Structural and Functional Insights into the DNA Replication Factor Cdc45 Reveal an Evolutionary Relationship to the DHH Family of Phosphoesterases

Ivet Krastanova1, Vincenzo Sannino1, Heinz Amenitsch2, Opher Gileadi3, Francesca M. Pisani4,5, and Silvia Onesti1,2

From the 1Structural Biology Laboratory, Sincrotrone Trieste, Trieste 34149, Italy, the 2Istituto di Biochimica delle Proteine, Consiglio Nazionale delle Ricerche, Napoli 80131, Italy, the 3Institut of Biophysics and Nanosystems Research, Austrian Academy of Sciences, Schmiedlstrasse 6, Graz 8042, Austria, and the 4Structural Genomics Consortium, Oxford OX3 7DQ, United Kingdom

Background: Although Cdc45 is a key replication factor, there are no biochemical or structural studies on the isolated protein.

Results: We report the first purification and biochemical characterization of human Cdc45, as well as the first structural data on the isolated Cdc45 by small angle x-ray scattering.

Conclusion: Cdc45 is related to the RecJ/DHH family of phosphoesterases and binds single-stranded DNA.

Significance: The similarity has important evolutionary implications.

Cdc45 is an essential protein conserved in all eukaryotes and is involved both in the initiation of DNA replication and the progression of the replication fork. With GINS, Cdc45 is an essential cofactor of the Mcm2–7 replicative helicase complex. Despite its importance, no detailed information is available on either the structure or the biochemistry of the protein. Intriguingly, whereas homologues of both GINS and Mcm proteins have been described in Archaea, no counterpart for Cdc45 is known. Herein we report a bioinformatic analysis that shows a weak but significant relationship among eukaryotic Cdc45 proteins and a large family of phosphoesterases that has been described as the DHH family, including inorganic pyrophosphatases and RecJ ssDNA exonucleases. These enzymes catalyze the hydrolysis of phosphodiester bonds via a mechanism involving two Mn2+ ions. Only a subset of the amino acids that coordinates Mn2+ is conserved in Cdc45. We report biochemical and structural data on the recombinant human Cdc45 protein, consistent with the proposed DHH family affiliation. Like the RecJ exonucleases, the human Cdc45 protein is able to bind single-stranded, but not double-stranded DNA. Small angle x-ray scattering data are consistent with a model compatible with the crystallographic structure of the RecJ/DHH family members.

Cdc45 is an essential factor required for the establishment (1–3) and progression (4–7) of the DNA replication fork in eukaryotic cells. As many other DNA replication factors, Cdc45 is more abundant in proliferating cells, whereas it is almost absent from long term quiescent, terminally differentiated, and senescent cells (8). Several genetic studies, two-hybrid screens, and co-immunoprecipitation analyses revealed that Cdc45 interacts with a number of other replication factors, including the Mcm2–7 complex, GINS, MCM10, replication protein A, DNA polymerases α, δ, and ε, the origin recognition complex subunit 2, and TopBP1 (for a review see Ref. 9). In a variety of eukaryotic organisms Cdc45 has been found to stably associate with Mcm2–7 and GINS to form a complex (the CMG),4 which is believed to act as the DNA helicase at the replication fork (7, 10–13). This hypothesis has been reinforced by the demonstration that the Drosophila melanogaster CMG complex can be reconstituted by co-producing its protein components in baculovirus-infected insect cells and is found to possess a robust DNA-unwinding activity in vitro, whereas the Mcm2–7 complex alone is almost completely unable to unwind duplex DNA (14). This analysis suggests that Cdc45 and GINS are helicase auxiliary factors whose association with the Mcm hetero-hexameric ring is absolutely required to reconstitute an active complex.

The critical biological function played by Cdc45 is underscored by its ubiquitous distribution and high degree of sequence conservation from yeast to man. Nevertheless, not much is known on the exact role of Cdc45 either in the pre-initiation complex or the CMG helicase, due to the lack of biochemical studies on the isolated protein. The analysis of the primary structure of Cdc45 has failed to reveal the presence of any significant similarity to known protein family or any characteristic sequence motif.

Due to the complexity of the eukaryotic DNA replication machinery, most of the information on the structure and biochemistry of replication proteins has been inferred from the study of the simpler archaeal system. In fact, Archaea possess a simplified version of the Mcm (15, 16) and GINS (17–19) complexes, but so far no archaenal homologue of Cdc45 has been identified.

