

Cdc45: the missing RecJ ortholog in eukaryotes?

Luis Sanchez-Pulido* and Chris P. Ponting

Department of Physiology, Anatomy and Genetics, MRC Functional Genomics Unit, University of Oxford, Oxford OX1 3QX, UK

Associate Editor: Burkhard Rost

ABSTRACT

Summary: DNA replication is one of the most ancient of cellular processes and functional similarities among its molecular machinery are apparent across all cellular life. Cdc45 is one of the essential components of the eukaryotic replication fork and is required for the initiation and elongation of DNA replication, but its molecular function is currently unknown. In order to trace its evolutionary history and to identify functional domains, we embarked on a computational sequence analysis of the Cdc45 protein family. Our findings reveal eukaryotic Cdc45 and prokaryotic RecJ to possess a common ancestry and Cdc45 to contain a catalytic site within a predicted exonuclease domain. The likely orthology between Cdc45 and RecJ reveals new lines of enquiry into DNA replication mechanisms in eukaryotes.

Contact: luis.sanchezpulido@dpag.ox.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on March 29, 2011; revised on May 11, 2011; accepted on May 27, 2011

1 INTRODUCTION

Archaea and eukarya possess DNA replication machineries that are evolutionarily related, indicative of a common ancestry prior to the divergence of these two ancient lineages (Leipe *et al.*, 1999). In particular, progression at eukaryotic replication forks requires a large multimolecular complex (the ‘replisome progression complex’) consisting of Cdc45, the GINS (go-ichi-ni-san) complex (Sld5, Psf1-3) and MCM2-7 (Aparicio *et al.*, 1997; Gambus *et al.*, 2006). Archaea possess homologs of only GINS subunits and MCM2-7, the macromolecular machine presumed to unwind DNA during replication (Marinsek *et al.*, 2006). In eukaryotes, the third component of this complex, Cdc45, is essential, ubiquitous and critical for the initiation and elongation of DNA replication (Aparicio *et al.*, 1997; Costa *et al.*, 2011; Diffley *et al.*, 1998; Mimura and Takisawa, 1998; Takisawa *et al.*, 2000; Tercero *et al.*, 2000).

Despite this fundamental contribution to eukaryotic DNA replication, orthologues of Cdc45 are not apparent among either archaea or bacteria. Is Cdc45, therefore, an innovation of early eukaryotic life, one that has since become essential in eukaryotic species from across this diverse kingdom (Leipe *et al.*, 1999)? Such innovations are very common. There is, for example, no known DNA repair protein possessing the same arrangement of protein domains across bacteria, archaea and eukaryotes, and only one repair protein, the 5′–3′ exonuclease RecJ (Lovett and Kolodner, 1989),

which is conserved in most bacteria and archaea, but that is absent from eukarya (Aravind *et al.*, 1999; Rajman and Lovett, 2000). For most prokaryotic species, RecJ is the only 5′–3′ exonuclease that is specific for single-stranded DNA (ssDNA). It is known to be involved in homologous recombination, base excision repair, mismatch repair and, of particular importance here, in the rescue of stalled replication forks (Chow and Courcelle, 2007; Courcelle and Hanawalt, 1999; Courcelle *et al.*, 2003; Han *et al.*, 2006; Lovett and Kolodner, 1989; Sutura *et al.*, 1999; Wakamatsu *et al.*, 2010; Yamagata *et al.*, 2002).

In order to better understand the molecular function of Cdc45, which remains far from clear, and to trace its evolutionary ancestry, we embarked on an extensive analysis of Cdc45 sequences.

2 RESULTS AND DISCUSSION

2.1 Computational protein sequence analysis

Our exhaustive database searches for Cdc45 homologs using BLAST (Altschul *et al.*, 1997) revealed, as expected, single copies in all eukaryotes, from animals to land plants and fungi. We then employed a protein domain hunting strategy based on profile-to-sequence comparisons using HMMer2 (Eddy, 1996) against the UniRef50 protein sequence database (Wu *et al.*, 2006). As queries, we used sequence regions conserved within the Cdc45 multiple sequence alignment generated with T-Coffee (Notredame *et al.*, 2000).

