6042–6051 *Nucleic Acids Research, 2007, Vol. 35, No. 18* *Published online 30 August 2007*
doi:10.1093/nar/gkm554

An archaeal orthologue of the universal protein Kae1 is an iron metalloprotein which exhibits atypical DNA-binding properties and apurinic-endonuclease activity *in vitro*

Arnaud Hecker[1], Nicolas Leulliot[2], Danièle Gadelle[1], Marc Graille[2], Anthony Justome[3], Pierre Dorlet[4], Céline Brochier[5,6], Sophie Quevillon-Cheruel[2], Eric Le Cam[3], Herman van Tilbeurgh[2] and Patrick Forterre[1,7,*]

[1]Institut de Génétique et Microbiologie, Univ. Paris-Sud, IFR115, UMR8621-CNRS, 91405 Orsay, France, [2]Institut de Biochimie et de Biophysique Moléculaire et Cellulaire, Univ. Paris-Sud, IFR115, UMR8619-CNRS, 91405 Orsay, France, [3]Institut Gustave Roussy, Interactions Moléculaires et Cancer, UMR8126-CNRS, 94805 Villejuif Cedex, France, [4]Institut de Chimie Moléculaire et des Matériaux, Univ. Paris-Sud, UMR8182-CNRS, 91405 Orsay, France, [5]Institut de Biologie Structurale et de Microbiologie, Laboratoire de Chimie Bactérienne, UPR9043-CNRS, 13402 Marseille Cedex 20, France, [6]Université de Provence - Aix-Marseille I, 13331 Marseille Cedex 3, France and [7]Institut Pasteur, Unité Biologie Moléculaire du Gène chez les Extrêmophiles, 25 rue du Dr Roux, 75724 Paris Cedex 15, France

Received April 25, 2007; Revised July 3, 2007; Accepted July 5, 2007

ABSTRACT

The Kae1 (Kinase-associated endopeptidase 1) protein is a member of the recently identified transcription complex EKC and telomeres maintenance complex KEOPS in yeast. Kae1 homologues are encoded by all sequenced genomes in the three domains of life. Although annotated as putative endopeptidases, the actual functions of these universal proteins are unknown. Here we show that the purified Kae1 protein (Pa-Kae1) from *Pyrococcus abyssi* is an iron-protein with a novel type of ATP-binding site. Surprisingly, this protein did not exhibit endopeptidase activity *in vitro* but binds cooperatively to single and double-stranded DNA and induces unusual DNA conformational change. Furthermore, Pa-Kae1 exhibits a class I apurinic (AP)-endonuclease activity (AP-lyase). Both DNA binding and AP-endonuclease activity are inhibited by ATP. Kae1 is thus a novel and atypical universal DNA interacting protein whose importance could rival those of RecA (RadA/Rad51) in the maintenance of genome integrity in all living cells.

INTRODUCTION

Recently, two groups have independently discovered in *Saccharomyces cerevisiae* two closely related protein complexes called either KEOPS (for Kinase, Endopeptidase and Other Proteins of Small size) or EKC (for Endopeptidase-like and Kinase associated to transcribed Chromatin). The KEOPS complex is involved in telomere uncapping and elongation (1) whereas the EKC complex is involved in transcription of essential eukaryotic genes (2). Among the proteins composing this complex, Bud32 and Kae1, are highly conserved in the living world. Bud32 is a Ser/Thr protein kinase present in Archaea and Eukaryotes, whereas Kae1 (for Kinase-associated endopeptidase 1) is present in the three domains of life. The association between Kae1 and Bud32 appears to be also highly conserved since their genes are fused in several archaeal genomes (3). Interestingly, the human orthologue of Bud32 is PRPK (p53-Related Protein Kinase) (4,5), suggesting that the human orthologue of Kae1 (called OSGEP for O-SyaloGlycoprotein EndoPeptidase) participates to the p53 regulatory network via functional interaction with PRPK.

The Kae1 protein and its homologues belong to the small set of about 60 universal proteins present in all members of the three domains of life (6). This protein was

*To whom correspondence should be addressed. Tel: +33 1 69 15 74 89; Fax: +33 1 69 15 78 08; Email: patrick.forterre@igmors.u-psud.fr
Correspondence may also be addressed to Herman van Tilbeurgh. Email: herman.van-tilbeurgh@u-psud.fr

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

placed in 2004 by Galperin and Koonin at the top of their list of 10 'known-unknown' proteins 'that should be priority targets for experimental study' (7). Indeed, this putative endopeptidase (of 'known' biochemical function) was the only protein of 'unknown' biological role present in all 70 genomes then available. The endopeptidase activity of Kae1/OSGEP proteins was initially derived from a study of Mellors and colleagues (8,9) in 1991, who reported that an O-syaloglycoprotein endopeptidase previously purified from *Pasteurella haemolytica* was encoded by a bacterial *kae1* homologue gene. Later on, Koonin and co-workers (10) identified in 1999 an Hsc70-actin-like fold (HALF) in Kae1-related proteins and have suggested that they were ATP-dependent proteases with chaperone activity. Therefore, it was suggested that yeast Kae1 could modulate the activity of the KEOPS/EKC complexes via its proteolytic activity (1). Recently, the structure of the protein YeaZ from *Salmonella typhimurium* (a bacterial homologue of yeast Kae1) has been solved (11), showing that structurally this protein indeed contains a HALF fold and belongs to the ASKHA (Acetate and Sugar Kinases/Hsc70/Actin) superfamily of phosphotransferases. However, purified YeaZ does not bind ATP and has no glycoprotease activity.

Here, we reported the phylogenetic analysis of Kae1-like proteins and confirmed the universality of this family of protein. We have also performed the biochemical and structural characterization of the Kae1 protein from the hyperthermophilic archaeon *Pyrococcus abyssi* (PAB1159), thereafter called Pa-Kae1. This protein could be a good model to study the biochemical function of eukaryotic Kae1 since Pa-Kae1 shares 42% and 48% identities with yeast Kae1 and human OSGEP, respectively. Although annotated as putative endopeptidase, we show that Pa-Kae1 is an iron-protein with a novel type of ATP-binding site. Surprisingly, this protein is devoid of endopeptidase activity *in vitro* but binds cooperatively DNA and exhibits a class I apurinic (AP)-endonuclease activity (AP-lyase).

MATERIALS AND METHODS

Phylogenetic analysis

All Pa-Kae1 (PAB1159) homologues (i.e. 767 sequences) were retrieved from 409 complete or nearly complete sequenced genomes at the NCBI (ftp.ncbi.nih.gov) and aligned. A phylogenetic tree was then reconstructed using 100 Kae1-related sequences representative of a subset of the 767 Kae1-related sequences. More details are available in Supplementary Data.

Protein purification, crystallization and structure determination

The sequence encoding Pa-Kae1 was amplified by PCR from *P. abyssi* strain GE5 (Orsay) genomic DNA. Recombinant protein was then expressed in *Escherichia coli* and highly purified for spectroscopic characterization and crystallization. Details of crystallization, data collection and structure determination of Pa-Kae1 can be found in Supplementary Materials and Methods. Structures have been deposited into the Brookhaven Protein Data Bank under the accession numbers 2IVN, 2IVO and 2IVP.