8 The work was supported by a Friuli Venezia Giulia grant (DAGEAST to I.K.) and by the Associazione Italiana per la Ricerca sul Cancro (Grant IG10646 to S. O. and Grant IG9087 to F. M. P.).
9 This article contains supplemental Figs. S1–S3 and Table S1.
1 Both authors contributed equally to this work.
2 To whom correspondence may be addressed. Tel.: 39-081-613-2292; Fax: 39-081-613-2277; E-mail: fm.pisani@ibp.cnr.it.
3 To whom correspondence may be addressed. Tel.: 39-040-375-8451; Fax: 30-40-938-0902; E-mail: silvia.onesti@elefetta.trieste.it.
4 The abbreviations used are: CMG, Cdc45/Mcm2–7/GINS complex; TEV, tobacco etch virus; SAXS, small angle x-ray scattering.
Experimental Procedures

Protein Expression and Purification—The nucleotide sequence encoding the human Cdc45 (hCdc45) was amplified using the Platinum PfX DNA polymerase (Invitrogen), using a cDNA clone from the mammalian Gene Collection (IMAGE ID 2964592) as a template, and the following primers: 5'-TTAA-GAAGGAGATACTGTTCTGGATTTCCGCA-AAG-3' and 5'-GATTGGAAGTAGAGGTTCTGCAGGA-CAGGAGAAATATGGCC-3'.

The amplified fragment was then Ligation independent cloned into the pNIC-CTHF vector (Structural Genomics Consortium, Oxford (20)) for protein expression as a fusion with a C-terminal His6-FLAG tag cleavable with tobacco etch virus (TEV) protease.

Expression of hCdc45 was carried out in *Escherichia coli* BL21 (DE3)-R3-Rosetta cells (SGC, Oxford) grown overnight at 25 °C in TB broth, following 0.1 mM isopropyl 1-thio-β-d-galactopyranoside induction. A frozen cell pellet (corresponding to 1 liter of culture) was resuspended in 100 ml of buffer A (50 mM sodium phosphate buffer, pH 7.4, 1 mM NaCl, 10% (v/v) glycerol, 2 mM β-mercaptoethanol) containing EDTA-free protease inhibitor mixture (Roche Applied Science), 50 units/ml benzonase nuclease (Novagen), and 1 mg/ml lysozyme (Sigma-Aldrich) and incubated for 2 h at 4 °C. Cells were further disrupted by sonication (Bandelin Sonopuls HD3200 ultrasonic homogenizer). The insoluble material was removed by centrifugation (17,000 rpm for 1 h at 4 °C, Beckman Allegra 64R), and the supernatant was loaded onto a 5-ml HiTrap chelating column (GE Healthcare) previously charged with Ni2+ (NiSO4) and equilibrated with buffer B. An imidazole step gradient (25 mM, 250 mM, and 500 mM) was applied, and protein elution was achieved using buffer A plus 250 mM imidazole. Fractions containing the hCdc45 were pooled and subsequently loaded onto a Superdex 200 16/60 size-exclusion column (GE Healthcare) equilibrated with buffer B (50 mM sodium phosphate, pH 7.0, 150 mM NaCl, 5% (v/v) glycerol, 2 mM β-mercaptoethanol). The eluted protein was incubated with TEV protease for overnight tag cleavage at 4 °C. 12% SDS-PAGE was carried out to check that the cleavage was complete. The hCdc45-TEV mix was loaded onto a 5-ml Heparin column (GE Healthcare), previously equilibrated with buffer B. The cleaved hCdc45 was eluted from the column applying a linear (150 mM to 1 M) NaCl gradient. Finally the protein was subjected to size-exclusion chromatography in buffer D (20 mM Tris-HCl, pH 7.9, 150 mM NaCl, 5% (v/v) glycerol, 2 mM β-mercaptoethanol).

EMSAs—The synthetic oligonucleotide used as single-stranded DNA had the following sequence: 5'-TCTACCT-GGACGACCGGTATATAGGCCCCCTATATAGGG-CACAGGTCCATCA-3'. A complementary synthetic oligonucleotide used to prepare the blunt DNA duplex had the following sequence: 5'-TGATTGACCTGCTGCGCCCTATATATAGGCCCCCTATATACCGGTGTCAGTCTAGGA-3'. The first oligonucleotide was labeled using T4 poly-nucleotide kinase and [γ-32P]ATP and annealed to a 2-fold molar excess of the cold complementary strand to prepare the double-stranded DNA ligand. For the DNA mobility shift assays, 10-μl mixtures were prepared that contained 100 fmol of 32P-labeled DNA in 20 mM Tris-HCl, pH 7.5, and the indicated amounts of protein (0.5–5 μg). Following incubation for 20 min at 27 °C, complexes were separated by electrophoresis through 5% polyacrylamide/bis gels (19:1) in 0.5× TBE (1× TBE (89 mM Tris Base, 89 mM Boric Acid, 2 mM EDTA, pH 8.3)). Gels were dried down and analyzed by phosphorimaging. Experiments were performed in triplicate, and the results were averaged. The *error bars* on the *graphs* are the standard error.

