REVIEW ARTICLE

Archaeal protein kinases and protein phosphatases: insights from genomics and biochemistry

Peter J. KENNELLY¹

Department of Biochemistry – 0308, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, U.S.A.

Protein phosphorylation/dephosphorylation has long been considered a recent addition to Nature's regulatory arsenal. Early studies indicated that this molecular regulatory mechanism existed only in higher eukaryotes, suggesting that protein phosphorylation/dephosphorylation had emerged to meet the particular signal-transduction requirements of multicellular organisms. Although it has since become apparent that simple eukaryotes and even bacteria are sites of protein phosphorylation/dephosphorylation, the perception widely persists that this molecular regulatory mechanism emerged late in evolution, i.e. after the divergence of the contemporary phylogenetic domains. Only highly developed cells, it was reasoned, could afford the high 'overhead' costs inherent in the acquisition of dedicated protein kinases and protein phosphatases. The advent of genome sequencing has provided an opportunity to exploit Nature's phylogenetic diversity as a vehicle for critically examining this hypothesis. In tracing the origins and evolution of

protein phosphorylation/dephosphorylation, the members of the Archaea, the so-called 'third domain of life', will play a critical role. Whereas several studies have demonstrated that archaeal proteins are subject to modification by covalent phosphorylation, relatively little is known concerning the identities of the proteins affected, the impact on their functional properties, or the enzymes that catalyse these events. However, examination of several archaeal genomes has revealed the widespread presence of several ostensibly 'eukaryotic' and 'bacterial' protein kinase and protein phosphatase paradigms. Similar findings of 'phylogenetic trespass' in members of the Eucarya (eukaryotes) and the Bacteria suggest that this versatile molecular regulatory mechanism emerged at an unexpectedly early point in development of 'life as we know it'.

Key words: Archaea, Archaeobacteria, phosphorylation/dephosphorylation.

MOLECULAR REGULATION AND 'LIFE AS WE KNOW IT'

For many years, scientists have intuitively grasped the inextricable link between molecular regulatory processes and 'life as we know it', i.e. the emergence of multicellular organisms whose capabilities greatly exceed the sum of their parts. The execution of the developmental programme responsible for directing the spatially and temporally correct differentiation of the vast spectrum of highly specialized cell types that comprise animals, plants and other higher organisms is critically dependent upon the broadcast, receipt and interpretation of a multiplicity of internal and external cues. The concomitant integration of these communities of cells into unified and intimately co-ordinated wholes is no less dependent upon the existence of extensive signal-transduction networks linking neuronal, hormonal and other intercellular messengers to their ultimate molecular targets inside the cell.

Molecular regulatory networks are not the exclusive province of 'higher' organisms, however. Recent studies have revealed that many bacteria possess signal-transduction circuits whose number, diversity and complexity belies their 'simple' organismal stereotype (reviewed in [1–5]). Upon reflection, the regulatory sophistication displayed by 'lower' organisms is not so surprising. Virtually all cells face the same fundamental challenges, whether they exist as autonomous entities or as members of an intimate, interdependent collective. Prominent among these is the need to

detect and appropriately react to an oftentimes heterogeneous population of surrounding cells that may include organismal partners, predators, prey, hosts, parasites, symbionts, etc.

MOLECULAR REGULATION AND ANCIENT LIFE

The first primitive micro-organisms, living in isolated niches on a nascent biotic planet, faced neither the pressures nor the opportunities presented by biological diversity. Thus it would appear logical to assume that the attainment of a 'critical mass' of basic functional attributes [6] constituted an absolute prerequisite for the development of dedicated molecular regulatory mechanisms. Only a fairly complex organism would possess a sufficiently diverse and robust molecular infrastructure to support the development of such mechanisms, as well as the potential for deriving significant benefits therefrom. Moreover, competition from other species, an obvious source of selective pressure for the development of sensor–response mechanisms and their associated molecular regulatory machinery, could not have arisen until primitive organisms had attained sufficient complexity to enable them to diversify from one another. Until this threshold had been reached, primitive organisms presumably were better served in the struggle for continuance by employing scarce internal assets and limited external resources to: (i) develop dedicated information-storage molecules; (ii) diversify their metabolism in order to access a wider spectrum of nutrient sources and to utilize them with greater efficiency; (iii) construct a protective membrane

Abbreviations used: LMM PTP, low-molecular-mass PTP [also known as low-molecular-weight (LMW) PTP]; MAP, methyltransferase activation protein; ORF, open reading frame; PTP, protein tyrosine phosphatase; cPTP, conventional PTP; Rap, regulator aspartyl phosphatase.

¹ e-mail pjkenne@vt.edu

barrier; and (iv) devise a means for storing nutrients in anticipation of a 'rainy day'.

The expectation that molecular regulatory mechanisms appeared as a refining tertiary attribute of life forms that had evolved into crude, but quite recognizable, versions of contemporary microbes [6] ignored the fact that the emergence of life, by its very nature, created an immediate market for co-ordination and control. Although the debate continues as to the ultimate nature of the first living organisms, e.g. metabolist (a set of self-sustaining autotrophic reactions) versus prebiotic polymers (e.g. RNA world) (reviewed in [7]), general agreement exists as to the quality that separated dynamic, but abiotic, entities from living ones. That quality is the ability to replicate: to synthesize physically and functionally autonomous copies of oneself. In a world of limited resources, however, replication constitutes a double-edged sword. Although the ability to create multiple copies of oneself may enhance the statistical probability that any one representative of the species will survive a future calamity, unchecked replication renders at least one such calamity inevitable: overpopulation. The survival-enhancing benefits of replication will be compromised further if, when one or more critical precursors are scarce, continued attempts to replicate yield a high proportion of incomplete or otherwise defective progeny, a not unlikely occurrence prior to the advent of molecular proof-reading mechanisms. Thus, as the expanding demands of unmoderated replication inevitably outstripped the supply of nutrients, the stresses of starvation would have been exacerbated by the wasteful consumption of those resources via the synthesis of dysfunctional progeny.

Replicative 'fail-safe' mechanisms thus represented an important attribute in the development of sustainable life forms. The acquisition of some means, however crude, for balancing the maintenance needs of the existing population with the opportunities generated via replication would have conferred a substantial survival benefit upon any life form so endowed. Thus molecular regulatory mechanisms may have emerged at a surprisingly early point in the evolutionary process. Presumably, the first such mechanisms would have taken the form of simple refinements to existing catalysts. For example, a crude form of feedback inhibition could be established by sufficiently increasing the affinity of an enzyme for its reaction product(s), while resource thresholds for certain processes could be established through the formation of multimeric enzymes that bound substrates in a co-operative manner. The limits of the active site, however, invariably constrained the range of potential regulators to substrates, products and their close structural relatives. Such mechanisms therefore were essentially local in scope and passive in nature, as the range of 'signals' to which they could respond was inevitably derived from the immediate process in which each enzyme participated.

Eventually, however, specialized mechanisms and dedicated enzymes emerged that enabled molecular regulation to escape the limitations imposed by an enzyme's active site. Allosterism and covalent modification represented a quantum leap in regulatory capability and sophistication as they: (i) expanded the range of potential regulatory variables beyond an enzyme's catalytic reactants and structural analogues thereof; (ii) enabled 'active' intervention into molecular processes independent of reactant levels; and (iii) opened avenues for the co-ordinate regulation of enzymes whose reactants shared no obvious structural similarity.

Not only did the emergence of allosterism and covalent modification make it possible to co-ordinate diverse metabolic processes involving radically different substrates, these 'active' regulatory mechanisms provided the means for developing signal-

transduction cascades capable of responding to membrane-impermeant factors impinging upon the exterior of the cell. Indeed, Monod, quoted by Perutz in [8], referred to his newly formulated concept of allosteric regulation as the second secret of life, the first being DNA.

PROTEIN PHOSPHORYLATION/DEPHOSPHORYLATION: A VERSATILE MOLECULAR REGULATORY MECHANISM

The virtues of the phosphoryl group as a protein modulator

The direct, active regulation of biological macromolecules is accomplished via two basic types of mechanisms: allosteric regulation and covalent modification. Some might argue that gene expression, the processing, translation, and degradation of mRNA, and protein degradation also should be ranked in this same category. However, molecular regulation via these 'mechanisms' is derivative in Nature, because the modulation of each can be traced ultimately to the allosteric and/or covalent regulation of the macromolecules that transcribe, process, translate or degrade.

Whether measured in terms of the sheer number or diversity of its targets, or its potential for creating highly sophisticated regulatory networks, protein phosphorylation/dephosphorylation ranks as Nature's premier mechanism for regulating the functional properties of proteins. The reasons for this are several. The high charge density, extreme hydrophilicity and strong propensity to form salt bridges with the guanidino moiety of arginine render the phosphoryl group an extremely potent agent for perturbing protein structure [9,10]. Coupled with the low structural demands involved in the creation of a phosphorylation site (i.e. a nucleophilic oxygen or nitrogen atom within an amino acid side chain surrounded by a few local recognition features [11]), phosphorylation imposes minimal structural demands upon the protein to be regulated. Mutagenic alteration of a single amino acid residue in firefly luciferase, for example, proved sufficient not only to create a recognition site for the cAMP-dependent protein kinase, but to confer phosphorylation-mediated control on the enzyme's catalytic activity *in vitro* [12]. Therefore, when considered from the perspective of engineering potential protein targets, phosphorylation offers considerable advantages relative to the construction of large and conformationally complex binding domains for allosteric ligands, and the associated means for transmitting an effector binding event to its ultimate site of action (reviewed in [13,14]).

Protein phosphorylation also offers exceptional potential as a mechanism for integrating multiple signals. The structural simplicity of a 'typical' phosphorylation site enables many proteins to accommodate multiple sites within their structure. Interactions among these sites provide one means for creating logic gates capable of processing environmental signals [15]. For a protein with two sites of phosphorylation, for example, an 'OR' gate is created if phosphorylation of either site is sufficient to provoke a functional change. If both sites must be phosphorylated for a functional change to occur, an 'AND' gate will be formed. If phosphorylation at one site blocks phosphorylation at a second, function-altering site, a 'NAND' gate will result. Another avenue for signal integration is provided by the overlapping substrate-recognition requirements of many protein kinases [11], which frequently enables a single phosphorylation site to be recognized by multiple protein kinases, each of which may respond to a different environmental signal.

The ability of phosphoproteins to act as loci for information processing has been greatly enhanced by the 'reversibility' of protein phosphorylation. The strong thermodynamic driving force and mechanistic facility of dephosphorylation, a simple

hydrolytic reaction catalysed by protein phosphatases, enables phosphoproteins to be rapidly restored to their original physical and functional states. In addition to providing the means for resetting these molecular switches to their basal 'ready' state (sometimes referred to as signal termination), protein phosphatases serve as another conduit for conveying sensory information to the phosphoprotein 'decision nodes' within a cell's information-processing machinery [16,17]. Moreover, because protein phosphatases act in an antagonistic fashion to protein kinases, the interplay between these enzymes determines not only the direction in which the phosphorylation state of a protein may change, but the rate, magnitude and duration of that change as well, thus adding a temporal dimension to the signal-response process.

Protein phosphorylation/dephosphorylation: an evolutionarily recent acquisition?