DNA-binding analysis and electrophoretic mobility shift assay

DNA-binding reactions were carried out in a mixture composed of single- or double-stranded DNA in Tris-HCl 20 mM pH 7.5, NaCl 100 mM, DTT 1 mM and purified Pa-Kae1 (0.5 nM to 20 μM). Incubation was performed for 10 min at 65°C and reaction products analysed by non-denaturating acrylamide gel electrophoresis. See Supplementary Data for more detailed protocol.

Microscopy

Pa-Kae1/DNA complexes were formed in binding buffer Tris-HCl 20 mM pH 7.5, NaCl 100 mM, DTT 5 mM at 75°C for 10 min with different types of DNA at molar ratio of 1 protein for 5 bp (or nucleotides). Analysis of complexes by electron microscopy is described in more detail in Supplementary Data.

AP-endonuclease activity

AP-endonuclease assays were performed at 72°C for 10 min in a 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT buffer complemented with DNA containing abasic site(s) and purified Pa-Kae1 at molar ratio of 1 protein for 2 bp (or nucleotides). Reactions were stopped and products analysed by electrophoresis. Detailed protocols concerning sample preparations and AP-endonuclease assay are available in Supplementary Data.

RESULTS

Phylogenetic analysis of Kae1 homologues

Although Kae1-like proteins are universal, the phylogenetic relationships among them had never been analysed. This was an important issue considering the biological importance of these proteins and the contradictions in the literature between the biochemical activities reported for different members of this family. We have identified 767 Kae1 homologues present at least in one copy in the 409 genomes available at the NCBI (Figure S1). We identified a single homologue in each archaeal genome whereas we detected one or two homologues in all eukaryotic and bacterial genomes. The phylogeny of the 767 Kae1 homologues recovered four distinct clusters corresponding to: (i) archaeal Kae1 (ii) eukaryotic orthologues of yeast Kae1 and human OSGEP, (iii) bacterial orthologues of the *E. coli* YgjD protein and eukaryotic orthologues of Qri7 and (iv) bacterial orthologues of the *E. coli* protein YeaZ (Figure S1). Interestingly, the eukaryotic orthologues of Kae1/OSGEP and the archaeal Kae1 proteins appear significantly more similar to each other than to the two other subfamilies. We thus suggest reserving the name Kae1 to the archaeal (ar-Kae1) and the Kae1/OSGEP eukaryotic subfamily (ek-Kae1), since the associated kinase Bud32/PRPK is not present in bacteria, and there is presently no indication to suggest that

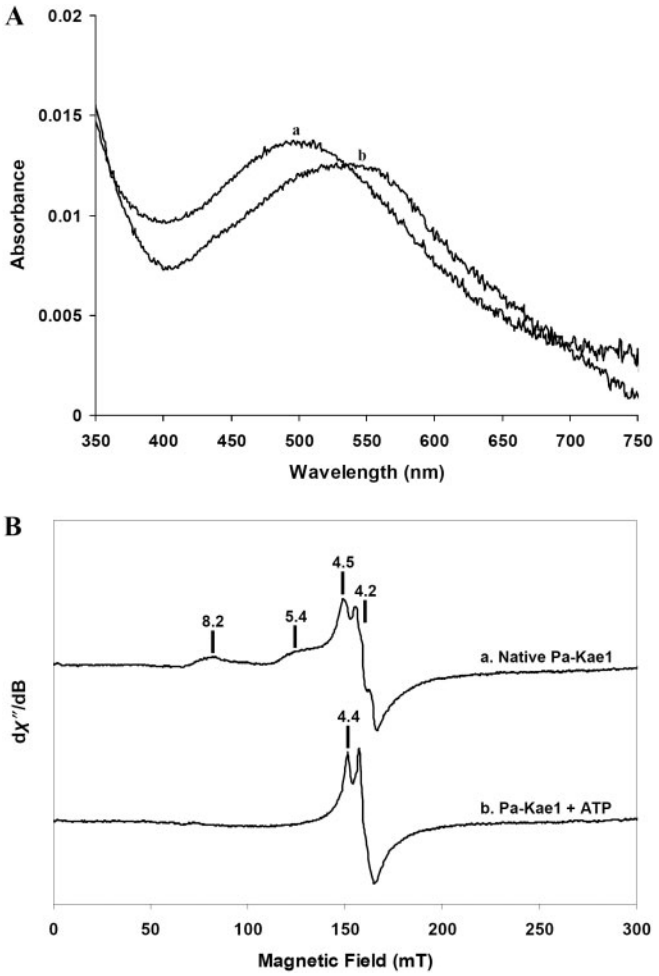

Figure 1. Spectroscopic analysis of Pa-Kae1. (**A**) Light absorbance spectra of native Pa-Kae1 alone (a), in presence of 10 times molar excess of ATP (b). Spectra were obtained using 100 µg of protein in Tris-HCl 20 mM pH 8.0, NaCl 200 mM buffer. (**B**) Electron paramagnetic resonance in absence (a) or presence (b) of ATP; microwave power 0.1 mW; microwave frequency 9.38 GHz; modulation amplitude 1 mT; modulation frequency 100 kHz; temperature 10 K. Some *g* values are indicated on the EPR spectra.

proteins from the second eukaryotic subfamily (Qri7) interact with a kinase.

Pa-Kae1 is an ATP-binding iron metalloprotein

We have expressed a recombinant histidine-tagged Kae1 (thereafter called Pa-Kae1) from the archaeon *P. abyssi* in *E. coli*. The purified protein (36.2 kDa) is present as a monomer in solution (data not shown). Strikingly, Pa-Kae1 protein samples have a pink colour, exhibiting a broad UV-visible absorbance band centred at 500 nm ($\varepsilon_{500nm} = 1830\,M^{-1}.cm^{-1}$, A_{280nm}/A_{500nm} ratio = 9.9 ± 1.2) (Figure 1A), reminiscent of that found in purple acid phosphatases (PAPs) (12). It is well documented that the characteristic purple colour of this subclass of phosphatases results from a charge-transfer transition from a tyrosinate to a Fe(III) ion (12,13). After adding 1 mM ascorbic acid to the Pa-Kae1 solution, the broad absorbance observed at 500 nm gradually declined and disappeared totally after ~45 min (Figure S2), suggesting that Pa-Kae1 indeed contained a Fe^{3+} atom that was reduced into Fe^{2+} by treatment with ascorbic acid. We confirmed the presence of an iron atom in purified recombinant Pa-Kae1 and further determined its iron content by atomic absorption spectroscopy, yielding 0.78 ± 0.02 mol of Fe/mol of protein. The EPR spectrum of Pa-Kae1 at pH 8.0 exhibited broad features at $g = 8.2$ and 5.4 with sharper features at $g = 4.5$ and 4.2, characteristic of a high-spin ($S = 5/2$) Fe(III) species (Figure 1B).

Since previous sequence analysis suggested the presence of an ATP-binding site in Kae1, we incubated Pa-Kae1 with ATP. This resulted both in an immediate colour change from pink to purple corresponding with a shift of the native absorption band at 500 nm to 540 nm (Figure 1A) and a dramatic change in the EPR spectrum (Figure 1B). The broader bands of the spectrum at $g = 8.2$ and 5.4 disappear and the sharper ones are shifted at $g = 4.4$ and 4.2 with a low field component visible at $g = 9.2$. These data suggested that Pa-Kae1 binds ATP in close vicinity of its iron-binding site.