Electrophoretic mobility shift assays (EMSAs) were carried out after incubation of hCdc45-His-FLAG-DNA complexes with anti-FLAG monoclonal antibodies (Abcam). For these experiments 10-μl mixtures were prepared that contained 50 fmol of 32P-labeled oligonucleotide in 20 mM Tris-HCl, pH 7.5, and 5 μg of hCdc45-His-FLAG. Following incubation for 20 min at 27 °C, anti-FLAG antibody was added (0.5, 1, and 2 μg in 2 μl of the following buffer: 10 mM sodium phosphate, pH 7.4, 150 mM NaCl, 50% glycerol; an equal volume of buffer (2 μl) was added into the samples where the antibody was omitted, and the incubation was continued for additional 30 min. The mixtures were subjected to electrophoresis, as previously described, and the gels were analyzed by phosphorimaging.

Analytical Gel Filtration and DNA-binding Activity of hCdc45—An aliquot of purified hCdc45 (1.9 mg) was loaded onto an analytical gel-filtration column (Bio-Sil SEC-250, BioRad). The column was developed with 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM β-mercaptoethanol, 5% (v/v) glycerol at a flow rate of 1.0 ml/min. Fractions (200 μl) were collected, and aliquots (5 μl) were run through an 8% (w/v) polyacrylamide-SDS gel. The gel-filtration column was calibrated using the following markers: tyroglobulin, bovine γ-globulin, chicken ovalbumin, equin myoglobin, and vitamin B-12. The DNA-biding activity of each peak fraction (aliquots of 1 μl) was analyzed by EMSAs using radiolabeled ssDNA as a probe, as previously described.

SAXS—The synchrotron scattering data were collected at the Austrian small angle x-ray scattering (SAXS) beamline of the electron storage ring ELETTRA (21) at a wavelength λ of 1.54 Å. A Pilatus 100K (Tectris, Baden, Switzerland) was used as a detector, and a sample distance of 0.75 m was set to resolve the momentum transfer, q (q = 4π sin(θ)/λ, with θ as half scattering angle), in the range from 0.17 to 4.5 nm⁻¹. Samples of recombinant hCdc45 at a concentration of 0.44, 0.90 mg/ml, and 1.85 mg/ml in buffer D were used. All samples were kept and measured in a 1.5-mm glass capillary (Glass, Berlin, Germany) at 8 °C. The three 30-s exposures have been averaged, because the comparison of the first and last pattern did not show any effect of radiation damage.

The primary data reduction was performed using IGOR Pro (Wavemeters, Lake Oswego, OR), which was also used to estimate the molecular mass of the solutes by comparing them to the scattering of water. The reduced data were treated further with the ATSAS program collection (22). The indirect Fourier transformation was calculated by using the program GNOM (23), which determines the distance distribution function, the radius of gyration Rg the forward scattering I(0), as well as the maximal dimension Dmax of the proteins. The results are summarized in supplemental Table S1, in which additionally the
molecular mass of each sample has been calculated. As the sample c (1.85 mg/ml) showed some evidence of aggregation, as seen by the SAXS pattern change at \( q < 0.4 \) \( \text{nm}^{-1} \) and the difference in the relative molecular mass, the low-\( q \)-data of sample b (0.9 mg/ml) has been merged with the high-\( q \)-data of the sample c (1.85 mg/ml; supplemental Fig. S1). The initial volume was a sphere with a \( D_{\text{max}} \) of 12.5 nm consisting of 5594 individual spheres (diameter 0.32 nm). The best 44 models have been averaged to obtain the final model with the program DAMAVER (25).