Surprisingly, this approach yielded statistically significant sequence similarity of Cdc45 orthologues with the RecJ family of DHH phosphoesterases (Aravind and Koonin, 1998). For example, a global profile sequence search with the N-terminal conserved region of the Cdc45 family (corresponding to amino acids 16–104 of UniProt: O75419_HUMAN) as query (represented by a green oval in Fig. 1A), identified the first RecJ family member in an archaeal sequence (amino acids 23–107 of UniProt: BIL6T5_KORCO from *Korarchaeum cryptofilum*) with a significant *E*-value of 0.002 (Fig. 1A–C). Reciprocally, the profile of the DHH RecJ N-terminal domain yielded a significant sequence similarity with a green algal Cdc45 protein (amino acids 24–115 of UniProt: A8HW43_CHLRE from *Chlamydomonas reinhardtii*; *E*-value = 0.1; Fig. 1C).

Additionally, predicted secondary structures (Jones, 1999) of the N-terminal region of the Cdc45 family are consistent with secondary structures from the known crystal structure of RecJ (Yamagata *et al.*, 2002) (Fig. 1E). We were unable to identify statistically significant sequence similarity in the evolutionarily conserved C-termini of the Cdc45 family to other protein sequences (represented by a red box in Fig. 1A; see Supplementary Figure).

*To whom correspondence should be addressed.