Although Kae1 homologues are systematically annotated as O-syaloglycoprotein endopeptidases, we could not demonstrate any peptidase activity of Pa-Kae1 on glycophoryn A (8), i.e. the O-sialoglycoprotein previously used as substrate to check the endopeptidase activity of the Kae1 protein from *Mannheimia haemolytica* (9). Although the spectroscopic properties of Pa-Kae1 are reminiscent of those of PAPs, we could not detect any phosphatase activity on phosphopeptides RRA(pT)VA (14), END(pY)INASL (15) and DADE(pY)LIPQQG (16), substrates that are used in standard serine/threonine or tyrosine protein phosphatases. In fact, Pa-Kae1 exhibited only a very low ATPase activity (ca. 1.5 pmol/min/µg) corresponding to a low auto-phosphorylation activity on serine, threonine and tyrosine residues (data not shown).

Overall structure of Pa-Kae1

The 3D crystal structure of Pa-Kae1 was solved to 3 Å resolution by the single wavelength anomalous dispersion (SAD) phasing method from a tungstate derivative and further refined to 1.6 Å resolution against native data. The fold of Pa-Kae1 classifies it as a member of the ASKHA superfamily (containing Acetate and Sugar Kinases, Hsc70, Actin) of phosphotransferases (17,18). Pa-Kae1 is made of two similar α/β subdomains (hereafter named I and II), each built around a five-stranded beta sheet with strand order β5, β4, β1, β2 and β3, with β2 antiparallel to the others (Figure 2A) and family-specific insertions (18). The α3 helix of each subdomain packs against the β-sheet of the opposing subdomain and forms a perpendicular α helical layer sandwiched between the two β sheets. Helices α1 and α2 pack against the exterior of each β-sheet. In subdomain II, there are three extra helices (αI, αII and αIII) inserted between β3 and α1, comprised between Pa-Kae1 residues Thr149 and Phe229 (coloured green in Figure 2A). Although the structures of Pa-Kae1 and bacterial homologue YeaZ from *Salmonella typhimurium* are globally similar,

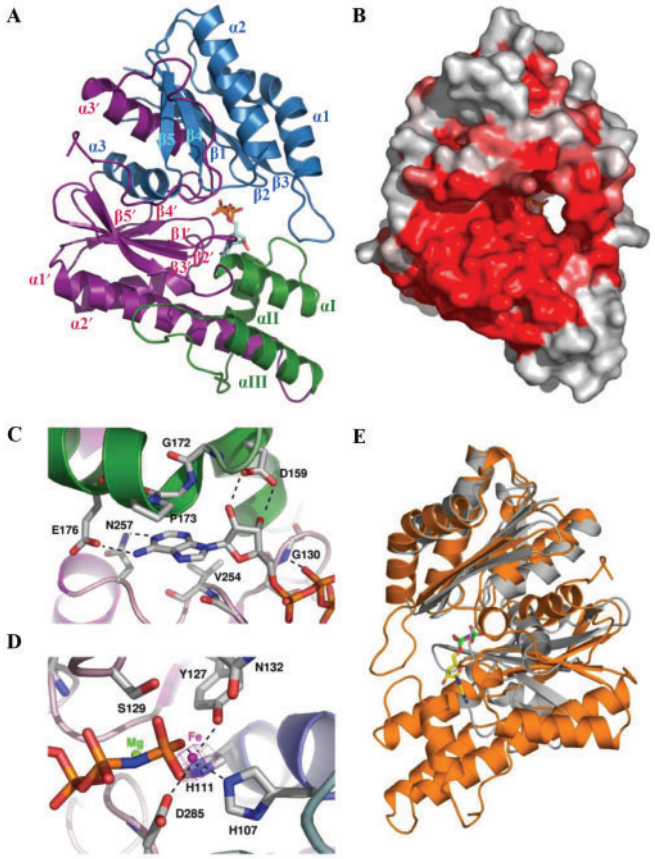

Figure 2. Structure of Pa-Kae1 in complex with AMPPNP. (**A**) Cartoon representation of the Pa-Kae1 monomer. The N-terminal and C-terminal subdomains are coloured blue and purple, respectively. The α-helical cap, inserted in the second subdomain, is shown in green. Bound AMPPNP is shown as sticks. (**B**) Amino acid sequence conservation projected on the molecular surface of Pa-Kae1 (shades of red for highly conserved). AMPPNP inside the tunnel is represented in sticks. (**C**) Detailed view of the base and ribose binding, with the same colour code as panel A. (**D**) Detailed view of the nucleotide phosphate-binding site. The residual anomalous map at 33σ of the dataset collected at the iron absorption edge is shown in purple. The Fe^{3+} atom has been placed at the centre of the density. (**E**) Superposition of Pa-Kae1 (yellow) and YeaZ (grey) on their N-terminal subdomains. AMPPNP from Pa-Kae1 is represented as sticks.

the relative orientation of both subdomains in the two proteins differs considerably. YeaZ and Pa-Kae1 superpose best on the N-terminal subdomain (r.m.s.d. 1.63 Å for 117 Cαs) and start to diverge from residue His111 in Pa-Kae1 corresponding to the helical exit of the subdomain (Figure 2E). The C-terminal subdomain of YeaZ superposes less well (r.m.s.d. 2.96 Å for 82 Cαs) and lacks the large helical insertion between β3 and α1. In YeaZ, the C-terminal subdomain has rotated over ~40° relative to the N-terminal one resulting in a much more open groove between both compared to Pa-Kae1.

The Pa-Kae1 structure in complex with AMPPNP revealed that the nucleotide is bound at the edge of the β-sheets, occupying a channel that is composed of loops and helices from both subdomains (Figure 2A). The adenine ring makes specific base interactions via its N6 and N1 atoms to the Glu176 and Asn257 side chains (Figure 2C). Both O2′ and O3′ hydroxyl groups from the nucleotide ribose moiety are hydrogen bonded to the Asp159 carboxylate. In addition, the 2′OH group of the ribose moiety interacts with the Gly172 amide group. The AMPPNP α-phosphate is contacted by the main chain amide from Asp285, the β-phosphate by the main chain amide from Gly130 and the γ-phosphate is interacting with His107 N2, Tyr127 OH, Ser129 OH and the Asp285 carboxylate groups. Comparison of Pa-Kae1, free and in complex with AMPPNP, shows no evidence for nucleotide induced structural rearrangements. Projection of sequence conservation on the molecular surface of Pa-Kae1 shows a strongly conserved patch near the entrance of the AMPPNP-binding tunnel (Figure 2B). This highly conserved patch suggests that this part of the protein surface may be involved in interactions with a substrate and/or another protein partner within a multi-protein complex.

Coupling of Fe and ATP binding

In an attempt to solve the structure of Pa-Kae1 using the iron anomalous signal, we collected a highly redundant dataset at the wavelength corresponding to the iron X-ray absorption edge. Analysis of the anomalous signal revealed a strong peak (at 33σ) in close proximity to Tyr127 (Figure 2D). This allowed us to unambiguously position the iron atom. The ferric ion (see spectroscopic analysis) is liganded by the δ-OH group of Tyr127, the N2 atoms of His107 and His111, the carboxylate of Asp285 and the γ-phosphate from the AMPPNP ligand. In our refined maps, we observe some residual density very near to both the Fe^{3+} site and the AMPPNP phosphate groups, which could indicate the presence of a partially occupied Mg^{2+} (Mg^{2+} was present in the crystallization liquor). Although the ATP-binding site location in Pa-Kae1 is identical to that of other members of the ASKHA family, the ATP interaction mode differs in one important aspect: the participation of the AMPPNP γ-phosphate group to the coordination sphere of a Fe ion (Figure 2D). This direct contact between the γ-phosphate of AMPPNP and a Fe(III) ion explains the spectroscopic shifts upon addition of ATP to the solution. Although a few members of the ASKHA family bind iron, these sites are far removed from and unrelated to the ATP-binding pocket (19). The direct Fe coordination of ATP is a crucial point in the evaluation of the biochemical function and mechanism of action of Pa-Kae1 and probably of Kae1 in general. The homologous regions/residues interacting with Fe/ATP in Pa-Kae1 are clearly absent in YeaZ, corroborating the fact that YeaZ does not bind nucleotides or metals (11) (Figure 2E).