**RESULTS AND DISCUSSION**

*Cdc45 Shows Sequence Similarity to Archaeal Proteins Belonging to the DHH Family of Phosphoesterases*—We carried out a bioinformatic analysis on Cdc45. Database searches with the Position-Specific Interactive BLAST algorithm, using the human Cdc45 sequence as search model and default parameters, identified weak but significant similarity to two archaeal sequences on the second iteration run. The proteins were annotated as “phosphoesterase domain-containing protein” from *Candidatus kororacecum cryptofilum OPF8* (E-value: \( 2 \times 10^{-4} \)) and “putative single-stranded DNA-specific exonuclease Rec” from *Methanocella paludicola SANAIE* (E-value: \( 3 \times 10^{-4} \). They both belong to the DHH family of phosphoesterases that was first described by Aravind and Koonin (26). Both sequences have been recently added to the databases, explaining why the similarity had not been previously identified. Although the sequence similarity detected by BLAST involved only the first 130–140 residues, multiple alignments between Cdc45 and putative archaeal orthologues show that some similarity can be noticed throughout the entire sequence (Fig. 1).

The DHH family is defined by a number of conserved sequence motifs and can be split into two distinct clusters: subfamily 1 (whose prototype is the *Escherichia coli* RecJ protein), is so far represented only in Archaea and Bacteria; subfamily 2 (which includes the *Drosophila melanogaster* Prune protein and the *Saccharomyces cerevisiae* exopolysphosphatase PPX1), is also found in Eukarya. Crystallographic analysis of some DHH proteins (27–29) has revealed the presence of a catalytic core formed by two domains: in the N-terminal domain invariant residues (aspatic acid and histidine residues) coordinate two metals ions (typically manganese ions), suggesting a two-metal mediated hydrolysis reaction.

Because Archaea possess multiple RecJ-like proteins, we have selected the sequences displaying the closest match to the *Candidatus* and *Methanocella* homologues. To distinguish the archaeal RecJ/DHH proteins that may be putative Cdc45 orthologues from other members of the DHH family, we will use the notation Rec\(^{Cdc45} \). However, most Archaea seem to possess two close paralogues of the putative Cdc45 counterpart. A study, focused on detecting the archaenal equivalent of *E. coli* RecJ, identifies two highly similar proteins from *Methanocaldococcus jannaschii* (MJ0977 and MJ0831). Although both were able to partially complement a RecJ mutant phenotype in *E. coli*, ssDNA nuclease activity could only be observed for the MJ0977 protein (30). Two highly similar Rec\(^{Cdc45} \) proteins (AF0699 and AF0735) were also found associated with a replication protein network in *Archaeoglobus fulgidus* (31). More recently, a RecJ-like single-stranded 5’-3’ DNA exonuclease from *Thermococccus kodakaraensis* was found to physically associate with GINS, and this association was found to stimulate its exonuclease activity (32).

A putative RecJ homologue has been reported to co-purify with Mcm and GINS from cellular extracts of *Sulfolobus solfataricus* (SSO295 (18)). This sequence is rather divergent from most of the archaeal proteins mentioned above and lacks some of the motifs; in particular, some of the conserved residues putatively responsible for the nuclease activity are absent. An open reading frame highly similar to SSO295 is present in most *Sulfolobales* genomes. Divergent sequences are also detected in other Crenarchaeota, such as *Aeropyrum pernix*.

Both the eukaryotic Cdc45 proteins and the archaeal Rec\(^{Cdc45} \) sequences match only to the RecJ catalytic core (domains I and II, as defined in the crystal structure of *Thermus thermophilus* RecJ (27), while lacking the 50-residues N-terminal extension as well as domains III and IV. These additional elements form a closed-ring structure that is predicted to encircle ssDNA.

Cdc45 sequences only partially retain the motifs typical of the DHH family that are involved in metal binding. Whereas in motif 1 both aspartic residues coordinating the metal are conserved (26,DVD), the DHH motif 3, which gives the name to the family, is mutated to "DTH. Although a number of residues belonging to motifs 2 and 4 are conserved suggesting that the protein fold is similar in those regions, the key residues for metal binding and catalysis are not conserved, with the aspartate of motif 2 mutated to asparagine, and the aspartate of motif 4 to a glutamine (Fig. 1). In contrast, most of the archaeal Rec\(^{Cdc45} \) comprise all of the motifs that are conserved in the bacterial RecJ exonucleases, with the exception of the putative orthologues from *Sulfolobales*, whose sequence is rather divergent and lacks most of the RecJ canonical residues.