Despite its obvious potential as a mechanism for both the simple switching of proteins between functionally distinct states and the construction of sophisticated organic computers for multivalent signal processing, protein phosphorylation/dephosphorylation was long stereotyped as a late addition to Nature's molecular regulatory arsenal. Initial attempts to detect protein phosphorylation in microbial organisms were frustrated by the unsuitability of the tools available, which were derived from early studies on mammalian systems (reviewed in [18]). On the other hand, during this same period of time numerous examples of elegant allosteric feedback mechanisms were being discovered in these same bacterial organisms (reviewed in [19–21]). The impression soon formed that allosteric regulation represented the *primaevae* active regulatory mechanism (reviewed in [22]). It was further reasoned that, although allosterism offered sufficient scope for development to meet the relatively straightforward regulatory requirements of unicellular life forms, it was inadequate to tackle the more complex demands imposed by the emergence of organisms comprising multiple differentiated cells. Thus the appearance of multicellular organisms and their associated neuroendocrine systems was assumed for a long time to provide the *raison d'être* for the emergence of protein phosphorylation/dephosphorylation, one that readily accommodated its apparent exclusivity to higher life forms.

In the 1980s, research on protein phosphorylation/dephosphorylation finally broke through the domain barrier. The initial 'crack' was opened by the discovery that, in enteric bacteria, isocitrate dehydrogenase was controlled by phosphorylation/dephosphorylation of a serine residue (reviewed in [23]). This crack was widened into a clear breach with the description of a modular signal-transduction paradigm called the two-component system. Two-component systems not only pervaded the members of the bacterial domain, but participated in the control of a multiplicity of cellular processes, ranging from chemotaxis to nitrogen metabolism (reviewed in [24,25]). Radical differences, however, were apparent in both structure and catalytic mechanism between the so-called 'eukaryotic' protein kinases that dominated the protein-phosphorylation processes of higher organisms and both the two-component histidine kinases and the bifunctional isocitrate dehydrogenase kinase/phosphatase found in Bacteria. The two-component and isocitrate dehydrogenase kinase/phosphatase systems soon were viewed as uniquely bacterial means for exploiting the superior biophysical properties of the phosphoryl group, ones that had emerged long after the eukaryotic and bacterial organisms had diverged from one another (reviewed in [1,4]). Hence their discovery tended to reinforce, rather than challenge, the basic perception that protein

phosphorylation/dephosphorylation constituted a relatively new regulatory mechanism.

A theoretical argument often mustered in support of the 'allosterism ancient/phosphorylation young' chronology was based upon comparisons of the crucial steps required in their development. The critical event in the establishment of allosteric regulation was the acquisition by the target protein of a ligand-binding domain suitable for transformation into a specialized sensory/regulatory domain. This would appear to be a relatively straightforward process, given that the genetic material already on hand encoded macromolecules that bound a wide range of substrates, products, cofactors, etc. Simple gene fusion and duplication events could readily produce proteins containing proto-allosteric domains (reviewed in [26]). On the other hand, protein phosphorylation/dephosphorylation required the development of two types of specialized regulatory enzymes, i.e. protein kinases and protein phosphatases. The development of both enzymes was critical if phosphorylation was to successfully compete against either (i) allosteric regulation, because the binding of dissociable ligands is, by its very nature, (ii) reversible, or an alternative but irreversible covalent modification process, i.e. the partial proteolysis of proenzymes.

The high initial 'overhead' costs associated with protein phosphorylation/dephosphorylation thus appeared to preclude its emergence until living organisms had exhausted the potential offered by other, less 'capital-intensive', molecular regulatory mechanisms. Therefore it was not until relatively recently that systematic efforts were mounted to exploit phylogenetically diverse organisms as a means for tracing the origins and development of protein phosphorylation/dephosphorylation. With the discovery in bacterial organisms of protein kinases and protein phosphatases traditionally associated with eukaryotes, and vice versa, long-held views concerning the history of this versatile regulatory mechanism have been changing (reviewed in [1]). In the sections below are summarized the first steps in the exploration of protein phosphorylation/dephosphorylation in the members of the third phylogenetic domain, the Archaea.

THE ARCHAEA : THE THIRD DOMAIN OF LIFE?

Nature's 'time machine'

Given the importance of molecular regulatory mechanisms in the development and smooth functioning of the complex organisms that we encounter during our daily existence, e.g. plants and animals, it is only natural to ask: when did active molecular regulatory mechanisms emerge? What form(s) did they take, and what functions did they target? Phylogenetic analysis offers a powerful tool for approaching these questions, as clues to their ancestry and history reside within the genome of every organism. However, the complexities of the evolutionary process frequently render phylogeny-based 'molecular archaeology' a challenging task.

Phylogeny: from morphology to molecules

Classic models of evolution postulated the existence of two distinct phylogenetic domains, commonly referred to as the eukaryotes and prokaryotes [27]. Under this paradigm, all contemporary life forms traced their origins to the appearance of a cellular organism that proved sufficiently robust and adaptable to endure, often referred to as the universal ancestor or last common ancestor. At some point thereafter, a physical or molecular event occurred that segregated one of the descendants of this universal ancestor from its 'brethren', enabling each to embark upon a unique path of development. From this ancient

Figure 1 Modern models of phylogeny

Shown on the left is a simplified schematic representation of the three-domain tree of Woese [29]. The hatched arrows illustrate the endosymbiotic events that gave rise to mitochondria and chloroplasts in eukaryotic organisms. Shown on the right is a simplified schematic representation of Doolittle's phylogenetic 'bush' [33], in which the broken arrows illustrate the vigorous 'commerce' in lateral-gene exchange, including endosymbiotic events, that have produced the apparent phylogenetic scrambling evident from whole genome sequences. This representation also incorporates the concept of a genetically interactive ancestral community or collective [155], as opposed to a clonal universal ancestor, as the root of the contemporary tree of life. Note that the line segments are not drawn to scale, and thus provide no quantitative indication of the degree of divergence between specific domains or kingdoms. AC, ancestral community; An, animals; Cy, cyanobacteria; F, fungi; G[−], Gram-negative bacteria; G⁺, Gram-positive bacteria; Ha, halophiles; Me, methanogens; Pl, plants; Pr, protists; Th, thermophiles; UA, universal ancestor.

bifurcation arose the prokaryotes and eukaryotes. The former were constrained by their retention of the gross characteristics of the universal ancestor, i.e. a single circular chromosome, an uncompartimentalized cell interior, and unicellular nature. The progenitors of the eukaryotes, on the other hand, unlocked the secrets of specialization at the subcellular level through the acquisition of prokaryotic endosymbionts (e.g. the predecessors of the mitochondria and chloroplasts) and the construction of other functionally specialized organelles, most notably the nucleus (reviewed in [28]). Many eukaryotes extended the process of functional compartmentation and specialization to the cellular level, leading to the formation of multicellular organisms that comprised a variety of differentiated cell types.

It is important to note that the original basis of the classic two-domain model was morphological, not genetic, in nature [27]. Although this represented a reasonable way for classifying those physically complex organisms that exhibited a wide range of primary and secondary structural characteristics to compare and contrast, these criteria proved difficult to apply to physically simple micro-organisms imbued with relatively few discernible morphological features. Thus the organisms comprising the classic prokaryotic domain were grouped together by default, not because they resembled one another so much as because they did not resemble eukaryotes.

With the advent of molecular biology, scientists procured the ability to directly and quantitatively analyse heredity by sequencing the genes by which it is encoded. Comparisons of the sequences of rRNAs, ubiquitous macromolecules participating in an essential cellular process, largely confirmed the coherence and organization of the morphologically defined eukaryotic domain, now sometimes referred to as the Eucarya. However, classification on the basis of rRNA sequences demolished the notion that the prokaryotes constituted a unified monophyletic domain. Instead, Olsen and Woese [29] proposed that three phylogenetic domains existed: the Eucarya and two distinct domains comprising 'prokaryotic' organisms, the Bacteria (originally referred to as the Eubacteria) and the Archaea (originally referred to as the Archaeobacteria). The Bacteria included most of the prokaryotic organisms commonly encountered in a typical course in microbiology: *Escherichia coli*,

Anabaena, *Bacillus subtilis*, *Pseudomonas aeruginosa*, *Salmonella typhimurium*, etc. The Archaea, on the other hand, were rich in extremophilic organisms that thrived in 'hostile' environmental niches, e.g. thermophiles, anaerobes and halophiles. Many archaeons also were imbued with exotic metabolic capabilities, e.g. methanogenesis and sulphur oxidation. Their extremophilic nature suggested to some investigators that the members of this 'third domain of life' constituted a set of 'living fossils', direct descendants of the organisms that once had dominated the volcanic, anaerobic, environment of primaevial Earth [30].

The shifting branches of the three-domain tree

Perhaps the most surprising outcome of Woese's analysis was the placement of the Archaea on the same major branch of the phylogenetic tree as the Eucarya, rather than with the superficially similar Bacteria (Figure 1, left panel). In other words, it was proposed that the first bifurcation in the evolution of contemporary organisms segregated the Bacteria away from an archaeal/eucaryal line of descent. As 'first cousins' of the eukaryotes, it was therefore expected that clues as to the nature of microbial proto-eukaryotes still resided within the Archaea. However, as scrutiny of the Archaea intensified, an increasing number of 'bacterial' features were discerned that called into question the position of the archaeal branch of the phylogenetic tree. While phylogenetic analysis using molecular markers other than rRNAs frequently led to a recapitulation of the original Woesian tree, other markers indicated that (i) the Archaea and Bacteria resided on the same major branch of the phylogenetic tree, or that (ii) the positions of the Archaea and Bacteria *vis-à-vis* the Eucarya should be reversed, or that (iii) the Eucarya were chimaeric organisms that had suddenly emerged through the fusion of an archaeon with a bacterium (reviewed in [31]).

Kandler [32] summarized the paradoxical relationship among the members of the three phylogenetic domains in the form of a triangle illustrating the manner in which each domain both resembled and differed from its fellows to similar degrees with regard to many fundamental features, providing no clear evidence of a hierarchy (Figure 2). With the advent of genomics, it became possible to analyse phylogenetic relationships on a global level

Figure 2 Kandler's triangle

Kandler [32] devised the triangular Figure shown here to represent the lack of a clear hierarchical relationship among the three phylogenetic domains delineated by the sequences of their rRNA: the Archaea, Bacteria and Eucarya. Each apex of the triangle represents one of the domains, as indicated. Major molecular features shared by two of the three domains are indicated next to the line segment joining the two domains in which they are found. For example, the Bacteria and Eucarya each contain lipids containing fatty acid side chains joined to glycerol head groups via ester linkages (ester lipids), whereas archaeal lipids contain polyisoprenoid side chains joined to their head groups by ether linkages. The Archaea and Eucarya share homologous transcription factors, such as EF-Tu (elong factors), not found in Bacteria. The Bacteria and Archaea each possess closed-circular genomes (circular DNA) in which genes for functionally related proteins are organized into spatially contiguous transcriptional units (operons), whereas the Eucarya do not. The Bacteria and Eucarya share a homologous family of phosphofructokinases that generally utilize ATP as a phospho-donor substrate (ATP-PFK), whereas then-known archaeal phosphofructokinases were members of a distinct protein family that generally utilized ADP as the phospho-donor substrate. Similarly, the Archaea and Eucarya share a so-called vacuolar-type H^+ -ATPase (V-ATPase) not found in the Bacteria.

rather than relying on a single, arbitrarily selected, molecular marker. These examinations revealed an extensive history of vigorous traffic in lateral gene transfer within, between and among the members of the Woesean phylogenetic domains, which has blurred traditional distinctions. The unexpected genetic permeability of domain boundaries has transformed what was a formerly orderly and linear phylogenetic tree into a tangled and interconnected web that W. F. Doolittle [33] has likened to a phylogenetic 'bush' (Figure 1, right panel).

