The universal Kae1 protein and the associated Bud32 kinase (PRPK), a mysterious protein couple probably essential for genome maintenance in Archaea and Eukarya

Arnaud Hecker*, Marc Graille†, Edwige Madec‡, Danièle Gadelle*, Eric Le Cam§, Herman van Tilbergh† and Patrick Forterre*‡

*Université Paris-Sud, Institut de Génétique et Microbiologie, IFR115, UMR8621-CNRS, 91405 Orsay Cedex, France, †Université Paris-Sud, Institut de Biochimie et de Biophysique Moléculaire et Cellulaire, IFR115, UMR8619-CNRS, 91405 Orsay, France, ‡Université des Sciences et Technologies de Lille, Laboratoire de Génétique des Enveloppes Bactériennes, CNRS UMR 8576, 59655 Villeneuve d’Ascq, France and §Université Paris-Sud, Institut Gustave Roussy, Interactions Moléculaires et Cancer, UMR8126-CNRS, 94805 Villejuif Cedex, France

Abstract
The similarities between essential molecular mechanisms in Archaea and Eukarya make it possible to discover, using comparative genomics, new fundamental mechanisms conserved between these two domains. We are studying a complex of two proteins conserved in Archaea and Eukarya whose precise biological role and biochemical function remain unknown. One of them is a universal protein known as Kae1 (kinase-associated endopeptidase 1). The second protein is a serine/threonine kinase corresponding to the proteins Bud32 in Saccharomyces cerevisiae and PRPK (p53-related protein kinase) in humans. The genes encoding the archaeal orthologues of Kae1 and PRPK are either contiguous or even fused in many archaeal genomes. In S. cerevisiae, Kae1 and Bud32 (PRPK) belong to a chromatin-associated complex [KEOPS (kinase, endopeptidase and other proteins of small size)/EKC (endopeptidase-like kinase chromatin-associated)] that is essential for telomere elongation and transcription of essential genes. Although Kae1 is annotated as O-sialoglycoprotein endopeptidase in most genomes, we found that the Kae1 protein from Pyrococcus abyssi has no protease activity, but is an atypical DNA-binding protein with an AP (apurinic) lyase activity. The structure of the fusion protein from Methanocaldococcus jannaschii revealed that Kae1 maintains the ATP-binding site of Kae1 in an inactive configuration. We have in fact found that Kae1 inhibits the kinase activity of Bud32 (PRPK) in vitro. Understanding the precise biochemical function and biological role of these two proteins (which are probably essential for genome maintenance) remains a major challenge.

Introduction
Until now, most studies on Archaea have focused on systems already known and well described in Eukarya or Bacteria (for reviews, see recent monographs edited by Garrett and Klenk [1] and Cavicchioli [2]). From these initial studies, Archaea have turned out to be a goldmine for molecular biologists. In particular, Archaea, being very similar to eukaryotes in term of fundamental molecular mechanisms, have been especially useful as simplified models for eukaryotic molecular biology. Hence, the discovery of topoisomerase IIB led to the identification of the protein that triggers meiotic recombination in Eukarya [3]. More recently, previous knowledge on the crenarchaeal SSB (single-stranded DNA-binding protein) led to the discovery of a second type of SSB in humans [4]. The accumulation of genomic data now makes it possible to follow a more challenging strategy, i.e. to use genome data mining to detect molecular systems shared by Archaea and Eukarya which are either poorly characterized or as yet unknown. During the analysis of the genome of the archaean Pyrococcus abyssi [5], we noticed the presence of many genes that have only archaeal and eukaryotic homologues and whose functions were uncharacterized. We established a list of such proteins, that we called PACEs (proteins of Archaea conserved in Eukarya) [6]. This new line of research led us to characterize a PACE whose biological function remained unknown: a kinase known as PRPK (p53-related protein kinase) in humans and Bud32 in yeast. In the course of our study, we found that, in some archaea, this PACE is fused to a protein known as Kae1 (kinase-associated endopeptidase 1) in yeast. Whereas this kinase is present...
only in Archaea and Eukarya, the Kae1 protein is universal. Biochemical analysis of these two proteins led us to conclude that they are involved in a still unknown mechanism required for the maintenance of genome integrity.