**Threading Algorithms Confirm the Presence of a RecJ/DHH-like Core Fold in hCdc45**—A variety of threading/fold recognition algorithms were used to verify the similarity among hCdc45 and DHH family members. The Protein Fold Recognition Server Phyre (33) identifies a three-dimensional similarity between hCdc45 and a number of DHH protein structures such as the RecJ exonuclease from *T. thermophilus* (PDB code: 1IR6), as well as various inorganic pyrophosphatases (PDB codes: 1WPN, 1I74, and 2HAW), but the similarity was restricted to the first 110 amino acids. When a multiple sequence alignment, including a number of Cdc45 sequences, was used as input for the HHpred server (34), a similarity (including the first 315 residues) was detected to an inorganic pyrophosphatase from *Bacillus subtilis* belonging to the DHH family (PDB code: 1WPN). The profile-profile alignment and fold-recognition server FFAS03 (35) was also used both with the full-length hCdc45 sequence and a number of fragments corresponding to putative domains. A score below the recommended threshold (−9) was obtained when the first 140 amino acid residues of hCdc45 were used as input, matching the N-terminal domain of a number of manganese-dependent inorganic pyrophosphatases from the DHH family, with the *B. subtilis* inorganic pyro-
phosphatase giving the best agreement (10.5, PDB code: 1K23). Using as input longer hCdc45 fragments still provides a match with the DHH proteins, although the scores get progressively higher, indicating a lower degree of confidence; using the full-length protein it is possible to detect some similarity to 1K23. The sequence-structure homology recognition server FUGUE (36) unambiguously identifies both the T. thermophilus RecJ (1IR6) as well as the Streptococcus gordonii inorganic pyrophosphatase (1K20) structures as the two best hits along the entire sequence with a degree of confidence higher than 99%. When using as input the sequences of a variety of archaeal RecJCdc45 proteins, all the servers predicted a strong structural similarity to the bacterial RecJ proteins, along the entire sequence, as expected from the significant sequence homology and the conservation of the characteristic motifs.

A central region in both Cdc45 (residues 252–363 in hCdc45) and archaeal RecCdc45 (residues 188–275 in the sequence from T. kodakaraensis) proteins, appears as a long insertion into the RecJ/DHH core. This region is reasonably well conserved between eukaryotic and archaeal proteins. Secondary structure predictions suggest a helical fold, and threading algorithms tend to find matches with helical bundle proteins, such as acyl-CoA-binding proteins and helix-turn-helix transcription factors. An additional insertion is unique to the eukaryotic Cdc45 proteins (residues 108–175 in the human sequence), and the central part includes a large number of charged residues (aspartate, glutamate, arginine, and lysine), thus suggesting a partially structured region (Fig. 1).

Based on the results of both the sequence analysis and the threading algorithms, we have produced a sequence alignment.
that summarizes the putative relationships among bacterial RecJ ssDNA exonucleases, archaeal RecJ\textsuperscript{Cdc45} and the eukaryotic Cdc45 (Fig. 1). In the first half of the molecule (up to motif IV) the similarity is strong enough to be detected based on sequence alone, whereas for the second half the alignment relies on the threading data, which identify similarity patterns in the absence of high sequence conservation. Consistent with the threading results, Fig. 1 shows the conservation of patterns of hydrophobic residues and an excellent match between the experimentally determined α-helices and β-strands of RecJ and the predicted Cdc45 secondary structure elements.

Although the correspondence is good overall, there are a few uncertainties in the central region of the proteins. For example it is difficult to establish unambiguously whether the first Cdc45 insertion occurs after β5 (as depicted in Fig. 1) or after α7. In the same way the second insertion (common to eukaryotic and archaeal proteins) may slide from the current position in Fig. 1 (after α10) to an alternative position after α11. However the match is very convincing in the first half of domain I as well as in domain II, including the long helix connecting the two domains (α14), suggesting that the relationship extends throughout the entire sequence.

While we were finalizing this report, a bioinformatic analysis (37) was published suggesting a similarity between Cdc45 and RecJ proteins, limited to the first 100 amino acid residues. Our report is consistent with that one (37) but further extends the analysis by showing that the sequence and structural similarity covers the entire length of the protein.

Biochemical Characterization of the Recombinant hCdc45—

We produced in bacterial cells hCdc45 as a fusion protein with a C-terminal His\textsubscript{6}-FLAG tag using TEV protease and purification over a heparin column (Fig. 2). Gel-filtration chromatography of purified hCdc45 was performed using a Bio-Sil SEC-250 column as described under “Experimental Procedures.” Peak fractions were analyzed by SDS-PAGE (5 μg/fraction) and used in a gel shift experiment (1 μl/fraction). EMSA with hCdc45-His-FLAG in the presence of anti-FLAG antibody. The assays were carried out by adding increasing amounts of a monoclonal anti-FLAG antibody (0.5, 1, and 2 μg, lanes 2 and 6, 3 and 7, and 4 and 8, respectively) into mixtures containing the single-stranded DNA probe with hCdc45-His-FLAG (lanes 6–8; 5 μg of protein) or without the recombinant protein (lanes 2–4; see text for details). A black arrow indicates the Cdc45-DNA complex, whereas the white arrows identify the ternary complexes with the anti-FLAG antibody.
Although many of the residues that in RecJ/DHH proteins are involved in Mn\(^{2+}\) binding and catalysis are not conserved in hCdc45, we examined the possibility that the few remaining aspartate and histidine residues (namely Asp-26, Asp-28, Asp-99, and His-101) may possibly coordinate one metal ion. We have therefore used both inductively coupled plasma/atomic emission spectroscopy and atomic absorption spectroscopy to test whether the purified protein contains manganese, magnesium, or zinc, but we failed to confirm the presence of any of these metals (data not shown).