Pa-Kae1 is a DNA-binding protein

In order to further define the biochemical role of Pa-Kae1, we investigated the genome context of the corresponding gene. We noticed that *kae1* was located in several archaeal genomes nearby genes encoding proteins involved in nucleotide metabolism and/or DNA repair (data not shown). We also noticed that the genes encoding the human and mouse *kae1* orthologues are located near those encoding the eukaryotic homologue of Endonuclease

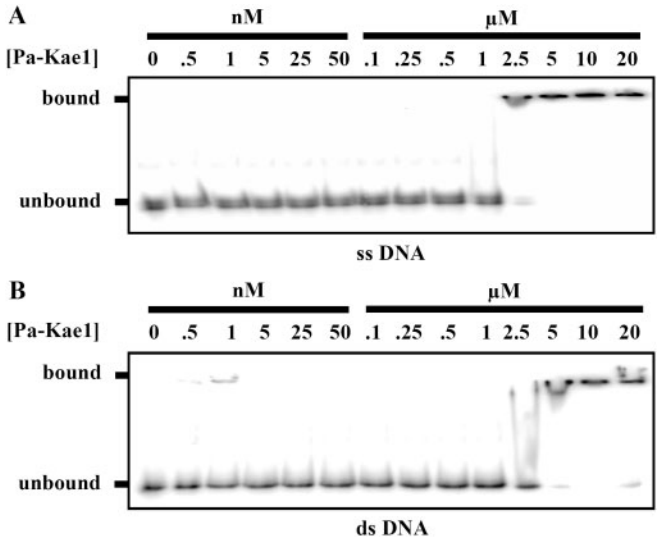

Figure 3. Gel shift DNA-binding assay of Pa-Kae1. Proteins were incubated at the indicated concentrations (0–20 μM) with single (**A**) or double-stranded (**B**) 65-mers oligonucleotide. Free and protein-bound DNAs were separated by acrylamide gel electrophoresis as described in Materials and Methods section.

III, i.e. the major mammalian apurinic endonuclease, APE1 (20). Interestingly, the two genes seem to be transcribed from a single divergent promoter (21,22). Although the structure of Pa-Kae1 did not reveal obvious DNA-binding features, these observations prompted us to test experimentally whether Kae1 interacts with DNA.

Incubation of increasing concentrations of Pa-Kae1 for 10 min at 65°C with single- or double-stranded oligomers of 65 bases at high protein/DNA ratio resulted in a dramatic shift of the oligomer position on polyacrylamide gels. The transition between free DNA and protein–DNA complexes occurred at 2.5 or 5 μM of Pa-Kae1 for single- or double-stranded DNA respectively, corresponding to four proteins per base (Figure 3). Interestingly, the formation of complexes between Pa-Kae1 and DNA was reduced by ATP concentrations above 0.25 mM, and the normal migration pattern of free DNA was nearly restored in the presence of 10 mM ATP (Figure S3). However, DNA (either single- or double-stranded) did not stimulate the low ATPase activity associated to Pa-Kae1.

We further investigated the DNA-binding properties of Pa-Kae1 by transmission electron microscopy, combining non-invasive spreading of biomolecules to high contrast obtained in dark field mode (Figure 4). Complexes

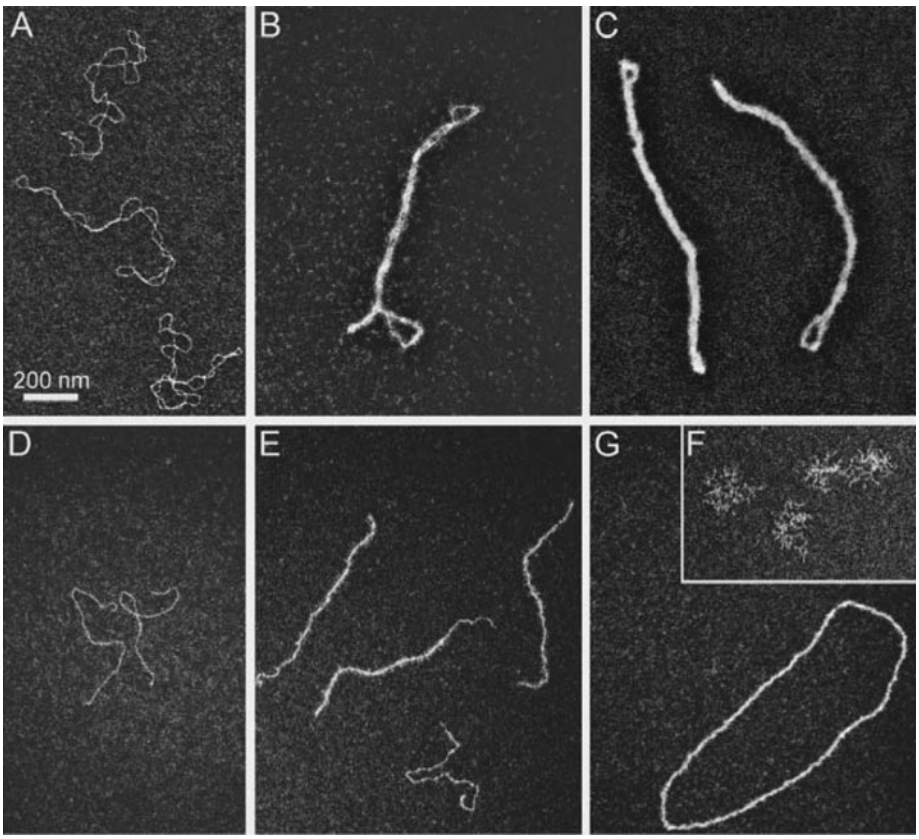

Figure 4. Pa-Kae1–DNA complexes analysed by annular dark field transmission electron microscopy. (**A**) Control negatively supercoiled pBR322 plasmid DNA. (**B**) Pa-Kae1-pBR322 nucleofilament. Plasmid DNA (1.5 μM in bp) was incubated 10 min at 23°C with Pa-Kae1 (300 nM) corresponding to a molar ratio of 1 protein for 5 bp. (**C**) Same as (B) but with an increased protein–DNA molar ratio (1 protein for 2 bp). (**D**) Control linear 1440-bp double-stranded DNA. (**E**) Linear DNA (1.5 μM in bp) of 1440 bp (36) was incubated 10 min at 23°C with Pa-Kae1 (300 nM) corresponding to a molar ratio of 1 protein for 5 bp. (**F**) Control circular single-stranded DNA (ΦX174) showing that the spreading method maintains secondary structures onto EM support grid. (**G**) ΦX174 single-stranded DNA (1.5 μM in nucleotide) was incubated with Pa-Kae1 (300 nM) corresponding to a molar ratio of 1 protein for 5 nt. Scale bar equals 200 nm.