Detection and analysis of putative serine/threonine kinases in *P. abyssi*

Because of our long interest in archaeal DNA replication [7,8], we were *a priori* interested in protein kinases from Archaea at the beginning of this project. Many eukaryotic DNA replication [9] and/or repair [10] proteins are regulated by cycles of phosphorylation–dephosphorylation, and we wondered whether one or several archaeal kinases could be involved in similar regulatory pathways. First, we noticed the presence of two putative serine/threonine kinases in our list of PACEs (PAB0405 and PAB1013). Then, a systematic screening of the *P. abyssi* genome for the presence of genes harbouring typical kinase signatures (subdomains I–VIII) [11] permitted us to identify a third putative candidate, PAB1047. These three proteins (presenting all motifs typical of serine/threonine protein kinases) exhibit good similarities with the so-called RIO kinases [12] (Figure 1A). The three *P. abyssi* putative kinase genes were cloned and expressed in Escherichia coli. The recombinant purified proteins were tested on MBP (myelin basic protein) [13,14], a useful substrate for *in vitro* kinase assays. As shown in Figures 1(B) and 1(C), all of them exhibit autophosphorylation and kinase activities on MBP (strongly stimulated by Mn2+ in the case of PAB1013 and PAB1047). The relationships between those archaeal kinases and their eukaryotic homologues showed that PAB1047 belongs to a subfamily that includes the Bud32 protein in *Saccharomyces cerevisiae* [15,16] and PRPK (p53-related protein kinase) in humans [17] (Figure 2). Both Bud32 and PRPK phosphorlyate anti-oncogen p53, and PRPK partially complements Bud32 deficiency in yeast [18]. Since p53 is an extremely important protein in vertebrates that is essential for maintenance of genome integrity, it was interesting to know whether the archaeal orthologue of PRPK could also be involved in some function related to genome stability.

Kae1: a universal protein associated with Bud32/PRPK

By analysing the genomic context of the genes encoding Bud32/PRPK, we realized that the gene encoding PAB1047 is often fused in archaeal genomes with another gene whose homologue in the bacterium *Pasteurella haemolytica* was supposed to code for an OSGEP (O-sialoglycoprotein endopeptidase) [19,20]. This latter gene corresponds to the PAB1159 protein in *P. abyssi* and has been called Kae1 in *S. cerevisiae*. Functional interaction between Kae1 and the kinase Bud32 (the orthologue of PAB1047 in *S. cerevisiae*) was demonstrated previously by two-hybrid analyses [21]. The fusion of the gene encoding these two proteins in several archaeal genomes suggested that their functional and structural interactions are probably conserved in Archaea and Eukaryota, making the study of this protein couple in Archaea a fascinating objective (Figure 2).

Interestingly, Kae1 homologues are present not only in Archaea and Eukarya, but also in Bacteria. Strikingly, Kae1 homologues are, in fact, present in all genomes which have been sequenced to date [22,23], with the single exception of the highly reduced genome of the endosymbiotic bacteria *Carsonella ruddii* and *Salvia musselli*. Kae1 thus belongs to the limited set of approx. 60 universal proteins that were probably already present in the LUCA (last universal cellular ancestor) of all extant life. As a matter of fact, Galperin and Koonin [22] put Kae1 as number one in their top ten list of the most important proteins to be studied, because it is the only universal protein whose biological function remains unknown.