We also carried out activity assays to check whether hCdc45 displays either pyrophosphatase or exonuclease activity. Consistently with the absence of metal ions and some of the putative catalytic residues, we were unable to detect any of the above enzymatic activities (data not shown).

Both its role in DNA replication and the putative similarity with a single-stranded DNA exonuclease, such as RecJ, suggested that Cdc45 could be a DNA-binding protein. We used EMSAs to evaluate the DNA-binding properties of hCdc45 (Fig. 2B). The purified recombinant protein binds single-stranded synthetic oligonucleotides with a weak but detectable affinity, comparable with the one observed for the *Drosophila* GINS complex (14) or for the full-length *T. thermophilus* RecJ (27). No increase in affinity was observed when a fork-containing DNA molecule was used as a ligand in the EMSAs, and negligible binding was detected to short blunt double-stranded DNA (Fig. 3C). Addition of Mg\(^{2+}\), Mn\(^{2+}\), or Zn\(^{2+}\) ions into the buffer, as well as variation of pH in the range 5.5–8.5, were found to have no effect on the hCdc45 DNA-binding activity (data not shown).

FIGURE 3. Small angle X-ray scattering data. A, the experimental SAXS profile (log intensity as a function of the momentum transfer) of hCdc45 (red points) is compared with the theoretical scattering curves calculated from the *ab initio* model (blue line). B, final model reconstructed from the scattering curve.
A number of RecJ-like proteins have been found associated with the DNA replication machinery in a variety of archaeal organisms, such as S. solfataricus (18), A. fulgidus (31), T. kodakaraensis (32). In particular, the T. kodakaraensis RecJCdc45 homologue has been shown to possess a 5′-3′ exonuclease activity that is highly stimulated by physical interaction with the GINS15 subunit (32). It has been proposed that this enzyme could participate in maturation of the Okazaki fragments in a pathway that is redundant with the FEN1- and Dna2-dependent pathways. Out of the multiple RecJ-like proteins present in archaeal organisms, it is plausible that some are associated with the DNA replication machinery and have a direct role in genome duplication; whereas others are “true” RecJ homologues, being involved in DNA repair/recombination transactions.

Based on our biochemical and bioinformatic analyses we suggest that Cdc45 originated from an ancestral 5′-3′ exonuclease loosing during evolution its catalytic activity and only retaining the ability to bind single-stranded DNA. The proteins from Sulfolobales may have followed a similar evolutionary path, as demonstrated by the absence of key catalytic residues and their tight association with GINS15. Association with the replication fork and GINS, in particular, may have led Cdc45 to loose its enzymatic activity and to acquire the function of “chaperoning” the lagging strand that is sterically excluded from the central channel of the MCM2–7 complex, as suggested from the cryoelectron microscopy of the CMG complex (38).

**Acknowledgments**—We thank Roberto Udisti (University of Florence) for determination of metal ions in protein samples by inductively coupled plasma-atomic emission spectroscopy, Spartaco De Gennaro and Achille Palma (University of Naples) for atomic absorption spectroscopy analysis. We are grateful to Alessandro Costa and James Berger (University of California at Berkeley) for sharing results before publication.

**REFERENCES**

1. Aparicio, O. M., Weinstein, D. M., and Bell, S. P. (1997) Components and dynamics of DNA replication complexes in S. cerevisiae. Redistribution of MCM proteins and Cdc45p during S phase. *Cell* 91, 59–69
2. Owens, J. C., Detweiler, C. S., and Li, J. J. (1997) CDC45 is required in conjunction with CDC7/DBF4 to trigger the initiation of DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* 94, 12521–12526
3. Uchiyama, M., Arai, K., and Masai, H. (2001) *Sna41goa1*, a novel mutation causing G1/S arrest in fission yeast, is defective in a CDC45 homolog and their tight association with GINS15. Association with the replication fork and GINS, in particular, may have led Cdc45 to lose its enzymatic activity and to acquire the function of “chaperoning” the lagging strand that is sterically excluded from the central channel of the MCM2–7 complex, as suggested from the cryoelectron microscopy of the CMG complex (38).