were formed at a molar ratio of 1 protein for 5 bp (or nucleotides). In agreement with the gel shift assay, we observed extensive binding to double-stranded (supercoiled pBR322 and linear double-stranded DNA, Figure 4B–C and E) and single-stranded (ΦX174, Figure 4G) DNA substrates. Pa-Kae1 polymerized along both DNA forms leading to the formation of dense well-ordered nucleofilaments that rigidify DNA. This polymerization appears to be cooperative as illustrated by the presence of totally coated DNA molecules next to free or partially coated ones (Figure 4E). Onto double-stranded DNA, increasing protein DNA molar ratio (1 protein for 2 bp) leads to the thickening of nucleofilaments and formation of spindle-shaped structures (Figure 4C). Intrinsic curvature contained in linear double-stranded DNA fragments disappears (compare Figure 4D and E) and plectonemes contained in negatively supercoiled DNA are removed as the molecules become completely covered (Figure 4A and C). These observations show that DNA conformational changes take place upon Pa-Kae1 binding although no significant DNA lengthening or shortening was measured. Pa-Kae1 exhibits a helix destabilizing activity towards single-stranded DNA, which removes secondary structures. Indeed, Pa-Kae1 completely covers and extends ΦX174 single-stranded circular DNA at a molar ratio of 1 protein for 5 nt (compare Figure 4G and F). Similar spreading was observed for gp32 or *E. coli* ssb at a molar ratio of 1 protein for 30 nt (23,24).

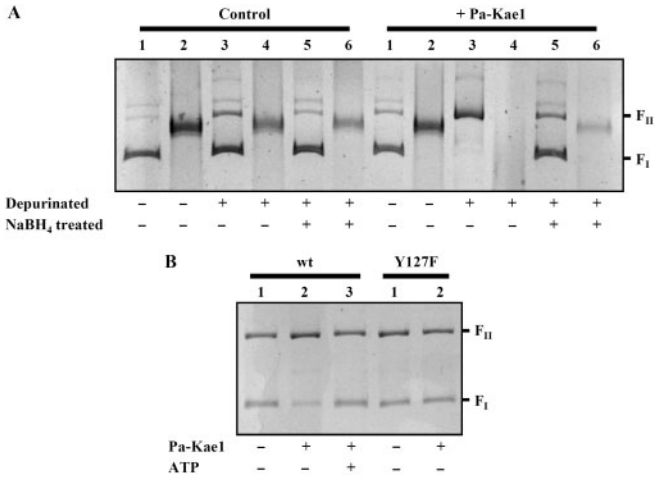

Figure 5. AP-endonuclease activity of Pa-Kae1. (**A**) Detection of AP-endonuclease activity of Pa-Kae1. AP-endonuclease activity was tested on untreated or depurinated DNA at 72°C for 10 min. Samples are then treated by SDS and Proteinase K to dissociate DNA/Pa-Kae1 complexes 30 min at 45°C and analysed by TAE-agarose gel electrophoresis. (1) untreated pBR322 DNA, (2) untreated ΦX174 DNA, (3) depurinated pBR322 DNA, (4) depurinated ΦX174 DNA, (5) depurinated pBR322 DNA treated by $NaBH_4$, (6) depurinated ΦX174 DNA treated by $NaBH_4$ incubated with or without Pa-Kae1 at a molar ratio of 1 protein for 2 nt (or base pairs). (**B**) Effect of the Y127F mutation on Pa-Kae1 activity. Depurinated pBR322 (1) was incubated with wild-type Pa-Kae1 and Y127 mutant as indicated in absence (2) or in presence (3) of ATP 10 mM for 10 min at 72°C. Products are then treated by SDS and Proteinase K for 30 min at 45°C and analysed by TAE-agarose gel electrophoresis.

Pa-Kae1 is a novel type of class I AP-endonuclease

Since, the *kae1* gene is located in the vicinity of the *ape* gene (apurinic endonuclease) in Human and rodents, we thought to check the effect of Pa-Kae1 on depurinated DNA. We thus prepared circular double-stranded depurinated DNA (negatively supercoiled pBR322 DNA), and circular single-stranded depurinated DNA (ΦX174 virion DNA) by incubation for 15 min at 60°C in sodium citrate buffer (25) (Figure 5A, lanes 1–2: untreated pBR322 and ΦX174 DNA and lanes 3–4: depurinated pBR322 and ΦX174, respectively). We found that Pa-Kae1 binds very strongly on depurinated DNA producing high molecular weight structures that did not enter into the agarose gel (data not shown). However, the DNA could be recovered after treatment of the protein/DNA complexes with sodium dodecylsulphate and proteinase K (Figure 5). Surprisingly, all depurinated supercoiled pBR322 (F_I) DNA was cleaved into a relaxed nicked form (F_{II}) after incubation with Pa-Kae1 (compare lanes 3 with and without Pa-Kae1), indicating that Pa-Kae1 exhibited an AP-endonuclease activity. Similarly, depurinated ΦX174 was completely degraded, leading to a smear, as expected for a single-stranded DNA cleaved at multiple AP-sites (compare lanes 4). The nicking activity on apurinic DNA was only observed at temperatures above 60°C and increased at 72°C (data not shown). The AP-endonuclease activity of Pa-Kae1 was stimulated by Mg^{2+} ions and was completely inhibited by EDTA, suggesting that this activity is dependent of a divalent ion bound to the protein. Furthermore, the AP-endonuclease activity was inhibited by ATP at the concentrations that prevented DNA binding, strengthening the idea that this activity was intrinsic to the Pa-Kae1 protein (Figure 5B, compare lanes 1–2 and 3, wild type).

AP-endonuclease activities have been classified into class I and class II according to their mechanism of cleavage (26). Class I AP-endonucleases (also called AP-lyases) cleaves 3′ to the AP site by a β-elimination mechanism producing a single-nucleotide gap flanked by a 5′-phosphate and a 3′ α,β-unsaturated aldehyde, whereas Class II AP-endonucleases catalyse hydrolysis of the sugar-phosphate backbone 5′ of the lesion, producing a single-nucleotide gap flanked by a free 3′ hydroxyl group and 5′-2-deoxyribose-5-phosphate (5′-dRP). It is possible to discriminate between class I and II by testing their activity on a depurinated DNA treated by sodium borohydride ($NaBH_4$) which reduces the AP site, since only class II AP-endonucleases can cleave such substrate (27). As shown in Figure 5, Pa-Kae1 is unable to cleave depurinated DNA, which has been reduced by $NaBH_4$ (either double-stranded pBR322 DNA or single-stranded ΦX174 DNA) (compare lanes 5 and lanes 6).