The archaeal Kae1 protein is the prototype of a novel class of DNA-binding proteins with atypical AP (apurinic) endonuclease activity

We first expressed and purified a recombinant Kae1 (hereafter called Pa-Kae1) from the archaeon *P. abyssi* in *E. coli*. Strikingly, Pa-Kae1 protein samples have a pink colour, exhibiting a broad UV–visible absorbance band centred at 500 nm reminiscent of that found in PAPs (purple acid phosphatases) [24]. The characteristic purple colour of this subclass of phosphatases results from a charge-transfer transition from a tyrosine residue to a Fe(III) ion [24,25], indicating that Pa-Kae1 is a metalloprotein. We obtained the three-dimensional crystal structure of Pa-Kae1 which has shown that this protein belongs to the ASKHA [acetate and sugar kinases/Hsc70 (heat-shock cognate 70)/actin] superfamily of phosphotransferases [23]. The structure of Pa-Kae1 in complex with 3′-P(NH)ppA (adenosine 5′-(β,γ-imido)triphosphate) revealed an unusual linkage between the iron and the γ-phosphate of the nucleotide. This mode of interaction between ATP and iron was never previously observed in other proteins. The amino acids involved in this interaction (*His*127, *His*137, *Tyr*127 and Asp289) are strictly conserved in all eukaryotic Kae1 sequences and nearly all archaeal ones, indicating that this novel type of iron/ATP-binding site is common to all Kae1 proteins. Mutation of these two histidine residues in the yeast protein is lethal [26]. On the other hand, mutation of *Tyr*127 removes the iron atom from the Pa-Kae1 protein *in vitro* [23].

Despite the presence of an ATP-binding site and some characteristic features shared by Kae1 and PAPs, Pa-Kae1 only exhibits an autophosphorylation activity on serine, threonine and tyrosine residues. We failed to detect *in vitro* any phosphatase or kinase activities, but we found a very low ATPase activity corresponding to an autophosphorylation of the protein. In particular, we did not detect the expected OSGEP activity, leading us to conclude a misannotation of this gene and other homologues. The protease activity previously detected in *Pasteurella haemolytica* was indeed never confirmed in other organisms. It was even suggested by the authors of the first annotation that the identification...
Figure 1  Detection and analysis of putative serine/threonine kinases in *P. abyssi*

(A) Alignment of PAB0405, PAB1013 and PAB1047 serine/threonine kinases with RIO kinases from *A. fulgidus*. Consensus sequences for each kinase subdomains (subdomains I-XI) as described by Hanks and Hunter [11] are shown. Upper-case letters, universally conserved amino acid residues; lower-case letters, highly conserved amino acid residues; o, positions conserving non-polar residues; *, positions conserving polar residues; x, any amino acid; +, positions conserving small residues with near-neutral polarity. Question marks designate motifs for subdomains for which no homologous sequences were apparent. (B) Autophosphorylation of PAB0405, PAB1013 and PAB1047. Purified kinase (approx. 1 μg) was subjected to autophosphorylation for 30 min at 72°C in 20 μl of a buffer containing 50 mM Hepes/HCl (pH 7.5), 100 mM NaCl and 25 μM [γ-32P]ATP (1 μCi, 6000 Ci/mmol) supplemented with MgCl₂ and/or MnCl₂. Radiolabelled proteins were separated by SDS/PAGE. Gels were then soaked in 16% trichloroacetic acid for 10 min at 90°C to eliminate radiolabelled nucleotides [32], before staining with Coomassie Blue and finally analysing by autoradiography. Lanes 1, 5 mM MgCl₂; lanes 2, 5 mM MnCl₂; lanes 3, 5 mM MgCl₂ and 5 mM MnCl₂. (C) Phosphotransferase activity of PAB0405, PAB1013 and PAB1047. Purified kinases (approx. 1 μg) were assayed under the same conditions as in (B), but with or without 50 μg of MBP as substrate. Lanes 1, PAB0405; lanes 2, PAB1013, lanes 3, PAB1047; lane C, MBP alone (control).
of this protein as OSGEP might have been misleading [27]. In the absence of biochemical results, we turned to genome context analysis, hoping to obtain clues about Kae1 activity. We noticed that the kae1 gene is frequently located close to genes encoding proteins involved in nucleotide metabolism and DNA repair in archaeal genomes. We thus tested whether Pa-Kae1 could interact with DNA. Our first observation was that Kae1 greatly disturbed the electrophoretic migration of a plasmid on an agarose gel (Figure 3A) and that this disturbed migration was abolished after protease treatment to remove Kae1. It was the first time that we observed such a dramatic effect with any one of the DNA-binding proteins studied in our laboratory. Using this simple assay, we found that Kae1 binds both single- and double-stranded DNA. We then tested the effect of ATP on DNA binding since Kae1 is an ATP-binding protein. To our surprise, DNA binding was completely inhibited at high ATP concentrations.