As can be seen from this brief outline, many fundamental concepts regarding phylogeny remain in a state of flux. The predominantly linear and divergent process of molecular evolution originally inferred from Darwin's seminal studies has been revealed to be a heterogeneous and pluralistic phenomenon that has given rise to a multidimensional phylogenetic matrix. However, the evidence accumulated to date largely supports the robustness of the bacterial, archaeal and eucaryal domains as a scaffold for understanding phylogenetic/evolutionary relationships [34]. Despite the unexpected plasticity of major portions of the genome, the comparison of phylogenetically diverse organisms thus remains the nearest available equivalent to a biological 'time machine' for the exploration of fundamental questions concerning the origins and evolution of the molecular processes that take place inside cells.

PROTEIN PHOSPHORYLATION IN THE ARCHAEA

Studies of global protein phosphorylation using radiolabelled phosphate

Protein phosphorylation was first reported in a member of the 'third domain of life', the extreme halophilic archaeon *Halo-*

bacterium halobium (since renamed *H. salinarium*) in 1980 [35]. *H. salinarium* is a photosensitive organism that uses light-driven ion pumps to generate cellular energy. Stimulated by Wang and Koshland's [36] report of protein kinase activity in the bacterium *S. typhimurium*, Spudich and co-workers [35] analysed the spectrum of phosphorylated polypeptides that could be extracted from *H. salinarium* that had been cultured in the presence of [32 P] P_i . A group of phosphorylated polypeptides (approx. 10) were evident on autoradiograms of SDS/polyacrylamide gels. Three of these were present only in cells that had been cultured in the absence of light, namely a membrane-associated species with an M_r of approx. 110 000, and two soluble species with M_r values of approx. 83 000 and 62 000. Efforts to remove the phosphoryl groups from the two larger polypeptides using acid or hydroxylamine proved unavailing, indicating that they were linked to the proteins by ester bonds: a finding consistent with phosphorylation of the hydroxy side chains of serine, threonine and/or tyrosine residues [35]. The effect of these treatments on the other phosphoproteins was not reported.

Intriguingly, when illumination was discontinued for cells that had been grown in the light, phosphorylation of the polypeptides with approximate sizes of 62, 83 and 110 kDa became evident within 30 s and reached maximal levels within 5 min [35,37]. On the other hand, the radiolabelled phosphate disappeared from these polypeptides with comparable rapidity when cells were switched from darkness to light. The speed with which these changes occurred, coupled with the lack of any noticeable differences in the polypeptide population visualized using Coomassie Blue, indicated that light was somehow modulating the degree of phosphorylation, rather than the abundance, of the polypeptides approximately sized 62, 83 and 110 kDa. Further analyses revealed that the dephosphorylation of these proteins was mediated by the collapse of the pH gradient established by the light-driven proton-pumping action of bacteriorhodopsin [37].

In 1984, Skorko [38] established the existence of protein phosphorylation in a second member of the Archaea, the extreme acidothermophile *Sulfolobus acidocaldarius*. This organism is encountered in volcanic hot springs, where it oxidizes elemental sulphur. SDS/PAGE revealed the presence of several prominently visible phosphopolypeptides with approximate M_r values of 31 500, 42 000, 52 000 and 71 000 in cells grown in the presence of [32 P] P_i . Association of 32 P radioactivity with these polypeptides was labile to incubation with alkali or alkaline phosphatase, but not with snake-venom phosphodiesterase, consistent with linkage via a phosphomonoester bond. When [γ - 32 P]ATP was added to cell extracts of *S. acidocaldarius*, an even greater number of phosphorylated polypeptides was evident. Phosphoamino acid analysis revealed that a significant fraction of the radioactive phosphate was bound to serine and threonine residues, suggesting that at least some of these polypeptides had been phosphorylated through the action of a protein kinase(s). The gross protein kinase activity in extracts appeared to markedly increase when cells entered stationary phase. A subsequent study revealed that a polypeptide of approximate size 40 kDa became phosphorylated on threonine when an isolated ribosomal fraction from *S. acidocaldarius* was incubated with [32 P]polyphosphate [39].

At least some of the protein phosphorylation events that take place in *S. acidocaldarius* exhibited the dynamism suggestive of regulatory potential. Using two-dimensional electrophoresis, Osorio and Jerez [40] observed the presence of more than 20 32 P-labelled proteins in cells grown in the presence of [32 P] P_i . One of these, a polypeptide with an M_r of approx. 36 000, was observed to undergo a dramatic increase in its apparent stoichiometry of phosphorylation when cultures of *S. acidocaldarius* were trans-

ferred to media containing growth-limiting concentrations of phosphate.

Studies of global protein phosphorylation using antibodies against phosphotyrosine

In 1997, the range of archaeons in which protein phosphorylation had been detected was extended further to include a second extreme acidothermophile (*Sulfolobus solfataricus*), a second extreme halophile (*Haloferax volcanii*) and an anaerobic methanogen (*Methanosarcina thermophila* TM-1) [41]. These studies employed phosphoamino acid-directed antibodies to provide the first direct evidence for the presence of phosphotyrosine in archaeal proteins. More recently, three tyrosine-phosphorylated polypeptides were isolated and partially sequenced (see below) from lysates of the hyperthermophilic archaeon *Thermococcus kodakaraensis* [42]. These findings were particularly noteworthy inasmuch as many investigators maintain that protein tyrosine phosphorylation is the exclusive province of higher eukaryotes [43].

Identification of individual phosphoproteins

Although the studies described above clearly established that proteins within a broad spectrum of archaeons were subject to covalent phosphorylation, frustratingly little progress has been made in ascertaining precisely which archaeal proteins and, by inference, what cellular processes are targeted by this covalent modification process. Unfortunately, only a small handful of phosphoproteins have been identified to date, including several 'red herrings' representing trapped phosphoenzyme intermediates rather than the products of regulatory protein kinases ([44], and M. B. Potter and P. J. Kennelly, unpublished work).

The first archaeal phosphoprotein of any type to be identified was the methyltransferase activation protein (MAP), from the methanogenic archaeon *Methanosarcina barkeri* [45]. MAP activates a key enzyme in the metabolic pathway by which *M. barkeri* transforms carbon dioxide to methane, methanol:5-hydroxybenzimidazolylcobamide methyltransferase, also known as MT1. Activation of MT1 by MAP was dependent upon ATP or other nucleoside triphosphates *in vitro*. However, non-hydrolysable ATP analogues, such as adenosine 5'-[α,β -methylene]triphosphate ('AMPPCP') or adenosine 5'-[β,γ -imido]triphosphate ('AMPPNP') proved to be ineffective as ATP substitutes. Incubation of purified MAP with [γ - 32 P]ATP revealed that it could phosphorylate itself to near-stoichiometric levels and that, in the absence of added ATP, the autophosphorylated protein activated equimolar quantities of MT1. However, substoichiometric quantities of MAP completely activated MT1 when exogenous ATP was present, suggesting that the former protein acts either by transferring its phosphoryl group to MT1 or by evoking a stable physical or conformational change in MT1 via a mechanism involving the hydrolysis of the enzyme-bound phosphoryl group.

In 1998, Condo et al. [46] reported that an aminopeptidase of approximate size 90 kDa from *S. solfataricus* became phosphorylated when cell lysates were incubated with [γ - 32 P]ATP. This protein was first identified by virtue of its tight association with the 60 kDa chaperonin in this archaeon. When the chaperonin-aminopeptidase complex was immunoprecipitated before incubation with [γ - 32 P]ATP, no phosphorylation was observed, suggesting that the aminopeptidase was targeted by an exogenous protein kinase present in the lysate. Prior incubation of cell lysates with ATP had no effect on the quantity of aminopeptidase activity recovered in immunoprecipitates, however; nor was the

chemical nature of the bond linking the phosphoryl group to the protein determined.

More recently, it has been reported that the Cdc6 proteins from *Methanobacterium thermoautotrophicum* and *Pyrobaculum aerophilum* could be phosphorylated by incubating them with [γ - 32 P]ATP, but not [α - 32 P]ATP, *in vitro* [47]. Phosphorylation took place on a serine residue and appeared to be self-catalysed, because variants in which the predicted Walker-A nucleotide-binding motif had been mutagenically altered did not undergo phosphorylation. By analogy to its homologue from fission yeast, Cdc18, archaeal Cdc6 was postulated to act as an initiator of replication that binds DNA either individually or as part of a multi-protein complex. Autophosphorylation of the members of this protein family appeared to be highly conserved, and therefore functionally important, because yeast Cdc18 also underwent autophosphorylation on serine [47]. Intriguingly, both single- and double-stranded DNA inhibited the ability of Cdc6 to autophosphorylate, presumably via binding to a predicted winged-helix domain near the protein's C-terminus.

Jeon et al. [42] extracted three tyrosine-phosphorylated polypeptides from a lysate of the hyperthermophile *T. kodakaraensis* using a substrate-trapping mutant of a potential protein tyrosine phosphatase (PTP), *Tk*-PTP. The N-terminal sequences of these proteins were determined by the Edman method. Since the genome sequence of *T. kodakaraensis* has yet to be determined, these sequences were used to search the genomes of three other hyperthermophilic archaeons, *Pyrococcus abyssi*, *P. furiosus* and *P. horikoshii*, for potential proteins possessing similar N-termini. On the basis of these comparisons, the three tyrosine-phosphorylated polypeptides were tentatively identified as the β -chain of a phenylalanyl-tRNA synthetase, RtcB, a conserved protein of unknown function whose gene may be co-transcribed with that of an RNA 3'-terminal phosphate cyclase in *E. coli*, and a phosphomannomutase. The apparent M_r values of these phosphotyrosyl polypeptides were approx. 64000, 53000 and 27000 respectively.

In our own laboratory, we have observed that incubation of a partially purified membrane extract from the archaeon *S. solfataricus* with [γ - 32 P]ATP resulted in the appearance of approximately a dozen phosphorylated polypeptides on two-dimensional gels (B. H. Lower and P. J. Kennelly, unpublished work). A virtually identical pattern of phosphorylation was observed when [γ - 32 P]ATP was replaced with [γ - 32 P]GTP. One of these polypeptides, which was phosphorylated on a threonine residue, was identified by mass peptide profiling as the product of open reading frame (ORF) *ss0469*. A recombinantly produced form of this protein, named SsoPK3, exhibited protein serine kinase activity (see below). Recombinant SsoPK3 displayed neither the propensity to self-phosphorylate nor the ability to utilize GTP as an alternative phospho-donor substrate, suggesting that it had been phosphorylated by the threonine-preferring, ATP/GTP-utilizing protein kinase activity detected previously in the partially purified membrane fraction [48].