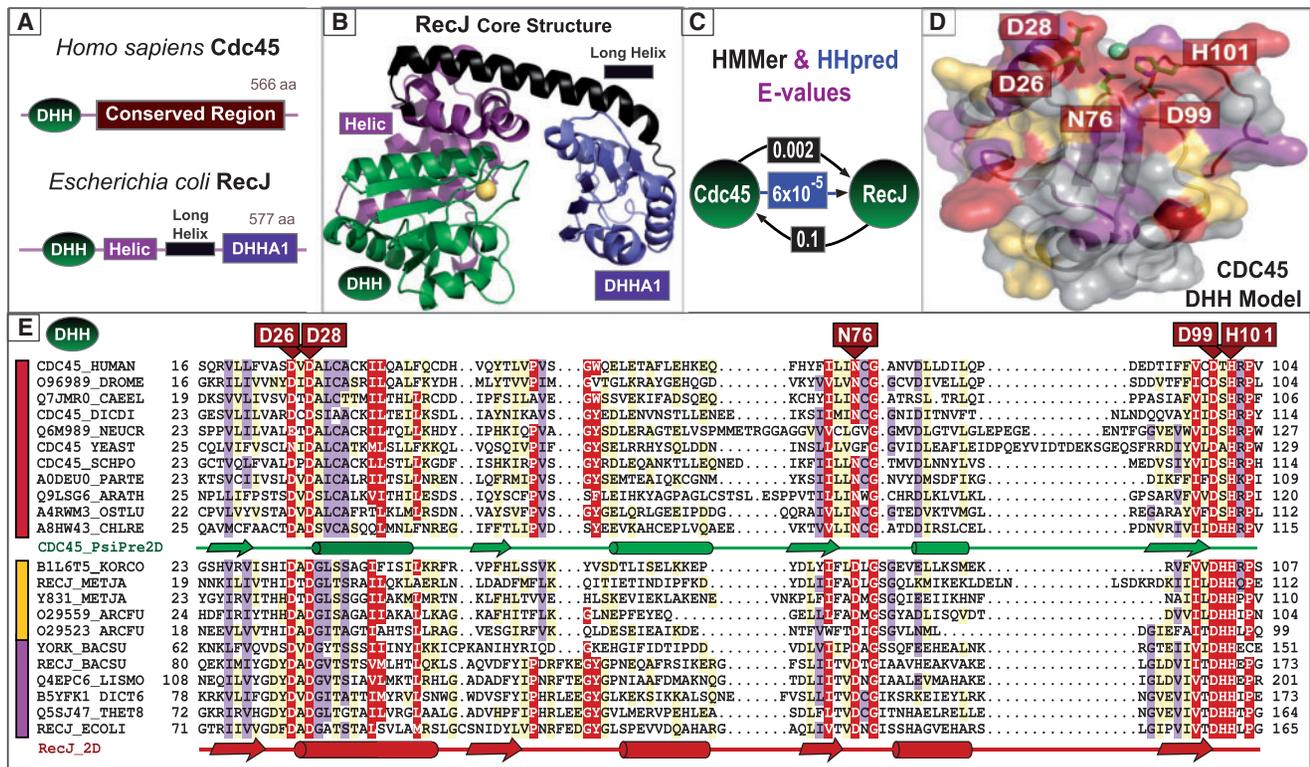


Fig. 1. Sequence analysis of the Cdc45 and RecJ protein families. (A) Domain architecture of human Cdc45 and *Escherichia coli* RecJ proteins. For the *E. coli* RecJ protein, domains were assigned according to the RecJ core structure (Yamagata *et al.*, 2002) and the Pfam domain database (Finn *et al.*, 2008). Proteins are drawn approximately to scale. (B) Ribbon diagram of the *Thermus thermophilus* RecJ core structure (PDB-ID: 1IR6) (Yamagata *et al.*, 2002) colored by its domain architecture: DHH domain (in green), helical domain (in purple), long connecting alpha helix (in black) and DHHA1 domain (in violet). The ribbon diagram was drawn using the Pymol protein structure visualization program (<http://www.pymol.org/>). (C) HMMer and HHpred comparison *E*-values between Cdc45 and RecJ families. Numbers correspond to HMMer profile versus sequence (gray boxes) and HHpred profile versus profile (blue box) comparison *E*-values from global profile search results (Eddy, 1996; Soding *et al.*, 2005). Arrows indicate the profile search direction. (D) Homology model of the DHH domain of human Cdc45. The surface coloring scheme indicates average BLOSUM62 scores (which are correlated with amino acid conservation) identical to those used for the alignment. Highly conserved residues that form part of its predicted active center are labeled and side chains shown. A manganese ion (from PDB-ID: 1IR6) is represented by the green sphere. The human Cdc45 DHH domain structural model was created using Modeller (Sali and Blundell, 1993) based on a RecJ structure (Yamagata *et al.*, 2002) and presented using Pymol. (E) Representative multiple sequence alignment of the DHH domain from Cdc45 and RecJ families. The alignment was produced using a combination of T-Coffee (Notredame *et al.*, 2000) and the profile-to-profile alignment program HHalign (Soding *et al.*, 2005). The limits of the protein sequences included in the alignment are indicated by the residue positions provided at each side. It was represented with the program Belvu (Sonnhammer and Hollich, 2005) using a coloring scheme indicating the average BLOSUM62 scores (which are correlated with amino acid conservation) of each alignment column: red (>1.55), violet (between 1.55 and 0.8) and light yellow (between 0.8 and 0.3). Residues predicted to form part of the active center of human Cdc45 DHH domain are indicated by red boxes above the alignment. Phyletic ranges of Cdc45 and RecJ families are indicated by colored bars to the left of the alignment: red (eukarya), yellow (archaea) and purple (bacteria). PsiPred secondary structure predictions (Jones, 1999) for the Cdc45 family are shown (in green) below the Cdc45 family alignment. Secondary structures from *T. thermophilus* RecJ (PDB-ID: 1IR6) are shown (in red) below the alignment (Yamagata *et al.*, 2002). Alpha helices and beta strands are indicated by cylinders and arrows, respectively. Sequences are named with their Uniprot identifications (Wu *et al.*, 2006). Species abbreviations: HUMAN, *Homo sapiens*; DROME, *Drosophila melanogaster*; CAEEL, *Caenorhabditis elegans*; DICDI, *Dictyostelium discoideum*; NEUCR, *Neurospora crassa*; YEAST, *Saccharomyces cerevisiae*; SCHPO, *Schizosaccharomyces pombe*; PARTE, *Paramecium tetraurelia*; ARATH, *Arabidopsis thaliana*; OSTLU, *Ostreococcus lucimarinus*; CHLRE, *Chlamydomonas reinhardtii*; KORCO, *Korarchaeum cryptofilum*; METJA, *Methanocaldococcus jannaschii*; ARCFU, *Archaeoglobus fulgidus*; BACSU, *Bacillus subtilis*; LISMO, *Listeria monocytogenes*; DICT6, *Dictyoglossus thermophilus*; THET8, *Thermus thermophilus*; and, ECOLI, *Escherichia coli*.

