.

Reidt, U., Reuter, K., Achsel, T., Ingelfinger, D., Lührmann, R. & Ficner, R. (2000). *J. Biol. Chem.* **275**, 7439–7442.

Reidt, U., Wahl, M. C., Fasshauer, D., Horowitz, D. S., Lührmann, R. & Ficner, R. (2003). *J. Mol. Biol.* **331**, 45–56.

Rost, B., Yachdav, G. & Liu, J. (2004). *Nucleic Acids Res.* **32**, W321–W326.

Sarkar, P., Reichman, C., Saleh, T., Birge, R. B. & Kalodimos, C. G. (2007). *Mol. Cell*, **25**, 413–426.

Scanlan, M. J., Chen, Y.-T., Williamson, B., Gure, A. O., Stockert, E., Gordan, J. D., Türeci, Ö., Sahin, U., Pfreundschuh, M. & Old, L. J. (1998). *Int. J. Cancer*, **76**, 652–658.

Schlatter, D., Thoma, R., Küng, E., Stihle, M., Müller, F., Borroni, E., Cesura, A. & Hennig, M. (2005). *Acta Cryst.* D**61**, 513–519.

Schlegel, J., Redzic, J. S., Porter, C. C., Yurchenko, V., Bukrinsky, M., Labeikovsky, W., Armstrong, G. S., Zhang, F., Isern, N. G., DeGregori, J., Hodges, R. & Eisenmesser, E. Z. (2009). *J. Mol. Biol.* **391**, 518–535.

Schneuwly, S., Shortridge, R. D., Larrivee, D. C., Ono, T., Ozaki, M. & Pak, W. L. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 5390–5394.

Schönbrunner, E. R., Mayer, S., Tropschug, M., Fischer, G., Takahashi, N. & Schmid, F. X. (1991). *J. Biol. Chem.* **266**, 3630–3635.

Seoighe, C. & Wolfe, K. H. (1999). *Gene*, **238**, 253–261.

Shieh, B.-H., Stamnes, M. A., Seavello, S., Harris, G. L. & Zuker, C. S. (1989). *Nature (London)*, **338**, 67–70.

Smith, M. R., Willmann, M. R., Wu, G., Berardini, T. Z., Möller, B., Weijers, D. & Poethig, R. S. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 5424–5429.

Stegmann, C. M., Lührmann, R. & Wahl, M. C. (2010). *PLoS One*, **5**, e10013.

Stewart, D. E., Sarkar, A. & Wampler, J. E. (1990). *J. Mol. Biol.* **214**, 253–260.

Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.

Takahashi, N., Hayano, T. & Suzuki, M. (1989). *Nature (London)*, **337**, 473–475.

Tan, Y.-J., Oliveberg, M., Otzen, D. E. & Fersht, A. R. (1997). *J. Mol. Biol.* **269**, 611–622.

Teigelkamp, S., Achsel, T., Mundt, C., Göthel, S. F., Cronshagen, U., Lane, W. S., Marahiel, M. & Lührmann, R. (1998). *RNA*, **4**, 127–141.

Than, M. E., Hof, P., Huber, R., Bourenkov, G. P., Bartunik, H. D., Buse, G. & Soulimane, T. (1997). *J. Mol. Biol.* **271**, 629–644.

Ulrich, A., Andersen, K. R. & Schwartz, T. U. (2012). *PLoS One*, **7**, e53360.

Wahl, M. C., Will, C. L. & Lührmann, R. (2009). *Cell*, **136**, 701–718.

Wang, Y., Han, R., Zhang, W., Yuan, Y., Zhang, X., Long, Y. & Mi, H. (2008). *FEBS Lett.* **582**, 835–839.

Wang, P. & Heitman, J. (2005). *Genome Biol.* **6**, 226.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P. & Burge, C. B. (2008). *Nature (London)*, **456**, 470–476.

Wang, X., Zhang, S., Zhang, J., Huang, X., Xu, C., Wang, W., Liu, Z., Wu, J. & Shi, Y. (2010). *J. Biol. Chem.* **285**, 4951–4963.
Watashi, K., Ishii, N., Hijikata, M., Inoue, D., Murata, T., Miyanari, Y. & Shimotohno, K. (2005). *Mol. Cell*, **19**, 111–122.
Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. (2009). *Bioinformatics*, **25**, 1189–1191.
Wolfe, K. H. & Shields, D. C. (1997). *Nature (London)*, **387**, 708–713.
Xu, C., Zhang, J., Huang, X., Sun, J., Xu, Y., Tang, Y., Wu, J., Shi, Y., Huang, Q. & Zhang, Q. (2006). *J. Biol. Chem.* **281**, 15900–15908.
Yaffe, M. B., Schutkowski, M., Shen, M., Zhou, X. Z., Stukenberg, P. T., Rahfeld, J.-U., Xu, J., Kuang, J., Kirschner, M. W., Fischer, G., Cantley, L. C. & Lu, K. P. (1997). *Science*, **278**, 1957–1960.
Yang, H.-P., Zhong, H.-N. & Zhou, H.-M. (1997). *Biochim. Biophys. Acta*, **1338**, 147–150.
Yurchenko, V., Zybarth, G., O'Connor, M., Dai, W. W., Franchin, G., Hao, T., Guo, H., Hung, H.-C., Toole, B., Gallay, P., Sherry, B. & Bukrinsky, M. (2002). *J. Biol. Chem.* **277**, 22959–22965.
Zhao, Y., Chen, Y., Schutkowski, M., Fischer, G. & Ke, H. (1997). *Structure*, **5**, 139–146.
Zhou, Z., Licklider, L. J., Gygi, S. P. & Reed, R. (2002). *Nature (London)*, **419**, 182–185.
Zhou, X. Z., Lu, P. J., Wulf, G. & Lu, K. P. (1999). *Cell. Mol. Life Sci.* **56**, 788–806.
Zydowsky, L. D., Etzkorn, F. A., Chang, H. Y., Ferguson, S. B., Stolz, L. A., Ho, S. I. & Walsh, C. T. (1992). *Protein Sci.* **1**, 1092–1099.