ARCHAEOAL PROTEIN KINASES AND PROTEIN PHOSPHATASES: A GENOMIC OVERVIEW

Introduction

No codons exist for the modified amino acids phosphoserine, phosphotyrosine, phosphohistidine, etc. Nor, with a few notable exceptions such as the response regulator domains that constitute the cognate substrates of two-component histidine kinases, do protein-phosphorylation sites display sequence features that are both sufficiently unique and consistently conserved to permit the reliable identification of phosphoproteins from genome sequence

Table 1 Summary of ORFs encoding known or potential protein kinases, protein kinase/phosphatases or protein phosphatases in published archaeal genomes

The Table below summarizes the number of ORFs in published archaeal genome sequences that may potentially encode members of the established protein kinase, protein kinase/phosphatase and protein phosphatase paradigms listed. Sources utilized to compile this Table include the published genomes and their associated websites [49–57], The Institute for Genome Research's Comprehensive Microbial Resource [153] and selected reviews [4,58,59,77,82,86,87]. Where sources disagree on the number of potential representatives of a particular enzyme paradigm, the higher estimate is presented. The numbers in parentheses indicate the number of those ORFs whose products have been demonstrated to exhibit the functional capabilities inferred from their DNA-derived amino acid sequence. Abbreviations: AceK, the isocitrate dehydrogenase/kinase phosphatase; Cdc25, the Cdc25 dual-specificity protein phosphatases; CheZ, the putative chemotactic response regulator phosphatase; ePK, eukaryotic protein kinases; HisK, histidine kinases and their derivative protein serine/threonine kinases; HPrK, the HPr kinase/phosphatase; MHCK, myosin II heavy chain kinase; PPM, PPM-family protein phosphatases; PPP, PPP-family protein phosphatases.

Archaeon	Genome (Mbp)	Protein kinases		Kinase/phosphatases		Protein phosphatases						
		HisK	ePK	MHCK	AceK	HPrK	CheZ	PPP	PPM	cPTP	LMM PTP	Cdc25
<i>A. pernix</i> K1	1.67	—	1	—	—	—	—	1	—	—	—	—
<i>A. fulgidus</i>	2.18	15	4	—	—	—	—	1	—	—	1	—
<i>Halobacterium</i> NRC1	2.57	14 (1)	—	—	—	—	—	—	—	—	—	—
<i>M. thermoautotrophicum</i>	1.75	16	3	—	—	—	—	—	—	1	1	—
<i>M. jannaschii</i>	1.66	—	4	—	—	—	—	—	—	3	—	—
<i>P. horikoshii</i> OT3	1.80	1	4	—	—	—	—	—	—	1	—	—
<i>S. solfataricus</i> P2	2.99	—	8 (2)	—	—	—	—	1 (1)	—	1	—	—
<i>T. acidophilum</i>	1.56	—	—	—	—	—	—	—	—	—	—	—
<i>T. volcanium</i>	1.58	—	3	—	—	—	—	1	1	—	—	—

data. However, this is not the case with regard to many of the potential protein kinases and protein phosphatases that may catalyse protein phosphorylation/dephosphorylation events in the Archaea. Decades of intensive research involving both eukaryotic and bacterial organisms have led to the molecular characterization of three distinct families of protein kinases, two bifunctional protein kinase/phosphatase paradigms, five families of protein serine/threonine/tyrosine phosphatases, and at least three potential protein aspartate phosphatase paradigms (reviewed in [4]). In most cases, extensive sequence comparisons, complemented by empirical structure–function studies, have led to the identification of conserved sequence motifs that have proven to be of great utility for the identification of new family members.

The genome sequences of nine archaeons have been published as of this writing: *Aeropyrum pernix* [49], *Archaeoglobus fulgidus* [50], *Halobacterium* NRC1 [51], *M. thermoautotrophicum* [52], *Methanococcus jannaschii* [53], *P. horikoshii* [54], *S. solfataricus* [55], *Thermoplasma acidophilum* [56] and *Thermoplasma volcanium* [57]. Below are summarized the insights derived from homology searches regarding the nature of the protein kinases and protein phosphatases responsible for modulating the phosphorylation and dephosphorylation of archaeal proteins.

Two-component histidine kinases

Histidine kinases constitute the upstream component of a signal-transduction unit referred to as the two-component signalling paradigm (reviewed in [2,3,5]). Histidine kinases act via a Ping Pong mechanism in which they first phosphorylate themselves on a histidine residue using ATP, and then transfer the phosphoryl group to a conserved aspartate residue on the response regulator domain that constitutes the downstream element of this binary module. The two-component paradigm is present in most bacterial organisms, some of which possess upwards of 30 histidine kinases and a like number of cognate response regulators [58,59]. Most two-component signal-transduction cascades activate the expression of genetic operons whose protein products participate in the uptake and metabolism of specific nutrients, such as phosphate or nitrogen, or modulate adaptation to changes

in the osmolarity of the surrounding medium or other environmental stresses. In these signal-transduction cascades, the response regulator domain is usually fused to a transcriptional activator domain whose potency is thereby controlled via phosphorylation. One notable exception to this pattern is the Che system that guides bacterial chemotaxis, in which the phosphorylated response regulator protein modulates the activity of a flagellar motor.

Although the two-component system has been found in several eukaryotes, such as yeast and plants, mammals possess neither *bona fide* histidine kinases nor their cognate response regulators (reviewed in [60]). They do contain, however, highly divergent variants of the histidine kinase paradigm, collectively referred to as mitochondrial protein kinases [61]. Mitochondrial protein kinases, which lack the autophosphorylated histidine residue conserved among their two-component counterparts, catalyse the direct transfer of a phosphoryl group from ATP to serine residues on exogenous protein targets such as pyruvate dehydrogenase and α -ketoacid dehydrogenase. Other serine/threonine-specific offshoots of the histidine kinase paradigm include eukaryotic phytochromes [62] and a bacterial enzyme family whose members participate both in the modulation of sporulation and stress responses in *B. subtilis* [63,64] and of carbon metabolism in *Synechocystis* PCC6803 [65].

Homology searches have revealed that four of the nine currently published archaeal genome sequences contained ORFs potentially encoding examples of the histidine kinase paradigm: *A. fulgidus*, *Halobacterium* NRC1, *M. thermoautotrophicum* and *P. horikoshii* (Table 1). While only one potential histidine kinase was evident in *P. horikoshii*, the other three archaeons each contained an estimated 14–16 of them [58,59]. All four archaeons also contained ORFs potentially encoding two-component response regulator domains in numbers comparable with those of their presumed cognate kinases [58]. No ORFs encoding potential serine/threonine-specific variants of the histidine kinase paradigm, i.e. homologues lacking the autophosphorylated histidine residue, were reported. The strong similarities between archaeal two-component modules, including the presence of structural and functional homologues (see below) of the bacterial chemotaxis system in *A. fulgidus*, *Halobacterium* NRC1 and *P. horikoshii*, indicate that the Archaea acquired the two-component

paradigm through a series of lateral gene transfers from the bacterial organisms in which this signal-transduction system is believed to have originated [58,59].

Aspartate phosphatases

In Bacteria, multiple mechanisms have been proposed for the enzyme-catalysed dephosphorylation of the phosphoaspartate moieties on response regulator proteins (reviewed in [4]). Although several response regulators have been demonstrated to possess intrinsic autophosphatase activity [66–68], it is unclear whether the ability to self-dephosphorylate is universal. Other proposed mechanisms for the dephosphorylation of response regulator proteins include the intervention of bifunctional variants of the histidine kinase paradigm [68,69], as well as three families of dedicated protein aspartate phosphatases: CheZ [66], Spo0E [70] and the regulator aspartyl phosphatases (Raps) [70,71]. However, it has yet to be determined whether these latter proteins act catalytically, or simply stimulate the latent autophosphatase activity of the response regulators to which they bind. Thus far, no evidence for the presence of homologues of CheZ, Spo0E or the Raps has surfaced from the Archaea.

Eukaryotic protein kinases

The eukaryotic protein kinase superfamily, whose prototype is the catalytic subunit of the cAMP-dependent protein kinase, represents the predominant source of protein serine/threonine/tyrosine kinase activity in members of the Eucarya (reviewed in [72,73]). Approx. 2% of the ORFs within the genomes of *Saccharomyces cerevisiae* [74], *Caenorhabditis elegans* [75] and *Drosophila melanogaster* [76] encode known or potential eukaryotic protein kinases. Although eukaryotic protein kinases were believed for many years to reside exclusively in members of the Eucarya, several bacterial homologues have been identified and characterized during the last decade (reviewed in [77]). Interestingly, although the total number of ORFs encoding known or potential two-component histidine kinases in the Bacteria thus far significantly outstrips that for the eukaryotic protein kinase paradigm (for example, *Synechocystis* PCC6803 has more than 30 ORFs encoding potential histidine kinases, as compared with 12 encoding potential eukaryotic protein kinases), the number of bacterial genomes that contain known or potential eukaryotic protein kinases equals, and may even slightly exceed, those containing the 'bacterial' two-component system [77].

The members of the eukaryotic protein kinase family are characterized by a catalytic core domain approx. 280 amino acids in length containing 12 conserved sequence motifs, each ranging from 4–20 residues in length, referred to as subdomains I–V, VIa, VIb and VII–XI [72,73]. These subdomains are distributed among three structural units: an N-terminal nucleotide-binding domain containing subdomains I–IV, a C-terminal phosphotransfer and protein-substrate-binding domain containing subdomains VIa–XI, and an intervening linker containing subdomain V. Only a handful of the amino acid residues comprising these subdomains are universally conserved among the members of this extraordinarily large protein superfamily. Not surprisingly, the majority of these are found within those regions of the enzyme most directly involved in executing the core functions shared among all members of the superfamily, i.e. binding of their common nucleotide substrate (subdomains I–III), catalysis of phosphotransfer (subdomain VIb), interaction with their metal-ion cofactor, usually Mg^{2+} (subdomain VII), or binding of protein substrates (subdomain VIII) [72,73]. Subdomains IX and XI form a salt bridge between their respective

conserved aspartate and arginine residues that helps to support the conformational integrity of the C-terminal lobe. The other subdomains, i.e. IV, V, VIa and X, exhibit very weak sequence conservation (usually the presence and spacing of a few hydrophobic residues), reflective of their largely structural roles.

Seven of the nine archaeal genomes surveyed contained ORFs whose predicted products contained plausible candidates for all of the most functionally critical subdomains within the eukaryotic protein kinase paradigm, i.e. subdomains I, II, III, VIb, VII and VIII (Table 2). Thus the probability that these ORFs encode proteins with phosphotransferase activity would appear to be quite high. It also should be noted that, whereas it has been reported that the genome sequence of *Halobacterium* NRC1 contained two ORFs, *mak* and *prk*, encoding potential eukaryotic protein kinases [51], inspection of their DNA-derived amino acid sequences indicated that both lacked plausible candidates for such functionally critical subdomains as VIb, which forms the catalytic loop (P. J. Kennelly, unpublished work). For this reason, they were omitted from further consideration.