To determine whether fold recognition analysis would generate supporting results, we submitted the Cdc45 N-terminal region as query to HHpred, a tool based on profile–profile comparisons (Soding *et al.*, 2005). The profile generated for the Cdc45 N-terminal region matched the N-terminal DHH domain of *Thermus thermophilus* RecJ protein (PDB-IDs: 2ZXR & 1IR6) (Wakamatsu

et al., 2010; Yamagata *et al.*, 2002) with a highly significant *E*-value of 6×10^{-5} (estimated error rate < 3%) despite a low level of sequence identity (< 20%) to the human Cdc45 protein (Fig. 1C). Moreover, in support of the first match, the next most statistically significant (from an *E*-value = 4×10^{-4}) matches are to seven more distant members of the DHH superfamily of phosphoesterases

(PDB-IDs: 3DEV, 3DMA, 1WPN, 1K20, 2HAW, 2QB7 and 2EB0) (Aravind and Koonin, 1998; Fabrichny *et al.*, 2007; Merckel *et al.*, 2001; Ugochukwu *et al.*, 2007).

The significant *E*-values of HMMer searches, the consistency of secondary structure predictions and corroboration by profile-to-profile comparison and fold assignment (HHpred) methods, taken together provide strong evidence that Cdc45 and RecJ families share a homologous DHH domain located at their N-termini. The inclusion within the Cdc45 alignment of distantly related sequences from recently sequenced unicellular eukaryotes, such as algae, and the use of profile-to-profile comparison tools that allow more sensitive sequence searches, likely explain why this remote homologous relationship between Cdc45 and RecJ families had hitherto escaped detection.

2.2 Consideration of whether Cdc45 originated from an ancestral RecJ molecule

In HMMer and HHpred searches, Cdc45 sequences were found to be most similar to RecJ sequences than to any other DHH family protein. Closer similarity between Cdc45 and RecJ than to any other DHH proteins implies that these proteins may be orthologous, having been derived from a single homolog in the last common ancestor of all cellular life. Evidence for or against this proposal was not forthcoming from phylogenetic analysis of DHH sequences. This was because accurate alignment is only possible for these sequences over motifs I, II and III (Fig. 1E; Aravind and Koonin, 1998) and this provides insufficient information to generate either a stable phylogeny or robust bootstrap values. Nevertheless, it is notable that the phyletic distribution of RecJ among bacteria and archaea (Aravind *et al.*, 1999), to the exclusion of eukarya, is complementary with the eukarya-specific phyletic distribution of what we now report as its eukaryotic homolog, Cdc45 (see species distribution of Pfam family CDC45 accession: PF02724) (Finn *et al.*, 2008). We thus suggest that Cdc45 and RecJ are orthologues, having been derived from a single homolog in the last common ancestor of all cellular life. Orthology is also consistent with the observation that to date, other than Cdc45, no eukaryotic member of the DHH superfamily has been implicated in DNA repair or replication.