To confirm our results obtained with DNA depurinated by acid treatment, we tested the activity of Pa-Kae1 on oligonucleotides containing a single abasic site. We thus prepared single- and double-stranded 65-mers containing a uracil residue at position 22 and we produced an apurinic site by treating these oligomers with uracil-DNA glycosylase (UDG). As shown in Figure 6, Pa-Kae1 can cleave these 65-mers, leading to fragments of the expected

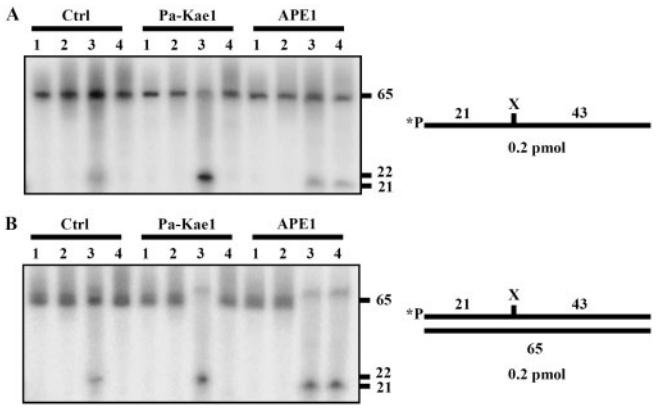

Figure 6. AP-endonuclease activity of Pa-Kae1 on oligonucleotides substrates. (**A**) AP-endonuclease activity was tested at 65°C for 10 min on a 65-mers oligonucleotide substrates (0.2 pmol) containing (1) adenine, (2) uracil, (3) abasic site or (4) THF at position 22. Products were analysed by denaturing acrylamide gel electrophoresis. The uncut 65-mers DNA substrate and cleaved 21 or 22-mers products are indicated on the right. Oligonucleotide containing abasic site or THF is uncut by APE1 (compare with oligonucleotides incubated without enzyme). (**B**) Same as (A) but using double-stranded 65-mers oligonucleotides. Double-stranded oligonucleotide containing abasic site or THF is cut by APE1 (control) producing a 21-mers fragment.

sizes if cleavage has occurred in 3′ (22-mers), indicating that the cleavage had indeed occurred at the AP site using a class I type mechanism (lanes 3). Also, as expected for a class I AP-endonuclease, Pa-Kae1 did not cleave single- or double-stranded oligomers in which the nucleosides at position 22 were replaced by tetrahydrofurane (THF), mimicking a reduced abasic site (compare lanes 3 and 4). As control, Figure 6B, lane 4 shows that our oligomers containing THF was cleaved by the human Class II AP-endonuclease APE1 (a double-stranded specific AP-endonuclease), generating a 21-mers products as expected for a cleavage in 5′. Taken together, our experiments thus demonstrate that Pa-Kae1 cleaves apurinic DNA by a mechanism similar to those of a class I AP-endonuclease (AP-lyase).

Unlike previously described AP-endonucleases of both class I and II, cleavage of apurinic DNA by Pa-Kae1 is only observed at high protein to DNA ratio. To determine more precisely the relationships between DNA binding and cleavage, we tested increasing concentrations of Pa-Kae1 on our 65-mers oligonucleotides containing a single AP site (Figure S4). Interestingly, one can notice a perfect correlation between the concentrations required for DNA binding and for cleavage (compare Figures 3 and S4). Cooperative binding of Pa-Kae1 to DNA thus appears to be a prerequisite for cleavage of the apurinic sites.

In order to determine if the iron atom of Pa-Kae1 is involved in its AP-endonuclease activity, we replaced tyrosine 127 by phenylalanine by site-directed mutagenesis. The broad UV-visible absorbance centred at 500 nm disappeared (Figure S5A) and the EPR signal (Figure S5B) was dramatically modified, indicating that the mutated protein has lost its iron atom and confirming that Tyr127 is essential to the binding and stabilization of the Fe(III) ion. The mutated proteins could still produce gel shift with DNA, indicating that the mutation did not affect DNA binding (data not shown). However, as shown in Figure 5B, the class I AP-endonuclease activity decreased in the Y127F mutant (compare lanes 2 wild-type and mutant Y127F), indicating that the iron atom linked to Pa-Kae1 is involved in the AP-endonuclease activity and confirming that this activity is not due to an *E. coli* contaminant.

DISCUSSION

Phylogenomics confirms the universality of Kae1-related proteins (ar-Kae1, ek-Kae1, Qri7, YgjD and YeaZ) suggesting their very ancient origin and major role in the three domains of life. Previous systematic genetic approaches have also shown that Kae1, in yeast (28), and YgjD, in *E. coli* and *B. subtilis* (29), are essential. We have found that Pa-Kae1 (a representative of ar-Kae1) exhibits unique features that highlight the need for further in-depth studies of all Kae1-related proteins in the three domains of life.

First, we found that Kae1 is a metalloprotein containing an iron atom directly linked to ATP via the gamma phosphate. This mode of interaction between ATP and iron was never previously observed in other proteins. The amino acids involved in this interaction (His107, His117, Tyr127 and Asp285) are strictly conserved in all ek-Kae1 sequences and nearly all archaeal ones, indicating that this novel type of Fe/ATP-binding site is common to all Kae1 proteins (Figure S6). Mutation of these two histidines in the yeast protein indeed abolished the activity of the EKC complex *in vivo* (2). It was previously hypothesized that these two histidines were used to coordinate a zinc atom and were the signature of a metalloprotease. In fact, purified Pa-Kae1 did not exhibit the glycoprotease activity supposed to be associated to this protein family. One cannot exclude that such activity requires specific substrate or reaction conditions. Ping Ping Ng *et al.* (30) recently reported that human Kae1 (OSGEP) performs cleavage of misfolded proteins in acute promyelocitic leukaemia (OSGEP). Nevertheless, a doubt still remains concerning this activity since it was measured from partially purified fractions. Moreover, the protease activity previously detected in *P. haemolytica* was never confirmed in other organisms. It was even suggested by the authors of the first annotation that the identification of Kae1 as a glycoprotease might have been misleading (31). Although the name Kae1 (Kinase-associated endopeptidase 1) initially referred to its assumed glycoprotease activity, we notice that Kae1 could also mean kinase-associated endonuclease 1.

Surprisingly, Pa-Kae1 turned out to be a completely novel type of DNA-binding protein. Pa-Kae1 binds to both single- and double-stranded DNA, whether linear or closed circular. The highly cooperative binding processes leads to the formation of rigid nucleofilaments. Protein–protein interactions that depend on DNA binding appear to play a major role in the polymerization of Pa-Kae1 along DNA. The high concentration of protein required to

observe DNA binding indicates that the affinity of a single molecule for DNA should be very low, in agreement with the absence of large positively charged region at the surface of the molecule. A few DNA-binding proteins exhibit DNA-binding features that resemble to those of Pa-Kae1, but which also display striking difference. For instance nucleofilament formation by DNA recombinases (UvsX, RecA, RadA and Rad51) requires ATP and increases DNA length whereas in the case of Pa-Kae1, DNA binding is inhibited by ATP and DNA length is not affected. Proteins of the Alba family also bind equally single-stranded DNA and double-stranded DNA without ATP requirement as in the case of Pa-Kae1 but they form spindle-shaped structures that reduce DNA length (32).

Interestingly, Pa-Kae1 induces conformational changes within double-stranded DNA that reduce curvature of linear and modify topological parameters of circular fragments, leading to the apparent relaxation of supercoiling. The mechanism producing these conformational changes is unusual since this apparent relaxation does not correlate with a significant increase in DNA length (by means of helical twist decrease) as it is the case for DNA recombinases, neither with loop formation (by means of writhe modification) as it is the case in nucleosome-like structures. The modifications in DNA conformation induced by Kae1 coating on DNA could be explained by the ability of DNA to accommodate high twist with no length increase by virtue of its surprisingly low twist-stretch modulus (33).