In the Eucarya, the vast majority of known or potential eukaryotic protein kinases are characterized by the presence of a basic amino acid within subdomain VIb. The nature of this amino acid has proven to be a strong indicator of whether the protein kinase in question is serine/threonine- or tyrosine-specific. Most protein serine/threonine kinases contain a lysine residue (shown in bold) two residues downstream of the essential, invariant aspartate of the catalytic loop: Asp-Xaa-Lys-Xaa-Xaa-Asn. Protein tyrosine kinases, on the other hand, generally have an arginine residue (also shown in bold) either two or four residues after the essential aspartate: Asp-Xaa-Arg-Xaa-Xaa-Asn (Src family) or Asp-Xaa-Xaa-Xaa-Arg-Asn (receptor kinases) [78]. With the exception of ORFs *sso2291*, *sso3207* and *sso3182* from *S. solfataricus*, each of which contains the conserved lysine residue characteristic of protein serine/threonine kinases, neither lysine nor arginine residues were evident in subdomain VIb of potential archaeal protein kinases (Table 2). This observation does not in itself preclude the possibility that these latter ORFs encode functional protein kinases. Eucaryal organisms also contain a handful of eukaryotic protein kinases lacking a lysine or arginine residue within the catalytic loop. Two of these, specifically PID261/BUD32 [79] and Rio1p [80], both from yeast, have been shown to possess protein serine/threonine kinase activity. PKN6, a eukaryotic protein kinase from the bacterium *Myxococcus xanthus*, autophosphorylates itself, despite the absence of a lysine or an arginine residue in its catalytic loop [81]. Moreover, the recombinant products of two of the archaeal ORFs in question, *sso0469* and *sso2387* from *S. solfataricus*, also exhibited protein serine kinase activity *in vitro* (see below). Thus, although an arginine or lysine residue represents a common feature of subdomain VIb in eukaryotic protein kinases, they are not essential for either phosphotransferase activity or the targeting of protein substrates.

When one examines the extreme C-terminal portion of many of these potential archaeal versions of the eukaryotic protein kinase paradigm, they appear to be stunted relative to their eukaryotic counterparts (Table 2 and [82]). In many cases, one or more of the last three subdomains appears to be completely lacking and/or the spacing between the putative subdomains must be reduced significantly below that observed in prototypical eukaryotic protein kinases in order to fit the established template. These disparities suggest that the fold of these C-terminal domains may differ somewhat from that found in prototypical eukaryotic protein kinases. Since the C-terminal lobe binds the phosphoacceptor substrate, it therefore must be asked whether some of these potential archaeal kinases target non-protein

Table 2 Salient features of potential archaeal members of the eukaryotic protein kinase superfamily

Listed below are those ORFs encoded by archaeal genomes whose predicted protein products display significant resemblance to the eukaryotic protein kinase paradigm. Shown are the consensus sequences for each of these subdomains (subdomains I–V in a; subdomains Via, Vlb and VII–XI in b), as described by Hanks and Hunter [73]. Symbols used to denote the various features are as follows: uppercase letters, universally conserved amino acid residues; lowercase letters, highly conserved amino acid residues; 'o', positions conserving non-polar residues; *, positions conserving polar residues; 'x', any amino acid; and +, positions conserving small residues with near neutral polarity. Question marks (?) designate motifs for subdomains for which no homologous sequences were apparent, whereas a dash (–) indicates the complete absence of that subdomain. Residue position within a representative catalytic domain is indicated using the sequence of the cAMP protein kinase α subunit from the mouse. Shown below each consensus sequence are the regions from the potential archaeal proteins postulated to correspond thereto. Residues that match the consensus are shown in uppercase letters; all the others are shown in lowercase letters. Abbreviations used for the organisms: *A. pern.*, *Aeropyrum pernix*; *A. fulg.*, *Archaeoglobus fulgidus*; *M. therm.*, *M. thermoautotrophicum*; *M. jann.*, *Methanococcus jannaschii*; *P. hori.*, *Pyrococcus horikoshii*; *S. solf.*, *Sulfolobus solfataricus*; *T. volc.*, *Thermoplasma volcanium*; aa, amino acids.

(a)

			Subdomain					
Organism	ORF	Length	I	II	III	IV	V	
			Consensus ...	ogxG ₅₂ xogxv	oaoK ₇₂ xo	E ₉₁ xxoo	h ₁₀₀ xxooxoxxxxo	oooo*oo ₁₂₃
<i>A. pern.</i>	<i>APE0751</i>	266 aa		IssG ₆₇ keskV	VAVK ₇₇ iY	E ₁₀₁ dipa	r ₁₀₉ rlIYeWarke	IIVMEFL ₁₄₉
<i>A. fulg.</i>	<i>AF1804</i>	258 aa		IGtG ₅₈ keanV	MAVK ₈₀ iY	E ₁₀₀ rfdM	k ₁₁₀ ekVFiWteke	VLLMEFI ₁₅₀
<i>M. therm.</i>	<i>mt1005</i>	535 aa		eycG ₂₈₅ dnGam	LvYK ₃₀₀ Yr	E ₃₂₀ VdVp	?	?
<i>M. jann.</i>	<i>MJ0444</i>	290 aa		VnsG ₈₅ keavV	rAVK ₁₀₃ vY	E ₁₄₃ frnL	?	VLVMDFV ₁₇₂
<i>P. hori.</i>	<i>PH0445</i>	219 aa		IkgG ₇ aeaki	YpLK ₂₇ vI	E ₅₅ arIL	k ₆₃ elgVnVpyvF	IIVMEYI ₁₂₇
<i>S. solf.</i>	<i>ss00469</i>	582 aa		tGsG ₁₉₅ ksnla	pdtK ₂₁₅ vV	E ₂₃₅ vpsr	H ₂₆₁ vipddIidir	LtVADLI ₃₀₆
<i>S. solf.</i>	<i>ss02291</i>	554 aa		nkiG ₂₄₂ dYevV	YAMK ₂₇₁ ip	E ₂₉₇ finL	t ₃₀₈ knVVqLfais	YIVMELM ₃₄₈
<i>S. solf.</i>	<i>ss02387</i>	583 aa		MsiG ₃₂₄ vtGag	esMK ₃₄₆ iV	E ₃₅₅ irLA	H ₄₁₇ ggattFhayn	VYIGeKI ₄₆₇
<i>T. volc.</i>	<i>TVN0776</i>	510 aa		aisG ₁₂₀ aslgq	VvVK ₁₃₉ vn	E ₁₆₈ nFLY	?	VLVMEYI ₂₃₂

(b)

Organism	ORF	Subdomain							
		Vla	Vlb	VII	VIII	IX	X	XI	
		Consensus ...	+oxooh ₁₅₈	oohrD ₁₆₆ ok+xNooo	oko+D ₁₈₄ fgo+	+pE ₂₀₈ oo	D ₂₂₀ oo+ogooo	ooxxo ₂₅₀	R ₂₈₀ x+
<i>A. pern.</i>	<i>APE0751</i>	?	LVHaD ₁₉₅ LseyNIlV	pwIiD ₂₁₂ vsqA	AeE ₂₂₆ FL	?	MgfeF ₂₄₇	R ₂₆₂ gA	
<i>A. fulg.</i>	<i>AF1804</i>	vkrLYq ₁₈₈	LVHaD ₁₉₆ LseyNIMY	VyFiD ₂₁₂ mGqA	maE ₂₂₅ SY	D ₂₃₁ VrNIirFF	MLkeV ₂₅₅	–	
<i>M. therm.</i>	<i>mt1005</i>	aIgrLH ₄₄₄	IIHgD ₄₅₀ LtGsNIll	VvFiD ₄₆₉ FGgl	sLE ₄₉₄ st	e ₅₀₃ CFSrvLeg	–	R ₅₂₅ ei	
<i>M. jann.</i>	<i>MJ0444</i>	yFkIIK ₁₉₈	LVHGD ₂₁₄ LseyNIlV	pvFiD ₂₃₁ FGqS	?	D ₂₆₄ cnykdLYk	MikqL ₂₉₀	–	
<i>P. hori.</i>	<i>PH0445</i>	lIgrLH ₁₁₉	IVHGD ₁₂₇ LtTsNMIM	IyFiD ₁₄₄ FGLA	AmE ₁₆₉ st	e ₁₈₂ VLNgyeeV	?	R ₂₁₁ yr	
<i>S. solf.</i>	<i>ss00469</i>	?	ddiRD ₃₄₉ dsinNLLd	rkteD ₃₉₄ fgIe	iPE ₄₁₈ Lp	D ₄₄₆ npkVlFII	aIekL ₄₈₁	R ₄₉₈ lA	
<i>S. solf.</i>	<i>ss02291</i>	?	ALdVIH ₃₈₄	YVHL ₃₉₂ VKpqNIYF	VKLGD ₄₂₆ lGSa	qiE ₄₅₁ aa	D ₄₆₅ IFALGata	Ivera ₄₉₄	R ₅₁₃ qY
<i>S. solf.</i>	<i>ss02387</i>	?	dpksD ₄₉₆ afTiNLdA	nrIee ₅₁₄ aGLn	kPE ₅₆₄ ea	–	–	–	
<i>T. volc.</i>	<i>TVN0776</i>	mrmLLK ₂₆₇	IFHaD ₂₇₄ phpgNIsv	IiLyD ₂₉₂ FGMv	sdE ₃₂₁ II	D ₃₄₆ MaianFYg	VifqF ₃₇₆	R ₄₂₈ lT	

substrates. A precedent for such behaviour has been supplied by two bacterial homologues of the eukaryotic protein kinase paradigm that act on non-protein substrates: the aminoglycoside antibiotic kinases [83] and a lipopolysaccharide kinase [84]. None of the known or potential archaeal kinases contain plausible candidates for motif III, a conserved sequence feature unique to the antibiotic kinases [85], whereas a recombinant version of SsoPK3, the product of one of the most severely truncated ORFs, *ss00469* (Table 2), exhibited protein kinase activity *in vitro* (see below). However, until the enzymic properties and physiological roles of these potential protein kinases has been investigated more thoroughly, due caution must be exercised in extrapolating their functions on the basis of sequence comparisons.

Dendograms constructed from available sequence information suggest that the eukaryotic protein kinase superfamily may trace its origins to the epoch of the universal ancestor [86]. The ORFs encoding potential protein kinases in the Archaea appear to have been acquired via inheritance from the last common precursor of the combined archaeal/eucaryal line of descent, as no evidence has as yet been uncovered for acquisition by the Archaea of eukaryotic protein kinases via lateral gene transfer [86]. This conclusion must be viewed with caution, however, given the limited number of archaeal genomes analysed and the occurrence

of at least two such lateral gene-transfer events between the Eucarya and the Bacteria [86,87].

Protein serine/threonine phosphatases

To date, two major families of protein serine/threonine phosphatases have been discovered in living organisms, referred to as the PPPs and PPMs [88]. The members of the PPP family, which includes protein phosphatases 1, 2A and 2B (also called calcineurin), constitute the quantitatively dominant source of protein serine/threonine phosphatase activity in eukaryotic organisms [89,90]. Members of the PPP family also have been found in several bacterial organisms, including *E. coli* [91] and *Microcystis aeruginosa* [92], as well as in bacteriophage λ [93]. The members of the PPM family include protein phosphatase 2C and pyruvate dehydrogenase phosphatase in eukaryotes [88,94]. The PPMs are the most prevalent family of protein serine/threonine phosphatases in the Bacteria. Known or potential PPMs are present in slightly more than half of the bacterial genomes sequenced to date [4,77], where they participate in the control of sporulation and stress responses in *B. subtilis* [95,96] and *M. xanthus* [97], and the co-ordination of carbon metabolism in *Synechocystis* PCC6803 [65,98].