2.3 Structure and function of RecJ and Cdc45

The RecJ core structure contains two globular regions connected by a long alpha helix (Fig. 1B). Homology between Cdc45 and RecJ that we describe above corresponds to the N-terminal DHH domain which is named after three conserved residues (Asp, His, His) that contribute to catalysis and/or to binding to bivalent cations, such as manganese Mn^{2+} or Mg^{2+} . The Cdc45 DHH domain contains three (I to III) of the four conserved motifs that form part of the divalent cation coordination sphere of their phosphoesterase catalytic center (Aravind and Koonin, 1998; Fabrichny *et al.*, 2007; Halonen *et al.*, 2005; Merckel *et al.*, 2001; Sutura *et al.*, 1999; Tammenkoski *et al.*, 2007; Ugochukwu *et al.*, 2007; Wakamatsu *et al.*, 2010; Yamagata *et al.*, 2002). Aside from two aspartic acids of motif I, broadly conserved within the Cdc45 family (D26 and D28, see Fig. 1D and E), a common element to all these structures is an aspartic acid–histidine pair (D99 and H101, Fig. 1D and E), equivalent to flanking residues of the DHH triplet motif which lends the enzyme family its name (motif III). This triplet motif is also present in Cdc45 from the green algae *Chlamydomonas*

reinhardtii (Fig. 1E) and *Volvox carteri*. The aspartic acid–histidine pair is highly conserved in Cdc45, and in the DHH family has been postulated to act as a general acid, to protonate the oxygen bridging the phosphate groups thereby facilitating its hydrolysis (Ugochukwu *et al.*, 2007). However, the Cdc45 family exhibits a substitution of an aspartic acid, which is otherwise well conserved in the DHH family, with a non-conserved asparagine (N76). The aspartic acid, together with the central histidine of the DHH motif (motif III), form part of the conserved divalent cation coordination sphere identified in the known structures of the DHH superfamily. On one hand, the apparently incomplete phosphoesterase active center of Cdc45 might reflect substantial differences in the catalytic activities of RecJ and Cdc45. On the other, the Cdc45 active center may have further contributions from amino acids from within its apparently RecJ-dissimilar yet conserved C-terminal region, or even contributions from its interaction partners, in an analogous manner to arginine residues of Ras GTPase activator proteins that stabilize the transition state of the GTPase reaction (Ahmadian *et al.*, 1997).

Homology between RecJ and Cdc45 suggests that these molecules share molecular and/or cellular functions. Cdc45 is enriched at stalled replication forks (Pacek *et al.*, 2006) and has been described as a checkpoint of the DNA damage-dependent response with potential roles in genome stability (Broderick and Nasheuer, 2009; Liu *et al.*, 2006). Due to Cdc45 being an indispensable component of eukaryotic DNA replication forks and owing to its closer evolutionary relationship to RecJ than to any other DHH phosphoesterase, eukaryotic Cdc45 may, like prokaryotic RecJ, be involved in the rescue of stalled replication fork structures (Chow and Courcelle, 2007; Courcelle and Hanawalt, 1999; Courcelle *et al.*, 2003).

Others have proposed that the 5′–3′ exonuclease EXO1-mediated processing of stalled replication forks in eukaryotes (Cotta-Ramusino *et al.*, 2005; Qiu *et al.*, 1998) resembles the action of bacterial RecJ (Courcelle *et al.*, 2003; Han *et al.*, 2006). However, genetic evidence indicates that additional proteins possess an EXO1-like exonuclease activity involved in the rescue of stalled replication forks (Alam *et al.*, 2003).

Cdc45 may thus have retained only a subset of RecJ's diverse cellular roles. It may also be possible that the DHH domain of Cdc45 acquired a novel exonuclease activity that facilitates the initiation and elongation of the eukaryotic DNA replication fork. Nevertheless, preliminary studies have not revealed it to possess such an activity (data not shown).

Alternatively, another functional scenario for Cdc45 is suggested by a recent experimental finding that bacterial RecJ-like proteins are involved in the degradation of oligonucleotides, suggesting a potential role for both RecJ and Cdc45 in preventing genomic alterations during DNA replication (Bryan and Swanson, 2011; Wakamatsu *et al.*, 2011).