The DNA-binding properties of Kae1 were surprising, since there is no positively charged region at the surface of the Kae1 structure that could be obvious site of binding. However, we could confirm DNA binding by gel-shift assays using short oligonucleotides as substrates and by electron microscopy analyses [23]. Analysis of Kae1–DNA complexes by electron microscopy showed that Kae1 forms a well-ordered filament that rigidifies DNA. Kae1 can completely cover and rigidify a single-stranded DNA ring such as a single-stranded ΦX174 plasmid (Figure 3B). Unlike the situation with RecA, the formation of Kae1 filament is ATP-independent (instead, the filament is destabilized by ATP) and the length of the DNA does not change upon Kae1 binding [23].

In vertebrates, the genes encoding OSGEP (the mammalian orthologues of Kae1, see Figure 2) are often located just beside the genes encoding the major mammalian apurinic endonuclease APE1 (AP endonuclease 1) (also known as APEX, Hap-1 or Ref-1). In humans and mice, both genes are transcribed in opposite direction from the same promoter (Figure 3C) [28,29]. This observation prompted us to test the effect of Pa-Kae1 on depurinated DNA. We found that Pa-Kae1 binds depurinated DNA. Moreover, after protease treatment, the DNA was cleaved, indicating that Pa-Kae1 harbours AP endonuclease activity (Figure 3D). Further analyses showed that Pa-Kae1 is a class I AP endonuclease (or AP lyase), cleaving DNA 3′ to the AP site [23]. Pa-Kae1 is the first ASKHA protein with DNA-binding and AP endonuclease activities. This finding led us to suggest reserving the term Kae1 for kinase-associated endonuclease 1 instead of kinase-associated endopeptidase 1.

Although the mechanistic characteristic showed that Kae1 is a class I endonuclease (AP-lyase), Kae1 protein does not seem to be related to the two families of AP endonucleases described previously: the class I family (AP-lyases) with the example of Endo III (endonuclease III) in E. coli and the class II family (true AP-endonucleases) whose prototypes are Exo III (exonuclease III) in E. coli and APE1 in humans, and Endo IV in E. coli and Apn1 in S. cerevisiae. These proteins are structurally all different and exhibit different catalytic properties. The AP endonuclease activities of APE1 and Exo III are catalytic and are not affected by ATP, whereas Kae1 only cleaves depurinated DNA at a high protein/DNA ratio, when the protein completely covers the DNA molecule. It is unlikely that Kae1 acts as a bona fide AP endonuclease in the cell, considering the presence of powerful AP endonuclease activities available in all cell types. However, we think that the AP endonuclease activity of Kae1 reflects the fact that this protein can recognize DNA damage. Interestingly, the genes encoding Kae1 and PRPK (either fused or adjacent in most archaeal genomes) are often located in archaeal genomes adjacent to the gene encoding the archaeal Ham1 protein, a universal protein whose function is to hydrolyse abnormal nucleotides in vitro and in vivo. This strongly suggests that Kae1 and PRPK could be involved in some housekeeping functions.

**The Kae1 protein inhibits the kinase activity of Bud32 (PRPK)**

In order to analyse the interactions between Kae1 and Bud32 (PRPK) (Figure 2), we have recently solved the crystal structure of the MJ1130 fusion protein (Mj-Kae1/Bud32) from the archaeon *Methanocaldococcus jannaschii* [30].