Table 3 Salient features of known or potential archaeal members of the PPP family of protein phosphatases

Listed below are those ORFs encoded by archaeal genomes whose predicted protein products display significant resemblance to PPP protein phosphatase paradigm. Shown are the conserved sequence motifs characteristic of the PPP family for each of these subdomains, as described in Barton et al. [89], using position numbers derived from the catalytic subunit of sds 21, a PPP from *S. pombe*, and below them the corresponding sequences from the indicated archaeal proteins or potential ORF products. For an explanation of the symbols, see the legend to Table 2. Abbreviations used for organisms: *A. pern.*, *Aeropyrum pernix*; *A. fulg.*, *Archaeoglobus fulgidus*; *M. thphl.*, *Methanosarcina thermophila*; *P. abyssi*, *Pyrodicticum abyssi*; *S. solf.*, *Sulfolobus solfataricus*; *T. volc.*, *Thermoplasma volcanii*. aa, amino acids.

Organism	ORF/protein	Length	Motif		
			I	II	III
			Consensus ...		
			o*oGD ₆₁ o	ooooGD ₉₁ o	oorGNHE ₁₂₁
<i>A. pern.</i>	<i>APE0777</i>	266 aa	MGDIHG ₆ DY	MFLG ₃₈ DYIDRG ₃₈	MLRG ₆₉ NHE
<i>A. fulg.</i>	<i>AF1822</i>	268 aa	tvVGDVHa ₃₅ DI	aIFLGDY ₆₁ aDRG ₆₁	LLRG ₈₈ NHE
<i>M. thphl.</i>	<i>PP1-arch2</i>	268 aa	VmIGDIHG ₄₉ NL	FLFLGDYVDRG ₇₉	LLRG ₁₀₈ NHE
<i>P. abyssi</i>	<i>Py-PP1</i>	302 aa	LvaGDtHG ₃₈ YP	VVFLGDYVDRG ₆₈	LLRG ₉₇ NHE
<i>S. solf.</i>	<i>PP1-arch1</i>	293 aa	VFVGdtHG ₄₀ AI	IVFLGDYVDR ₆₉ e	VLRG ₁₀₄ NHE
<i>T. volc.</i>	<i>TVN1195</i>	229 aa	IiYsDIHa ₁₁ NL	aIFLGDVVDyG ₃₈	gVmGNHd ₆₁

Figure 3 DNA-derived amino acid sequence of TVN0703, a potential archaeal PPM

Shown is the DNA-derived amino acid sequence of ORF *TVN0703* from the archaeon *T. volcanium* [57]. Below it are shown the 11 conserved subdomains common to the PPM family of protein serine/threonine phosphatases [94], aligned with the most closely corresponding regions of TVN0703, with matches indicated using red lettering. The amino acid residues corresponding to those that bind either directly or through bridging-water molecules to the two bivalent metal ions present within the active site of human PP2C α [156] are portrayed using green lettering. #, Positions conserving charged residues; for an explanation of the other symbols, see the legend to Table 2.

The members of the PPP family are characterized by a set of three conserved sequence motifs containing sets of catalytically critical residues [89,99]. Four of the archaeal genomes published to date contain an ORF encoding a known [e.g. PP1-arch1 from *S. solfataricus* (see below)] or potential member of the PPP family of protein serine/threonine phosphatases (Table 3). In addition, functional PPP-family protein serine/threonine phosphatases have been cloned from the archaeons *M. thermophila* TM-1 [100] and *Pyrodicticum abyssi* TAG11 [101] (see below).

The level of sequence identity shared between those archaeal PPPs demonstrated to possess protein phosphatase activity and their eukaryotic counterparts ($\geq 30\%$) far exceeds that apparent between either archaeal and bacterial PPPs or eukaryotic and bacterial PPPs (typically $\leq 19\%$ in each case). The former value also closely approaches the global degree of sequence identity shared among the catalytic domains of the eukaryotic PPPs themselves (34–35%; [89]). However, it should be noted that one of the potential PPP ORFs described in Table 3, *TVN1195* from

T. volcanii, exhibited deviations in two 'universally' conserved residues that raise doubts as to whether its potential protein product would possess protein phosphatase activity.

Sequence comparisons indicate that the progenitor of the PPP superfamily dates back to the hypothetical universal ancestor, and that the PPPs present in contemporary organisms generally were acquired by simple inheritance, with little or no dissemination via lateral gene transfer [92].

A single ORF encoding a potential member of the PPM family of protein serine/threonine phosphatases has been discovered in the genome of *T. volcanii* [57]. The predicted protein product of this ORF contained virtually all of the conserved sequence motifs characteristic of established members of the PPM family (Figure 3), and thus is highly likely to possess protein phosphatase activity. Acquisition of this PPM most likely occurred by horizontal gene transfer from a bacterium. The Bacteria, in turn, are thought to have acquired their PPMs from the Eucarya [86].

PTPs

Three major families of PTPs have been identified to date: the conventional PTPs (cPTPs), the low-molecular-mass PTPs (LMM PTPs; also known as 'LMW' PTPs) and the Cdc25 family. The members of each family have converged upon a common catalytic mechanism and active-site motif, Cys-Xaa₃-Arg [102,103]. The cPTPs form the most numerous and structurally diverse family of protein phosphatases in eukaryotes [104,105]. The extended cPTP family includes dual-specific phosphatases that act on phosphoserine and phosphothreonine in addition to phosphotyrosine [106], as well as variants that hydrolyse inositol phospholipids [107,108] and RNA di- and triphosphates [109]. By contrast, only a small handful of LMM PTPs are present in eukaryotes, and no dual-specific variants have been reported [110]. Both cPTP and LMM PTPs are present in the Bacteria [111–113].

Several bacterial organisms contain arsenate reductases that are offshoots of the LMM PTP paradigm [82]. These enzymes not only share the characteristic Cys-Xaa₃-Arg active site sequence of the PTPs, but their mechanism of catalysis apparently involves the formation of an arsenocysteine intermediate analogous to the phosphocysteine formed by their hydrolytic counterparts [114]. Intriguingly, the arsenate reductase encoded by plasmid p1258 from *Staphylococcus aureus* recently was shown to exhibit weak (vestigial?) phosphohydrolase activity towards *p*-nitrophenyl phosphate [115].

The Cdc25 family comprises a handful of highly specialized dual-specific phosphatases that target adjacent phosphothreonine and phosphotyrosine residues on cyclin-dependent kinases [116]. Their sole physiological function is to modulate progression through the eukaryotic cell cycle. X-ray crystallography has revealed that Cdc25 possesses the same basic fold as does the sulphur-metabolizing enzyme rhodanese [117]. Whereas rhodanese is commonly found in bacterial organisms, no Cdc25-type phosphatases have been uncovered in the Bacteria. However, in an interesting parallel with LMM PTPs and bacterial arsenate reductases, the arsenate reductase of yeast has been determined to be a structural homologue of Cdc25 [82].

ORFs potentially encoding PTPs were evident in five of the nine archaeal genomes published to date (Table 4). Three archaeons contained potential cPTPs: *M. jannaschii*, *P. horikoshii* and *S. solfataricus*. One contained only a potential LMM PTP, *A. fulgidus*, whereas *M. thermoautotrophicum* contained one of each. No potential Cdc25 phosphatases have been identified in any archaeal genome. Only one potential archaeal PTP has been characterized to date: *Tk*-PTP from *T. kodakaraensis* (see below).

Table 4 Salient features of potential archaeal PTPs: conventional PTPs (a) and LMM PTPs (b)

Listed below are those ORFs encoded by archaeal genomes whose predicted protein products display significant resemblance to either the conventional PTP paradigm (a) or the LMM protein phosphatase paradigm (b). Also included is *Tk*-PTP, a member of the conventional PTP family from the archaeon *T. kodakaraensis* (*T. kod.*) that, in recombinant form, hydrolysed both free phosphotyrosine and free phosphoserine *in vitro* [42]. Shown are the conserved sequence motifs characteristic of the conventional and LMM PTPs, as described in Fauman and Saper [154], and below them the corresponding regions of the predicted products of the indicated ORFs. For an explanation of symbols and the other abbreviated organism names, see the legend to Table 2.

(a)			
Organism	ORF	Length	Motif
			Consensus ...
<i>M. therm.</i>	<i>mt1586</i>	233 aa	DX ₂₇ ooHCxagxxRTg
<i>M. jann.</i>	<i>MJ0215</i>	159 aa	DX ₃₃ agaCagGvmRS1
<i>M. jann.</i>	<i>MJ1098</i>	254 aa	DX ₂₆ VVsCigGhgRTG
<i>M. jann.</i>	<i>MLECL20</i>	147 aa	DX ₃₄ YfHCykykrRTS
<i>P. hori.</i>	<i>PH1732</i>	146 aa	DX ₂₆ VVsCigGhgRTG
<i>S. solf.</i>	<i>sso2453</i>	161 aa	DX ₂₆ YIHCygGsgRSG
<i>T. kod.</i>	<i>Tk</i> -PTP	147 aa	DX ₂₄ LvhCvgGigRTG
			DX ₂₆ LIHCmgGlgRSG
(b)			
Organism	ORF	Length	Motif
			Consensus ...
<i>A. fulg.</i>	<i>AF1361</i>	135 aa	ooC1GniCRSX ₉₉ DP
<i>M. therm.</i>	<i>mt1355</i>	130 aa	FVCInHtaRSX ₈₈ KD
			FICrnNsgRSX ₈₃ PD

The amino acid sequence of *Tk*-PTP conforms to the cPTP paradigm (Table 4).

Other protein serine/threonine/tyrosine kinases and phosphatases

Recognizable homologues of several protein serine/threonine/tyrosine kinase and/or phosphatase paradigms previously identified in members of the Bacteria and Eucarya have yet to be encountered in any member of the Archaea. The absentees include the isocitrate dehydrogenase kinase/phosphatase [23], the histidine-containing protein kinase/phosphatase [118,119], and the myosin heavy chain/elongation factor-2 protein kinase family [120]. These enzymes share three common characteristics: residency within either the Eucarya or Bacteria, but not both (and, in several cases, only a small subgroup thereof); no or very few isoforms within any given organism; and extreme specialization, i.e. the targeting of a protein substrate. The limited extent to which these families have proliferated may reflect a late start relative to other protein kinase or protein phosphatase paradigms, and/or competitive disadvantages arising from inherent limitations of their basic structure or mode of action. The potential development and proliferation of the bifunctional isocitrate dehydrogenase kinase/phosphatase paradigm, for example, presumably has been severely circumscribed by the use of a common active site to catalyse its antagonistic functions [4]. This feature almost inevitably restricts the range of action of this enzyme or any potential derivatives to one or more common substrates in a simple on-off mode.

Do novel protein kinases or phosphatases exist in the Archaea?

Surveys indicated that eight of the nine archaeal genome sequences published to date contained ORFs encoding known or potential protein kinases, the exception being *T. acidophilum*

(Table 1). Seven of these eight contained ORFs potentially encoding homologues of the eukaryotic protein kinase superfamily, while four contained potential two-component histidine kinases. The greater pervasiveness of the eukaryotic protein kinase paradigm is consistent with analyses indicating that the progenitor of this family was present in the immediate predecessor of the Archaea, whereas the two-component system was acquired at a later time via horizontal gene transfer from the Bacteria. Seven archaeons contained one or more ORFs encoding known or potential protein serine/threonine phosphatases or PTPs. No evidence for the presence of dedicated protein aspartate phosphatases was apparent in any of these archaeal genomes, suggesting that their response regulators are dephosphorylated via simple chemical hydrolysis, by the intrinsic autophosphatase activity of the response regulators themselves, and/or through the action of bifunctional histidine kinases, the last two of which have been observed in certain bacterial two-component modules.