Finally, a third striking feature of Pa-Kae1 is its AP-endonuclease activity. The AP-endonuclease activity of Pa-Kae1 requires the iron atom since it is abolished in the Y127F mutant. Inhibition by EDTA indicates that another metal is involved in the cleavage reaction. Indeed, we observed the presence of a second metal close to the iron-ATP active site in the Pa-Kae1 structure. A reduced apurinic site is not substrate of Pa-Kae1, indicating that this protein ressembles a class I AP-endonuclease (also called AP-lyase). However, Pa-Kae1 has no structural similarities with other class I AP-endonucleases, such as *E. coli* Endonuclease III. Furthermore, its mechanism of action is highly unusual (requirement for a high protein to DNA ratio, cooperative binding to DNA, inhibition by ATP) indicating that Pa-Kae1 is a completely novel type of AP-endonuclease. In fact, AP-endonuclease activity (neither DNA binding) has never been reported for any protein of the ASKHA superfamily.

Interestingly, both the DNA binding and AP endonuclease activities of Pa-Kae1 are inhibited by high concentrations of ATP, suggesting that ATP could modulate the function(s) of Kae1 *in vivo*. This also indicates that DNA should bind somehow near the ATP-site linked to the iron atom, in agreement with the involvement of the iron in the AP endonuclease activity of Pa-Kae1. However, there is presently no hint from the structure concerning the precise mode of DNA binding onto Pa-Kae1. Further work, including crystallization of a Pa-Kae1/DNA complex will be required to solve this problem.

Considering the high sequence similarity between archaeal and eukaryotic Kae1 proteins and their common interaction with homologous kinases, it is likely that the properties of Pa-Kae1 described here are shared by their eukaryotic orthologues. The DNA-binding capacity of Pa-Kae1 make senses considering the participation of its yeast homologue in complexes involved in transcription and telomeres formation in yeast. Our findings suggest that Kae1 modulates the activities of its partners in the KEOPS and EKC complexes via its DNA-binding activity. Indeed, yeast Kae1 can be cross-linked to DNA, suggesting that this protein physically interacts with chromatin (2). Libri and co-workers (2) have shown that the EKC complex is important in yeast for the establishment and maintenance of a chromatin structure competent for transcription, possibly through an ATP-dependent remodelling function borne by Kae1. Similarly, Downey and co-workers (1) proposed that KEOPS promotes an 'open' telomere conformation that allows telomerase access and promotes exonuclease degradation of the chromosome end following telomere uncapping. These authors have suggested that the proteolytic activity of Kae1 somehow regulates the activity of KEOPS/EKC complexes. One cannot completely exclude that Kae1 exhibits a proteolytic activity *in vivo* that escaped our analysis *in vitro*. Similarly, it remains to be shown that Kae1 binds DNA *in vivo* as *in vitro*. However, the participation of Kae1 to complexes involved in processes dealing with DNA (telomeres formation, transcription) led us to believe that the DNA-binding properties and AP-endonuclease activities that we have discovered *in vitro* are relevant for its biological role. We propose specifically that a major role of Kae1 is to anchor the KEOPS/EKC complex to transcription sites and chromosome ends via its DNA-binding properties and to induce conformational changes that could trigger chromatin remodelling.

The AP-endonuclease activity of Pa-Kae1 is especially interesting in this context, since it implies that Kae1 can recognize abasic sites, which are major factors of genome instability. It is possible that Kae1 can recognize other types of DNA-damaged nucleotides, as suggested by the colocalization of its gene in archaeal genomes with those encoding Ham1, an enzyme involved in the hydrolysis of non-canonical nucleotides. We thus propose that, in addition to its structural role, Kae1 is involved in checking DNA damage. Moreover, a functional linkage between Kae1 and PRPK, a kinase known to phosphorylate p53 (a major protein in the control of genome integrity), has been observed in Human (4). The participation of Kae1 to the p53 network would explain why inhibition of human *kae1* gene expression via RNA interference interferes with tumour cell growth and promotes apoptosis of APL tumour cells (30). Kae1 thus could be a major previously unrecognized component of checkpoint in mammals.

The presence of Kae1 homologues in all completely sequenced genomes without a single exception is especially challenging, since very few proteins involved in the metabolism of DNA belong to the set of universal proteins (6). In particular, the list of universal proteins only contains two DNA repair proteins, the recombinase Rad51 (RadA/RecA) and the two major proteins of the MRE complex, Mre11 (SbcC) and Rad50 (SbcD).

Several authors suggested that the last universal ancestor of the three domains of life, LUCA, had an RNA or RNA/DNA genome, and that most proteins involved in DNA metabolism only appeared after the divergence of the major cellular lineages (34,35). It will be especially important to determine if bacterial homologues of Kae1 also are DNA-binding proteins. We anticipate that the study of Kae1 proteins in archaea and eukaryotes and of their bacterial homologues will be a major task in future studies on DNA repair mechanisms and chromatin structure.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported by the Centre National de la Recherche Scientifique, by the Human Frontier Science Program (HSFP), the Association pour la Recherche sur le Cancer (ARC), the Agence Nationale de la Recherche (ANR) and by the European 3D-repertoire program (LSHG-CT-2005-512028). The authors are indebted to Alain Boussac from the URA CNRS 2096 CEA-Saclay for atomic absorption spectroscopy experiments and to Nathalie Ulryck for crystallization trials. Funding to pay the Open Access publication charges for this article was provided by Institut Pasteur.

Conflict of interest statement. None declared.