3 CONCLUSION

We have identified a statistically significant sequence similarity between the N-terminal conserved DHH domains of Cdc45 and RecJ families. We postulate that Cdc45 possesses a divalent cation (Mg^{2+} or Mn^{2+}) dependent 5′–3′ exonuclease activity important for DNA replication. Experimental approaches are now required to investigate these evolutionary, molecular and cellular hypotheses.

ACKNOWLEDGEMENTS

We are grateful to John Rouse (University of Dundee) for his insightful comments on Cdc45.

Funding: L.S.-P. and C.P.P. are supported by the Medical Research Council UK.

Conflict of Interest: none declared.

REFERENCES

- Ahmadian, M.R. *et al.* (1997) Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras. *Nat. Struct. Biol.*, **4**, 686–689.
- Alam, N.A. *et al.* (2003) Germline deletions of EXO1 do not cause colorectal tumors and lesions which are null for EXO1 do not have microsatellite instability. *Cancer Genet. Cytogenet.*, **147**, 121–127.
- Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Aparicio, O.M. *et al.* (1997) Components and dynamics of DNA replication complexes in *S. cerevisiae*: redistribution of MCM proteins and Cdc45p during S phase. *Cell*, **91**, 59–69.
- Aravind, L. and Koonin, E.V. (1998) A novel family of predicted phosphoesterases includes *Drosophila* prune protein and bacterial RecJ exonuclease. *Trends Biochem. Sci.*, **23**, 17–19.
- Aravind, L. *et al.* (1999) Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res.*, **27**, 1223–1242.
- Broderick, R. and Nasheuer, H.P. (2009) Regulation of Cdc45 in the cell cycle and after DNA damage. *Biochem. Soc. Trans.*, **37**, 926–930.
- Bryan, A. and Swanson, M.S. (2011) Oligonucleotides stimulate genomic alterations of *Legionella pneumophila*. *Mol. Microbiol.*, **80**, 231–247.
- Chow, K.H. and Courcelle, J. (2007) RecBCD and RecJ/RecQ initiate DNA degradation on distinct substrates in UV-irradiated *Escherichia coli*. *Radiat. Res.*, **168**, 499–506.
- Costa, A. *et al.* (2011) The structural basis for MCM2-7 helicase activation by GINS and Cdc45. *Nat. Struct. Mol. Biol.*, **18**, 471–477.
- Cotta-Ramusino, C. *et al.* (2005) Exo1 processes stalled replication forks and counteracts fork reversal in checkpoint-defective cells. *Mol. Cell*, **17**, 153–159.
- Courcelle, J. and Hanawalt, P.C. (1999) RecQ and RecJ process blocked replication forks prior to the resumption of replication in UV-irradiated *Escherichia coli*. *Mol. Gen. Genet.*, **262**, 543–551.
- Courcelle, J. *et al.* (2003) DNA damage-induced replication fork regression and processing in *Escherichia coli*. *Science*, **299**, 1064–1067.
- Diffley, J.F. (1998) Replication control: choreographing replication origins. *Curr. Biol.*, **8**, R771–R773.
- Eddy, S.R. (1996) Hidden Markov models. *Curr. Opin. Struct. Biol.*, **6**, 361–365.
- Fabrichny, I.P. *et al.* (2007) A trimetal site and substrate distortion in a family II inorganic pyrophosphatase. *J. Biol. Chem.*, **282**, 1422–1431.
- Finn, R.D. *et al.* (2008) The Pfam protein families database. *Nucleic Acids Res.*, **36**, D281–D288.
- Gambus, A. *et al.* (2006) GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat. Cell Biol.*, **8**, 358–366.
- Halonon, P. *et al.* (2005) Effects of active site mutations on the metal binding affinity, catalytic competence, and stability of the family II pyrophosphatase from *Bacillus subtilis*. *Biochemistry*, **44**, 4004–4010.