There was no obvious pattern to the distribution of those ORFs potentially encoding phosphomonoester-specific protein phosphatases, beyond the fact that they were found only in those archaeons that contained ORFs for potential eukaryotic protein serine/threonine/tyrosine kinases (Table 1). Some archaeons possessed only the presumably serine-/threonine-specific protein phosphatases of the PPP or PPM family, while others contained only PTPs (Table 1). In only two instances did an archaeon harbour an example of each: *M. thermoautotrophicum* and *S. solfataricus*. In several archaeons, specifically *M. thermoautotrophicum*, *P. horikoshii* and *S. solfataricus*, the ratio of potential protein serine/threonine/tyrosine kinases to potential counter-vailing protein serine/threonine/tyrosine phosphatases was quite high (4:1).

Are we to infer from the absence of stereotypical protein serine/threonine phosphatases in *M. jannaschii* and *P. horikoshii* that these organisms contain only protein tyrosine kinases, or that protein serine/threonine phosphorylation events in these organisms are not enzymically reversible? Or does this mosaic distribution of potential protein serine/threonine/tyrosine phosphatases in archaeal genomes, and the frequently large disparity in the number of potential protein serine/threonine/tyrosine kinases to potential protein serine/threonine/tyrosine phosphatases, imply the existence of some as-yet-unrecognized family of protein phosphatases? Does the lack of ORFs whose products exhibit recognizable resemblance to established protein kinases in *T. acidophilum* imply the existence of some novel archaeal protein kinase? Definitive answers to these questions must await further investigation. However, it should be noted that several members of the cPTP family in eukaryotic [106] and bacterial organisms [113] are dual-specific enzymes capable of hydrolysing protein-bound phosphoserine and phosphothreonine in addition to phosphotyrosine. Under the circumstances, it seems likely that several of their potential archaeal counterparts also may be dual-specific. The fact that every archaeon that contained a potential LMM PTP also contained a presumably serine/threonine-specific potential PPP lends circumstantial support to this argument, since no dual-specific members of the LMM PTP family have been reported to date.

The 'scatter-gun' pattern of known or potential protein phosphatases and, to a lesser degree, protein kinase ORFs presumably reflects a tendency of these organisms to undergo reductive evolution, i.e. to discard un-needed genetic material subsequent to their adaptation to highly specialized environmental niches [121]. The extremophilic nature of many archaeons severely limits the range of venues in which they can grow and prosper. The resulting environmental monotony has minimized

or, perhaps in the case of *T. acidophilum*, eliminated the demand for phosphorylation-based sensor-response cascades. It should be noted that the genomes of two other organisms, the bacteria *Buchnera* sp. and *Rickettsia prowazekii*, also appeared to be devoid of ORFs encoding recognizable homologues of known protein kinases or protein phosphatases [77]. However, the existence of uniquely archaeal solutions to the exploitation of phosphate as a protein regulatory tool cannot be discounted.

FUNCTIONAL CHARACTERISTICS OF ARCHAEL PROTEIN KINASES AND PROTEIN PHOSPHATASES

Two-component systems

In 1995, Rudolph and Oesterhelt reported the successful cloning of a gene encoding a two-component histidine kinase from the halophilic archaeon *H. salinarum* [122]. The DNA-derived amino acid sequence of the predicted gene product, CheA_{H.s.}, displayed approx. 33 % identity with the histidine kinases of the chemotactic signal-transduction systems of *E. coli* and *B. subtilis*. Further analysis indicated that the genes adjacent to that for CheA_{H.s.} encoded homologues of other core components of the *E. coli* chemotactic signal-transduction cascade, i.e. the CheY response regulator protein and the CheB methyltransferase responsible for modulating receptor sensitivity [123]. Deletion of the genes encoding CheA, CheB, or CheY in *H. salinarum* led to loss of chemotactic and phototactic migration.

Recombinantly produced CheA_{H.s.} phosphorylated itself when incubated with [γ -³²P]ATP, a process that was dependent upon the presence of the conserved histidine residue that constitutes the site of autophosphorylation on classic two-component histidine kinases [124]. Moreover, the ³²P label was observed to disappear from autophosphorylated CheA_{H.s.} upon incubation, CheY_{H.s.}, as would be expected if the phosphoryl group had been transferred to its presumed cognate substrate. This interpretation was strongly corroborated by the observation that alteration of the predicted site of phosphorylation on CheY_{H.s.}, Asp⁵³, to a non-phosphorylatable alanine ablated the ability of the protein to induce the net dephosphorylation of CheY_{H.s.}.

Intriguingly, expression of CheA_{H.s.} restored chemotactic swimming to *E. coli* from which the endogenous gene for CheA had been deleted [122]. Similarly, Jung et al. [125] were able to produce a phototactic strain of *E. coli* by reconfiguring the bacterium's endogenous chemotactic signalling pathway using the HtrI and HtrII histidine kinases and their associated rhodopsin photoreceptors from a halophilic archaeon, *Natronobacterium pharaonis*. Thus bacterial and halobacterial CheAs behave as functionally interchangeable modules in the pathways linking chemo- or photo-tactic sensors with flagellar motors. These findings corroborated phylogenetic analyses indicating that archaeal organisms imported the two-component signalling paradigm from members of the Bacteria through a series of lateral gene transfers [58,59].

I am not aware of any published studies in which either an archaeal histidine kinase or an archaeal response regulator protein has been tested for protein aspartate (auto)phosphatase activity.

Protein serine/threonine/tyrosine kinases

For many years, the characteristics of archaeal protein serine/threonine kinases have proven to be elusive. Although their presence could be readily detected by simply adding [γ -³²P]ATP to a cell lysate or other crude fraction, it was not until quite recently that scientists have begun to trace these activities to their polypeptide sources. In 1999, Bischoff and Kennelly [126] re-

ported that a polypeptide of approximate size 67 kDa with protein serine/threonine kinase activity could be renatured within the matrix of a polyacrylamide gel following SDS/PAGE. By co-polymerizing an exogenous protein, casein, into the gel matrix, the transfer of phosphate from [γ - 32 P]ATP to the immobilized substrate protein could be observed in the region containing the renatured protein kinase. Solubilization of this protein kinase activity from the particulate fraction required the use of non-ionic detergents such as Triton X-100 or octylglucoside, indicating that its source was an integral membrane protein.

In addition to casein, the membrane-associated protein kinase also catalysed its autophosphorylation, as well as phosphotransfer to several exogenous peptides or proteins *in vitro*, including histone H4, reduced, carboxyamidomethylated and maleylated lysozyme and MLC peptide, an oligopeptide modelled after the phosphorylation sites of myosin P light chains from smooth-muscle tissue [48]. Amino acid analysis indicated that, in every instance, phosphorylation took place exclusively on threonine residues, with the exception of histone H4, where some phosphoserine also was detected. Kinetic analyses comparing the phosphorylation of MLC peptide with that of a variant in which the phosphoacceptor threonine residue was replaced by serine revealed that the enzyme's preference for threonine was manifested as a > 20-fold difference in $V_{\max}$ [48].

In addition to its unusual selectivity for threonine, the membrane-associated protein kinase from *S. solfataricus* also exhibited an unusually broad specificity with regard to potential nucleotide substrates *in vitro*. While ATP clearly emerged as the most efficacious phospho-donor, measurable rates of phosphotransfer also were observed, in decreasing order, with GTP, ADP or GDP [48,126]. Little, if any, activity could be detected using purine monophosphates or a variety of pyrimidine nucleoside triphosphates. Although this behaviour is quite unusual, it is not without precedent, as a handful of established protein kinases, including casein kinase II [127], have been observed to utilize both ATP and GTP as phospho-donor substrates *in vitro*.

Although the membrane-associated protein kinase from *S. solfataricus* has steadfastly resisted all attempts to extract information regarding its amino acid sequence, it has recently been reported that the catalytic polypeptide is glycosylated [128]. Binding studies with lectins suggested that the oligosaccharide moiety (moieties) contained terminal mannose, but lacked terminal galactose or sialic acid. Although *S. solfataricus* synthesizes a protective external polysaccharide coating called the S layer [129], very little is known concerning protein glycosylation in this or any other archaeon. Available biochemical and genomic data do suggest, however, that molecular machinery exists in the Archaea that is homologous with the dolichol-mediated protein N-glycosylation pathway of eukaryotes [130].

In eukaryotes, glycosylation is generally limited to the extracellular domains of membrane-associated proteins and to those proteins that are secreted from the cell. To date, a similar pattern has been evident with regard to the glycosylation of archaeal proteins [129,131–134], suggesting that the membrane-associated protein kinase *S. solfataricus* possesses an extracellular domain. The observed dependence of protein kinase activity on pH indicates that the catalytic domain of the protein is intracellular [128], as this extreme acidothermophile maintains a cytoplasmic pH of ≈ 6 [135] despite the acidity (pH ≤ 3 ; [136]) of its external environment. It is therefore tempting to speculate that the glycosylated protein serine/threonine kinase from *S. solfataricus* exhibits a transmembrane topography consistent with the receipt and transmission of extracellular signals.

Recently, our laboratory has expressed and characterized the recombinant products of two of the eight ORFs encoded by

Table 5 Basic functional characteristics of archaeal PPPs

Summarized are a few of the physical and functional characteristics of those archaeal members of the PPP family that have been studied to date. For further details, see the accompanying text. aa, amino acids; n.d., not determined.

Phosphatase	Organism	Length	Activating metals	Substrate specificity		Toxin sensitivity
				S/T	Y	
PP1-arch1	<i>S. solfataricus</i>	293 aa	Mn, Ni, Co	+	–	–
PP1-arch2	<i>M. thermophila</i>	268 aa	Mn, Ni, Co	+	–	+
Py-PP1	<i>P. abyssi</i>	302 aa	Mn, Ni, Co	+	n.d.	+

the genome of *S. solfataricus* that displayed sequence features reminiscent of the 'eukaryotic' protein kinase paradigm. The protein products of ORFs *ss02387* and *ss0469* from *S. solfataricus* (Table 2) were designated SsoPK2 and SsoPK3 respectively. Both displayed ATP-dependent protein serine kinase activity, with little propensity to phosphorylate threonine residues, indicating that neither was the source of the membrane-associated protein kinase detected in previous studies (B. H. Lower and P. J. Kennelly, unpublished work). Mutagenic alteration of amino acid residues that had been aligned with functionally critical residues of established eukaryotic protein kinases, such as the aspartate of the catalytic loop (i.e. subdomain VIb), abolished catalytic function in both SsoPK2 and SsoPK3, further supporting their inclusion in this enzyme superfamily. In contrast with prototypical eukaryotic protein kinases, SsoPK2 and SsoPK3 each displayed a strong preference for Mn²⁺ over Mg²⁺ as a cofactor *in vitro*.