REFERENCES

1. Downey,M., Houlsworth,R., Maringele,L., Rollie,A., Brehme,M., Galicia,S., Guillard,S., Partington,M., Zubko,M.K. *et al.* (2006) A genome-wide screen identifies the evolutionarily conserved KEOPS complex as a telomere regulator. *Cell*, **124**, 1155–1168.
2. Kisseleva-Romanova,E., Lopreiato,R., Baudin-Baillieu,A., Rousselle,J.C., Ilan,L., Hofmann,K., Namane,A., Mann,C. and Libri,D. (2006) Yeast homolog of a cancer-testis antigen defines a new transcription complex. *EMBO J.*, **25**, 3576–3585.
3. Lopreiato,R., Facchin,S., Sartori,G., Arrigoni,G., Casonato,S., Ruzzene,M., Pinna,L.A. and Carignani,G. (2004) Analysis of the interaction between piD261/Bud32, an evolutionarily conserved protein kinase of *Saccharomyces cerevisiae*, and the Grx4 glutaredoxin. *Biochem J.*, **377**, 395–405.
4. Abe,Y., Matsumoto,S., Wei,S., Nezu,K., Miyoshi,A., Kito,K., Ueda,N., Shigemoto,K., Hitsumoto,Y. *et al.* (2001) Cloning and characterization of a p53-related protein kinase expressed in interleukin-2-activated cytotoxic T-cells, epithelial tumor cell lines, and the testes. *J. Biol. Chem.*, **276**, 44003–44011.
5. Facchin,S., Lopreiato,R., Ruzzene,M., Marin,O., Sartori,G., Gotz,C., Montenarh,M., Carignani,G. and Pinna,L.A. (2003) Functional homology between yeast piD261/Bud32 and human PRPK: both phosphorylate p53 and PRPK partially complements piD261/Bud32 deficiency. *FEBS Lett.*, **549**, 63–66.
6. Koonin,E.V. (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.*, **1**, 127–136.
7. Galperin,M.Y. and Koonin,E.V. (2004) 'Conserved hypothetical' proteins: prioritization of targets for experimental study. *Nucleic Acids Res.*, **32**, 5452–5463.
8. Abdullah,K.M., Lo,R.Y. and Mellors,A. (1991) Cloning, nucleotide sequence, and expression of the Pasteurella haemolytica A1 glycoprotease gene. *J. Bacteriol.*, **173**, 5597–5603.
9. Abdullah,K.M., Udoh,E.A., Shewen,P.E. and Mellors,A. (1992) A neutral glycoprotease of Pasteurella haemolytica A1 specifically cleaves O-sialoglycoproteins. *Infect. Immun.*, **60**, 56–62.
10. Aravind,L. and Koonin,E.V. (1999) Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. *J. Mol. Biol.*, **287**, 1023–1040.
11. Nichols,C.E., Johnson,C., Lockyer,M., Charles,I.G., Lamb,H.K., Hawkins,A.R. and Stammers,D.K. (2006) Structural characterization of Salmonella typhimurium YeaZ, an M22 O-sialoglycoprotein endopeptidase homolog. *Proteins*, **64**, 111–123.
12. Klabunde,T., Strater,N., Frohlich,R., Witzel,H. and Krebs,B. (1996) Mechanism of Fe(III)-Zn(II) purple acid phosphatase based on crystal structures. *J. Mol. Biol.*, **259**, 737–748.
13. Vincent,J.B. and Averill,B.A. (1990) An enzyme with a double identity: purple acid phosphatase and tartrate-resistant acid phosphatase. *FASEB J.*, **4**, 3009–3014.
14. Donella Deana,A., Mac Gowan,C.H., Cohen,P., Marchiori,F., Meyer,H.E. and Pinna,L.A. (1990) An investigation of the substrate specificity of protein phosphatase 2C using synthetic peptide substrates; comparison with protein phosphatase 2A. *Biochem. Biophys. Acta*, **1051**, 199–202.
15. Daum,G., Solca,F., Diltz,C.D., Zhao,Z., Cool,D.E. and Fischer,E.H. (1993) A general peptide substrate for protein tyrosine phosphatases. *Anal. Biochem.*, **211**, 50–54.
16. Zhang,Z.Y., Thieme-Sefler,A.M., Maclean,D., McNamara,D.J., Dobrusin,E.M., Sawyer,T.K. and Dixon,J.E. (1993) Substrate specificity of the protein tyrosine phosphatases. *Proc. Natl Acad. Sci. USA*, **90**, 4446–4450.
17. Buss,K.A., Cooper,D.R., Ingram-Smith,C., Ferry,J.G., Sanders,D.A. and Hasson,M.S. (2001) Urkinase: structure of acetate kinase, a member of the ASKHA superfamily of phosphotransferases. *J. Bacteriol.*, **183**, 680–686.
18. Hurley,J.H. (1996) The sugar kinase/heat shock protein 70/actin superfamily: implications of conserved structure for mechanism. *Annu. Rev. Biophys. Biomol. Struct.*, **25**, 137–162.
19. Locher,K.P., Hans,M., Yeh,A.P., Schmid,B., Buckel,W. and Rees,D.C. (2001) Crystal structure of the Acidaminococcus fermentans 2-hydroxyglutaryl-CoA dehydratase component A. *J. Mol. Biol.*, **307**, 297–308.
20. Robson,C.N. and Hickson,I.D. (1991) Isolation of cDNA clones encoding a human apurinic/apyrimidinic endonuclease that corrects DNA repair and mutagenesis defects in E. coli xth (exonuclease III) mutants. *Nucleic Acids Res.*, **19**, 5519–5523.
21. Ikeda,S., Ayabe,H., Mori,K., Seki,Y. and Seki,S. (2002) Identification of the functional elements in the bidirectional promoter of the mouse O-sialoglycoprotein endopeptidase and APEX nuclease genes. *Biochem. Biophys. Res. Commun.*, **296**, 785–791.
22. Seki,Y., Ikeda,S., Kiyohara,H., Ayabe,H., Seki,T. and Matsui,H. (2002) Sequencing analysis of a putative human O-sialoglycoprotein endopeptidase gene (OSGEP) and analysis of a bidirectional promoter between the OSGEP and APEX genes. *Gene*, **285**, 101–108.
23. Chrysogelos,S. and Griffith,J. (1982) Escherichia coli single-strand binding protein organizes single-stranded DNA in nucleosome-like units. *Proc. Natl Acad. Sci. USA*, **79**, 5803–5807.
24. Delius,H., Mantell,N.J. and Alberts,B. (1972) Characterization by electron microscopy of the complex formed between T4 bacteriophage gene 32-protein and DNA. *J. Mol. Biol.*, **67**, 341–350.
25. Xanthoudakis,S., Miao,G., Wang,F., Pan,Y.C. and Curran,T. (1992) Redox activation of Fos-Jun DNA binding activity is mediated by a DNA repair enzyme. *EMBO J.*, **11**, 3323–3335.
26. Levin,J.D. and Demple,B. (1990) Analysis of class II (hydrolytic) and class I (beta-lyase) apurinic/apyrimidinic endonucleases with a synthetic DNA substrate. *Nucleic Acids Res.*, **18**, 5069–5075.
27. Verly,W.G., Colson,P., Zocchi,G., Goffin,C., Liuzzi,M., Buchsenschmidt,G. and Muller,M. (1981) Localization of the phosphoester bond hydrolyzed by the major apurinic/apyrmidinic endodeoxyribonuclease from rat-liver chromatin. *Eur. J. Biochem./FEBS*, **118**, 195–201.
28. Winzeler,E.A., Shoemaker,D.D., Astromoff,A., Liang,H., Anderson,K., Andre,B., Bangham,R., Benito,R., Boeke,J.D. *et al.* (1999) Functional characterization of the S. cerevisiae

genome by gene deletion and parallel analysis. *Science*, **285**, 901–906.
29. Arigoni,F., Talabot,F., Peitsch,M., Edgerton,M.D., Meldrum,E., Allet,E., Fish,R., Jamotte,T., Curchod,M.L. *et al.* (1998) A genome-based approach for the identification of essential bacterial genes. *Nat. Biotechnol.*, **16**, 851–856.
30. Ng,A.P., Howe Fong,J., Sijin Nin,D., Hirpara,J.L., Asou,N., Chen,C.S., Pervaiz,S. and Khan,M. (2006) Cleavage of misfolded nuclear receptor corepressor confers resistance to unfolded protein response-induced apoptosis. *Cancer Res.*, **66**, 9903–9912.
31. Jiang,P. and Mellors,A. (2004) In Barrett,A.J., Rawlings,N.D. and Woessner,J.F. (eds), *Handbook of Proteolytic Enzymes*. Elsevier, London, pp. 977–980.
32. Lurz,R., Grote,M., Dijk,J., Reinhardt,R. and Dobrinski,B. (1986) Electron microscopic study of DNA complexes with proteins from the Archaebacterium *Sulfolobus acidocaldarius*. *EMBO J.*, **5**, 3715–3721.
33. Lionnet,T., Joubaud,S., Lavery,R., Bensimon,D. and Croquette,V. (2006) Wringing out DNA. *Phys. Rev. Lett.*, **96**, 178102.
34. Forterre,P. (2006) Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. *Proc. Natl Acad. Sci. USA*, **103**, 3669–3674.
35. Leipe,D.D., Aravind,L. and Koonin,E.V. (1999) Did DNA replication evolve twice independently? *Nucleic Acids Res.*, **27**, 3389–3401.
36. Beloin,C., Jeusset,J., Revet,B., Mirambeau,G., Le Hegarat,F. and Le Cam,E. (2003) Contribution of DNA conformation and topology in right-handed DNA wrapping by the Bacillus subtilis LrpC protein. *J. Biol. Chem.*, **278**, 5333–5342.