**FIGURE 4. SAXS data are consistent with a RecJ-like fold.** A, a schematic diagram summarizing the result of the bioinformatic analysis and showing the relationship between eukaryotic Cdc45 proteins, archaeal RecJCdc45, and bacterial RecJ single-stranded DNA exonucleases. The RecJ motifs are shown in red, the Cdc45 charged insertion is in magenta, and the helical archaeal/eukaryotic insertion is in yellow. The crystal structure of the core of *T. thermophilus* RecJ (PDB code: 2ZXP, domains I and II) is shown below the diagram, with domain I in blue, domain II in dark blue, and the connecting helix in green. The conserved residues in the seven RecJ motifs are shown in yellow. **B**, the ab initio calculated SAXS model for hCdc45 (depicted as gray light spheres) is superimposed to the crystal structure of the core of *T. thermophilus* RecJ, in blue. Highlighted in magenta and indicated by a magenta asterisk is the putative position of the insertion, which is unique to the eukaryotic Cdc45 orthologues; highlighted in yellow and indicated by a yellow asterisk is the position of the helical bundle insertion that is common to both archaeal and eukaryotic proteins (see Fig. 1). As an example, the helical domain of the acyl-CoA-binding protein (PDB code: 2FDQ) has been fitted to the map, consistently with the results of the threading algorithms. The two views are roughly related by a 90° rotation around a horizontal axis.

It has been long suggested that a pyrophosphatase activity may be associated to the replication machinery (40). Pyrophosphate hydrolysis could enhance the catalytic activity of replicative DNA polymerases, because removal of the pyrophosphate, a product of dNTPs incorporation, would drive forward the polymerization reaction.