---

**Figure 2 | Distribution of Kae1-Bud32 in the three domains of life**

The distribution of the universal Kae1 protein and the associated Bud32 kinase couple is shown in the archaea *M. jannaschii* and *P. abyssi*, in the eukarya *S. cerevisiae* and *H. sapiens* and in the bacteria *E. coli*. Note that Kae1 and Bud32 are sometimes fused in archaea (for example, the MJ1130 protein from *M. jannaschii*) and that the kinase is absent from bacteria.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Organism</th>
<th>&quot;O-sialoglycoprotein endopeptidase&quot;</th>
<th>Kinase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archaea</td>
<td><em>P. abyssi</em></td>
<td>PAB1159</td>
<td>PAB1047</td>
</tr>
<tr>
<td></td>
<td><em>M. jannaschii</em></td>
<td>MJ1130 (N-ter domain)</td>
<td>MJ1130 (C-ter domain)</td>
</tr>
<tr>
<td>Eukarya</td>
<td><em>S. cerevisiae</em></td>
<td>Kae1 (kinase-associated endopeptidase 1)</td>
<td>Bud32</td>
</tr>
<tr>
<td></td>
<td><em>H. sapiens</em></td>
<td>OSGEP (O-sialoglycoprotein endopeptidase)</td>
<td>PRPK (p53-related protein kinase)</td>
</tr>
<tr>
<td>Bacteria</td>
<td><em>E. coli</em></td>
<td>YgjD</td>
<td>/</td>
</tr>
</tbody>
</table>

© The Authors Journal compilation ©2009 Biochemical Society Transactions
Figure 3 | Biochemical characterization of Pa-Kae1

(A) Gel-shift DNA-binding assay of Pa-Kae1. Pa-Kae1 was incubated at the indicated concentrations (0–10 μM) with 100 ng of single-stranded φX174 plasmid. Free and protein-bound DNAs were separated on an agarose gel (1%). (B) Pa-Kae1-DNA complexes analysed by annular dark-field transmission electron microscopy. Panel a: control circular single-stranded DNA (φX174) showing secondary structures. Panel b: incubation of Pa-Kae1 (300 nM) with φX174 single-stranded DNA (1.5 μM in nucleotide) at a molar ratio of 1:5 protein/nucleotides induces DNA conformational change. (C) Organization of human osgep gene and human apex gene. Human osgep and apex genes are represented with exons indicated by boxes and introns indicated by solid lines. The exons are numbered from the 5′ end of each gene. Filled and open boxes represent translated and untranslated regions respectively. (D) AP endonuclease activity of Pa-Kae1 on oligonucleotide substrates. AP endonuclease activity was tested at 65°C for 10 min on a 65-mer oligonucleotide substrate (0.2 pmol) containing adenine (lanes 1), uracil (lanes 2), abasic site (lanes 3) or THF (tetrahydrofuran) at position 21 (lanes 4). Products were analysed by denaturing acrylamide gel electrophoresis. The uncut 65-mer DNA substrate and cleaved 21- or 22-mers products are indicated on the left. Oligonucleotide containing abasic site or THF is uncut by APE1 (compare with oligonucleotides incubated without enzyme). Sizes are indicated in bp.

(Figure 4A). The structure of the Mj-Kae1 moiety is identical with that of Pa-Kae1, whereas the structure of the Mj-Bud32 moiety is the smallest protein kinase for which the structure has been solved to date. The Mj-Bud32 kinase (a shortened version of classical RIO kinases) is attached to the opposite side of the ATP-binding groove of Mj-Kae1. Interestingly, although the fusion protein was incubated with a 10-fold molar excess of p[NH]PPA, no nucleotide could be observed in the binding pocket of the kinase moiety, whereas a nucleotide was clearly bound to the active site of the Mj-Kae1.
Figure 4 | Structure and biochemical characterization of MJ1130 protein

(A) Ribbon representation of MJ1130. Mj-Kae1 is coloured in light (subdomain I) and dark (subdomain I') green and Mj-Bud32 (ATP-binding domain) in light and dark (protein/peptide substrate and initiating phosphotransfer domain) red. Bound p[NH]ppA is shown as sticks. (B) Autophosphorylation of Mj-Bud32 is inhibited by Mj-Kae1. MJ1130, Mj-Bud32 and/or Mj-Kae1 (approx. 5 μg) were subjected to phosphorylation for 30 min at 72°C in 20 μl of a buffer containing 50 mM Hepes/HCl (pH 7.5), 500 mM NaCl, 10 mM MnCl2 and 25 μM [γ-32P]ATP (1 μCi, 6000 Ci/mmol). Radiolabelled proteins were separated by SDS/PAGE. Gels were then soaked in 16% trichloroacetic acid for 10 min at 90°C to eliminate radiolabelled nucleotides [32], autoradiographed (upper panel) and finally stained with Coomassie Blue (lower panel). Lanes 1, MJ1130 alone; lanes 2, Mj-Bud32 (a) incubated with Mj-Kae1 (b); lanes 3, Mj-Kae1 alone; lanes 4, Mj-Bud32 alone.