- Han, E.S. *et al.* (2006) RecJ exonuclease: substrates, products and interaction with SSB. *Nucleic Acids Res.*, **34**, 1084–1091.
- Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.
- Leipe, D.D. *et al.* (1999) Did DNA replication evolve twice independently? *Nucleic Acids Res.*, **27**, 3389–3401.
- Liu, P. *et al.* (2006) The Chk1-mediated S-phase checkpoint targets initiation factor Cdc45 via a Cdc25A/Cdk2-independent mechanism. *J. Biol. Chem.*, **281**, 30631–30644.
- Lovett, S.T. and Kolodner, R.D. (1989) Identification and purification of a single-stranded-DNA-specific exonuclease encoded by the recJ gene of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **86**, 2627–2631.
- Marinsek, N. *et al.* (2006) GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep.*, **7**, 539–545.
- Merckel, M.C. *et al.* (2001) Crystal structure of *Streptococcus mutans* pyrophosphatase: a new fold for an old mechanism. *Structure*, **9**, 289–297.
- Mimura, S. and Takisawa, H. (1998) *Xenopus* Cdc45-dependent loading of DNA polymerase alpha onto chromatin under the control of S-phase Cdk. *EMBO J.*, **17**, 5699–5707.
- Notredame, C. *et al.* (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, **302**, 205–217.
- Pacek, M. *et al.* (2006) Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. *Mol. Cell*, **21**, 581–587.
- Qiu, J. *et al.* (1998) *Saccharomyces cerevisiae* exonuclease-1 plays a role in UV resistance that is distinct from nucleotide excision repair. *Nucleic Acids Res.*, **26**, 3077–3083.
- Rajman, L.A. and Lovett, S.T. (2000) A thermostable single-strand DNase from *Methanococcus jannaschii* related to the RecJ recombination and repair exonuclease from *Escherichia coli*. *J. Bacteriol.*, **182**, 607–612.
- Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.
- Söding, J. *et al.* (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.*, **33**, W244–W248.
- Sonnhammer, E.L. and Hollich, V. (2005) Scoredist: a simple and robust protein sequence distance estimator. *BMC Bioinformatics*, **6**, 108.
- Sutera, V.A. Jr *et al.* (1999) Mutational analysis of the RecJ exonuclease of *Escherichia coli*: identification of phosphoesterase motifs. *J. Bacteriol.*, **181**, 6098–6102.
- Takisawa, H. *et al.* (2000) Eukaryotic DNA replication: from pre-replication complex to initiation complex. *Curr. Opin. Cell Biol.*, **12**, 690–696.
- Tammenkoski, M. *et al.* (2007) Kinetic and mutational analyses of the major cytosolic exopolyphosphatase from *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **282**, 9302–9311.
- Tercero, J.A. *et al.* (2000) DNA synthesis at individual replication forks requires the essential initiation factor Cdc45p. *EMBO J.*, **19**, 2082–2093.
- Ugochukwu, E. *et al.* (2007) The crystal structure of the cytosolic exopolyphosphatase from *Saccharomyces cerevisiae* reveals the basis for substrate specificity. *J. Mol. Biol.*, **371**, 1007–1021.
- Wakamatsu, T. *et al.* (2010) Structure of RecJ exonuclease defines its specificity for single-stranded DNA. *J. Biol. Chem.*, **285**, 9762–9769.
- Wakamatsu, T. *et al.* (2011) Role of RecJ-like protein with 5′-3′ exonuclease activity in oligo(deoxy)nucleotide degradation. *J. Biol. Chem.*, **286**, 2807–2816.
- Wu, C.H. *et al.* (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.*, **34**, D187–D191.
- Yamagata, A. *et al.* (2002) The crystal structure of exonuclease RecJ bound to Mn²⁺ ion suggests how its characteristic motifs are involved in exonuclease activity. *Proc. Natl Acad. Sci. USA*, **99**, 5908–5912.