However, an important clue about the function and the evolutionary origin of Cdc45 comes from the finding that some of its putative archaeal homologues display a much closer similarity to the DHH subfamily 1 that includes the RecJ single-stranded DNA exonucleases. Moreover, the ab initio model obtained from SAXS data is more consistent with a compact RecJ core, rather than the more “open” conformation of inorganic pyrophosphatases. A number of RecJ-like proteins have
(2007) Human Cdc45 is a proliferation-associated antigen. FEBs J. 274, 3669–3684.
9. Pospiech, H., Grosse, F., and Pisani, F. M. (2010) The initiation step of eukaryotic DNA replication. Subcell. Biochem. 50, 79–104.
10. Zou, L., and Stillman, B. (2000) Assembly of a complex containing Cdc45p, replication protein A, and Mcm2p at replication origins controlled by S-phase cyclin-dependent kinases and Cdc7p-Dbp4p kinase. Mol. Cell Biol. 20, 3086–3096.
11. Gambus, A., Jones, R. C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R. D., and Labib, K. (2006) GINS maintains association of Cdc45 with MCM in repressile progression complexes at eukaryotic DNA replication forks. Nat. Cell Biol. 8, 358–366.
12. Pacey, M., Tutter, A. V., Kubota, Y., Takisawa, H., and Walter, J. C. (2006) Localization of MCM2–7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. Mol. Cell 21, 581–587.
13. Moyer, S. E., Lewis, P. W., and Botchan, M. R. (2006) Isolation of the Cdc45/Mcm2–7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. Proc. Natl. Acad. Sci. U.S.A. 103, 10236–10241.
14. Ilves, I., Petojevic, T., Pesavento, J. J., and Botchan, M. R. (2010) Activation of the MCM2–7 helicase by association with Cdc45 and GINS proteins. Mol. Cell 37, 247–258.
15. Sakakibara, N., Kelman, L. M., and Kelman, Z. (2009) Unwinding the structure and function of the archaeal MCM helicase. Mol. Microbiol. 72, 286–296.
16. Costa, A., and Onesti, S. (2009) Structural biology of MCM helicases. Crit. Rev. Biochem. Mol. Biol. 44, 326–342.
17. Makarova, K. S., Wolf, Y. I., Mekhedov, S. L., Mirkin, B. G., and Koonin, E. V. (2005) Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. Nucleic Acids Res. 33, 4626–4638.
18. Marinsek, N., Barry, E. R., Makarova, K. S., Dionne, L., Koonin, E. V., and Bell, S. D. (2006) GINS, a central nexus in the archaeal DNA replication fork. EMBO rep. 7, 539–545.
19. Yoshimochi, T., Fujikane, R., Kawanami, M., Matsunaga, F., and Ishino, Y. (2008) The GINS complex from Pyrococcus furiosus stimulates the MCM helicase activity. J. Biol. Chem. 283, 1601–1609.
20. Savitsky, P., Bray, J., Cooper, C. D., Marsden, B. D., Mahajan, P., Burgess-Brown, N. A., and Gileadi, O. (2010) High-throughput production of human proteins for crystallization. The SGC experience. J. Struct. Biol. 172, 3–13.
21. Amenitsch, H., Bernstorf, S., Kriebchaum, M., Lombardo, D., Mio, H., Rappolt, M., and Laggner, P. (1997) J. Appl. Crystallogr. 30, 872–876.
22. Konarev, P. V., Petoukhov, M. V., Volkov, V. V., and Svergun, D. I. (2006) J. Appl. Crystallogr. 39, 277–286.
23. Svergun, D. I. (1992) J. Appl. Crystallogr. 25, 495–503.
24. Franke, D., and Svergun, D. I. (2009) J. Appl. Crystallogr. 42, 342–346.
25. Volkov, V. V., and Svergun, D. I. (2003) J. Appl. Crystallogr. 36, 860–864.
26. Aravind, L., and Koonin, E. V. (1998) A novel family of predicted phosphoesterases includes Drosophila prune protein and bacterial Rec exonuclease. Trends Biochem. Sci. 23, 17–19.
27. Yamagata, A., Kakuta, Y., Masui, R., and Fukuyama, K. (2002) The crystal structure of exonuclease RecB bound to Ms2+ ion suggests how its characteristic motifs are involved in exonuclease activity. Proc. Natl. Acad. Sci. U.S.A. 99, 5908–5912.
28. Wakamatsu, T., Kitamura, Y., Kotera, Y., Nakagawa, N., Kuramitsu, S., and Masui, R. (2010) Structure of RecB exonuclease defines its specificity for single-stranded DNA. J. Biol. Chem. 285, 9762–9769.
29. Fabrichnyi, I. P., Lehtiő, L., Salminen, A., Zyryanov, A. B., Baykov, A. A., Lahti, R., and Goldman, A. (2004) Structural studies of metal ions in family II pyrophosphatases. The requirement for a Janus ion. Biochemistry 43, 14403–14411.
30. Rajman, L. A., and Lovett, S. T. (2000) A thermostable single-strand DNase from Methanococcus jannaschii related to the RecA recombination and repair exonuclease from Escherichia coli. J. Bact. 182, 607–612.
31. Motz, M., Kober, I., Girardot, C., Loeser, E., Bauer, U., Albers, M., Moeckel, G., Minch, E., Voss, H., Kilger, C., and Koegl, M. (2002) Elucidation of an archaeal replication protein network to generate enhanced PCR enzymes. J. Biol. Chem. 277, 16179–16188.
32. Li, Z., Pan, M., Santangelo, T. J., Chemnitz, W., Yuan, W., Edwards, J. L., Hurwitz J., Reeve, J. N., and Kelman, Z. (2011) A novel DNA nuclease is stimulated by association with the GINS complex. Nucleic Acids Res. 39, 6114–6123.
33. Kelley, L. A., and Sternberg, M. J. (2009) Protein structure prediction on the Web. A case study using the Phyre server. Nature Protocols 4, 363–371.
34. Söding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33, W244–W248.
35. Jaroszewski, L., Rychlewski, L., Li, Z., Li, W., and Godzik, A. (2005) FFAS03. A server for profile–profile sequence alignments. Nucleic Acids Res. 33, W284–W288.
36. Shi, J., Blundell, T. L., and Mizuguchi, K. (2001) FUGUE. Sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. J. Mol. Biol. 310, 243–257.
37. Sanchez-Pulido, L., and Ponting, C. P. (2011) Cdc45. The missing RecJ ortholog in eukaryotes? Bioinformatics 27, 1885–1888.
38. Costa, A., Ilves, I., Tamberg, N., Petoejovic, T., Nogales, E., Botchan, M. R., and Berger, J. M. (2011) The structural basis for MCM2–7 helicase activation by GINS and Cdc45. Nat. Struct. Mol. Biol. 18, 471–477.
39. Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys. J. 76, 2879–2886.
40. Aravind, L., and Koonin, E. V. (1998) Phosphoesterase domains associated with DNA polymerases of diverse origins. Nucleic Acids Res. 26, 3746–3752.
41. Edgar, R. C. (2004) MUSCLE. Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.