moiety. Careful analysis of the structure suggested that the P-loop of the kinase active site of Mj-Bud32 is in an inactive conformation owing to its association with Mj-Kae1. The fusion protein indeed exhibits no detectable kinase activity, but only the low autophosphorylation activity typical of Pa-Kae1.

To analyse further the interaction between Kae1 and Bud32, we have independently expressed the two domains of the MJ1130 protein in E. coli. We observed that both domains exhibit an autophosphorylation activity (Figure 4B). Addition of the Mj-Kae1 moiety to Mj-Bud32 inhibits this autophosphorylation activity. Similarly, we have shown that the yeast Kae1 protein also inactivates the kinase activity of the yeast Bud32 in vitro [30]. The fact that overexpression of Bud32p in yeast is toxic for the cell [30] strengthens the idea that the kinase activity of Bud32/PRPK should be repressed. However, the physical linkage of Kae1 and Bud32/PRPK in several archael genomes raises the question of the role of the kinase activity in vivo. Indeed, this activity is a priori continuously repressed in these organisms, suggesting that the kinase activity is not part of the normal biological role of Bud32/PRPK or that some unknown regulatory factor(s) (including interactions with DNA and/or abnormal bases) could activate the kinase activity in vivo.

In Eukarya, the Kae1 and Bud32 proteins have recently been identified as member of a larger complex called KEOPS (for kinase, endopeptidase and other proteins of small size) or EKC (endopeptidase-like kinase chromatin-associated). These complexes are involved in telomere maintenance [31] and/or in the transcription of essential genes [26]. Other members of the KEOPS/EKC complexes, Pcc1, Pcc2 and Cgi121 in S. cerevisiae, are obvious candidates to regulate the activity of Kae1 and Bud32. We have recently detected homologues of Pcc1 and Cgi121 in archaea (P. Forterre, A. Hecker and C. Brochier-Armanet, unpublished work), suggesting that the KEOPS/EKC complex could have an ancient origin.

Conclusion

Further biochemical and biological characterizations of Kae1, Bud32 (PRPK) and other partners of the KEOPS/EKC complex should allow us to generate ideas on the role of these systems in ancient organisms (especially to determine their connection to the nature of ancestral genomes, either RNA or DNA). These studies are also essential for our understanding of the crucial role(s) of these strongly conserved systems in modern organisms. In particular, understanding the biological role of the universal Kae1 protein remains a major challenge. The importance of this protein has been completely ignored until now. The universal distribution of Kae1, its DNA-binding properties and its probable conserved role in genome maintenance are all reminiscent of the RecA proteins and its archael and eukaryotic homologues (RadA and Rad51 respectively). Amazingly, there are, at present, only about ten published papers on Medline which mention the Kae1 (OSGEP, YgjD) proteins, whereas more than 6000 references can be retrieved using RecA and Rad51 as key
words. There is also a striking discrepancy between the huge number of laboratories working on p53 (46938 publications in Medline) and the very low number of publications on PRPK and Bud32 (fewer than 12). Once more, the study of Archaea, together with comparative genomic analyses, has been critical in revealing a fundamental biological mechanism that had previously escaped the attention of biologists.

**Funding**

This work was supported by the Centre National de la Recherche Scientifique, the Human Frontier Science Program, the Association pour la Recherche sur la Cancer, the Agence Nationale de la Recherche and the European 3D-Repertoire programme (grant number LSHP-CT-2005-512028). A.H. was supported by the Human Frontier Science Program and the Agence Nationale de la Recherche.

**References**