SsoPK2, but not SsoPK3, was observed to phosphorylate itself when incubated with [γ - 32 P]ATP *in vitro*. Mass peptide profiling has traced at least a portion of the sites upon which the protein self-phosphorylates to an area that is aligned with the activation loop of prototypical eukaryotic protein kinases (B. H. Lower and P. J. Kennelly, unpublished work). Intriguingly, phosphorylation of this region by either an exogenous protein kinase or via an autocatalytic process constitutes an important step in the activation of many eucaryal representatives of the eukaryotic protein kinase superfamily [137–139].

Protein serine/threonine phosphatases

To date, protein serine/threonine phosphatases from three different archaeons have been characterized at the molecular level: PP1-arch1 from *S. solfataricus* [140,141], PP1-arch2 from *M. thermophila* TM-1 [100,142] and Py-PP1 from *P. abyssi* TAG11 [101]. Typical of the Archaea themselves, PP1-arch1, PP1-arch2 and Py-PP1 embodied a paradoxical mixture of 'eucaryal' and 'bacterial' features (Table 5).

Functional characteristics shared by archaeal and eucaryal PPPs included specificity for protein-bound phosphoserine and phosphothreonine residues, and the sensitivity of PP1-arch2 [100,142] and Py-PP1 [101], but not PP1-arch1 [140], to toxic secondary metabolites such as okadaic acid (reviewed in [143, 144]). Whereas the two toxin-sensitive archaeal PPPs, PP1-arch2 and Py-PP1, did not exhibit the spectacular nanomolar affinities for okadaic acid characteristic of protein phosphatases 1 and 2A from eukaryotes, their sensitivity to micromolar concentrations of these compounds mirrored that of protein phosphatase 2B, also known as calcineurin [143]. By contrast, studies to date indicate that bacterial PPPs are toxin-insensitive and display little propensity to discriminate between phosphorylated serine,

threonine, tyrosine or histidine residues on proteins or other polypeptides (reviewed in [77]).

On the other hand, PP1-arch1, PP1-arch2 and Py-PP1 each required the presence of an exogenous cofactor, in the form of a bivalent metal ion such as Mn^{2+} , to support activity. The fact that one or both of the metals that make up the presumed bimetallic centre within the active site of archaeal PPPs was freely dissociable *in vitro* mirrors the behaviour of bacterial PPPs [91–93]. The active-site metal ions of eukaryotic PPPs, by contrast, are bound tenaciously by these enzymes, and hence they do not require added cofactor *in vitro* [88]. Bivalent-metal-ion-dependent protein serine/threonine phosphatase activity has also been detected in the archaeons *H. volcanii* [145] and *H. salinarum* [146]. The latter enzyme also hydrolysed *p*-nitrophenyl phosphate and free phosphotyrosine, suggesting that it may act on protein-bound phosphotyrosine as well. In neither case has the amino acid sequence of the enzymes responsible been determined, however.

No evidence has emerged as yet for the existence in archaeal organisms of the types of regulatory/targeting subunits that modulate the activity and site of action of protein phosphatase 1 and 2A in eukaryotes [147–149]; nor were any obvious regulatory or targeting domains evident within their sequences. However, it would be premature to conclude that no such subunits or domains exist. Published studies of the functional characteristics of both PP1-arch2 and Py-PP1 focused almost entirely on the recombinant products of their cloned genes, eliminating any opportunity for the formation of hetero-oligomeric complexes. Although PP1-arch1 was isolated from *S. solfataricus* as a monomer, the method used to track the enzyme relied upon the measurement of its catalytic activity [141]. Hence any oligomeric forms that were present in low abundance and/or which exhibited low specific activity could readily have been overlooked.

PTPs

To date, only a single potential archaeal PTP has been characterized: *Tk*-PTP from *T. kodakaraensis* [42]. *Tk*-PTP is a member of the cPTP family. Recombinant *Tk*-PTP exhibited phosphohydrolase activity towards both free phosphotyrosine and free phosphoserine *in vitro*, suggesting that it might be a dual-specific protein phosphatase. Mutational alteration of either the cysteine or the arginine within the presumed active-site sequence, specifically Cys⁹³ or Arg⁹⁹, abolished catalytic activity as predicted. However, alteration of Asp⁶³, the putative active-site aspartate, reduced catalytic activity by less than half. The latter finding suggests either that Asp⁶³ does not fulfil the role of catalytic acid–base in this phosphatase or that some other residue was able to compensate for its loss in the mutagenically altered enzyme. Although *Tk*-PTP has yet to be tested against any phosphoprotein substrates, the successful isolation of three phosphotyrosine-containing proteins using a catalytically inactive Cys⁹³ → Ala ‘substrate-trapping’ mutant of the enzyme (see above) strongly suggests that it possesses protein phosphatase activity.

HOW OLD IS REGULATORY PROTEIN PHOSPHORYLATION?

Recent studies, aided in no small measure by the advent of whole genome sequencing, have established the existence of protein phosphorylation/dephosphorylation in the members of the third domain of life, the Archaea. Thus with the exception of a small handful of micro-organisms that may once have possessed, but since discarded, this capability, these findings establish the

universality of protein phosphorylation/dephosphorylation as a molecular regulatory mechanism. Equally as important, the Archaea contain several enzymic paradigms for catalysing the phosphorylation and dephosphorylation of proteins considered originally to be exclusively eukaryotic or bacterial in residency, including the eukaryotic protein kinases, two-component histidine kinases, the PPP family protein phosphatases, PPM family protein phosphatases and both the conventional and LMM families of PTPs. Similarly, many members of the Bacteria and Eucarya have been demonstrated to possess eukaryotic or bacterial protein kinases and phosphatases respectively [82]. Even when one considers the potential contributions of Nature’s surprisingly vigorous ‘commerce’ in lateral gene transfer, the phylogenetic pervasiveness of these paradigms implies a lengthy evolutionary history: one considerably longer than had been imagined even a few years ago.

How old, then, is protein phosphorylation/dephosphorylation? It once was reasoned that the high ‘overhead’ costs associated with protein phosphorylation/dephosphorylation, i.e. the development of dedicated protein phosphotransferases and phosphohydrolases, delayed its appearance until long after the emergence of allosteric regulatory mechanisms. A key assumption underlying this chronology was that the first life forms were polypeptide-based [150], and thus developed the ability to synthesize proteins before their acquisition of information-bearing oligonucleotides. However, more recent models of evolution postulate that early life forms were ribozymal in nature [151]. Presumably, these self-replicating RNA molecules expanded their catalytic capabilities by incorporating amino acid cofactors [152] as the first step in the delegation of catalytic and information storage functions to proteins and oligonucleotides respectively. If this chronology is correct, living organisms acquired the catalytic machinery necessary to make and break phosphoester and phosphoramidate bonds long before the appearance of proteins themselves, offering fertile ground for the emergence of protein phosphorylation/dephosphorylation as a molecular regulatory mechanism.

Of greater authority in dating the origins of protein phosphorylation, analysis of two of the molecular paradigms that catalyse the phosphorylation and dephosphorylation of proteins, the eukaryotic protein kinase paradigm [86] and the PPP protein phosphatase paradigm [92], suggests that they may trace their progenitors to the last common ancestor. When did these enzymes apply their abilities to the task of molecular regulation? The answer to this question must await the identification of a much greater proportion of the proteins, and hence processes, that are subject to regulatory control via protein phosphorylation/dephosphorylation in the Bacteria and, in particular, the Archaea. Much work and many discoveries remain as research on archaeal protein phosphorylation/dephosphorylation moves from the genomic into the proteomic era.

I acknowledge the support of National Science Foundation grant MCB 0077484.

REFERENCES

- Kennelly, P. J. and Potts, M. (1996) Fancy meeting you here! A fresh look at ‘prokaryotic’ protein phosphorylation. *J. Bacteriol.* **178**, 4759–4764
- Aizawa, S.-I., Harwood, C. S. and Kadner, R. J. (2000) Signaling components in bacterial locomotion and sensory perception. *J. Bacteriol.* **182**, 1459–1471
- Hoch, J. A. (2000) Two-component and phosphorelay signal transduction. *Curr. Opin. Microbiol.* **3**, 165–170
- Kennelly, P. J. (2001) Protein phosphatases – a phylogenetic perspective. *Chem. Rev.* **101**, 2291–2312
- Bouret, R. B. and Stock, A. M. (2002) Molecular information processing: lessons from bacterial chemotaxis. *J. Biol. Chem.* **277**, 9625–9628

- 6 Mushegian, A. R. and Koonin, E. V. (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10268–10273
- 7 Bada, J. L. and Lazcano, A. (2002) Some like it hot, but not the first biomolecules. *Science* **296**, 1982–1983
- 8 Perutz, M. F. (1989) Mechanisms of cooperativity and allosteric regulation in proteins. *Q. Rev. Biophys.* **22**, 139–236
- 9 Westheimer, F. H. (1987) Why nature chose phosphates. *Science* **235**, 1173–1177
- 10 Johnson, L. N. and Lewis, R. J. (2001) Structural basis for control by phosphorylation. *Chem. Rev.* **101**, 2209–2242
- 11 Kennelly, P. J. and Krebs, E. G. (1991) Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. *J. Biol. Chem.* **266**, 15555–15558
- 12 Sala-Newby, G. B. and Campbell, A. K. (1991) Engineering a bioluminescent indicator for cyclic AMP-dependent protein kinase. *Biochem. J.* **279**, 727–732
- 13 Mattevi, A., Rizzi, M. and Bolognesi, M. (1996) New structures of allosteric proteins revealing remarkable conformational changes. *Curr. Opin. Struct. Biol.* **6**, 824–829
- 14 Helmstaedt, K., Krappmann, S. and Braus, G. H. (2001) Allosteric regulation of catalytic activity: *Escherichia coli* aspartate transcarbamoylase versus yeast chorismate mutase. *Microbiol. Mol. Biol. Rev.* **65**, 404–421
- 15 Roach, P. J. (1991) Multisite and hierarchal protein phosphorylation. *J. Biol. Chem.* **266**, 14139–14142
- 16 Bray, D. (1995) Protein molecules as computational elements in living cells. *Nature (London)* **376**, 307–312
- 17 Weng, G., Bhalla, U. S. and Iyengar, R. (1999) Complexity in biological signaling systems. *Science* **284**, 92–96
- 18 Cozzone, A. J. (1988) Protein phosphorylation in prokaryotes. *Annu. Rev. Microbiol.* **42**, 97–125
- 19 Monod, J. (1966) From enzymatic adaptation to allosteric transitions. *Science* **154**, 475–483
- 20 Stadtman, E. R. (1966) Allosteric regulation of enzyme activity. *Adv. Enzymol. Relat. Areas Mol. Biol.* **28**, 41–154
- 21 Sanwal, S. D. (1970) Allosteric controls of amphibolic pathways in bacteria. *Bacteriol. Rev.* **34**, 20–39
- 22 Reichard, R. (2002) Ribonucleotide reductases: the evolution of allosteric regulation. *Arch. Biochem. Biophys.* **397**, 149–155
- 23 LaPorte, D. C. (1993) The isocitrate dehydrogenase phosphorylation cycle: regulation and enzymology. *J. Cell. Biochem.* **51**, 14–18
- 24 Bourret, R. B., Hess, J. F., Borkovich, K. A., Pakula, A. A. and Simon, M. I. (1989) Protein phosphorylation in chemotaxis and two-component regulatory systems of bacteria. *J. Biol. Chem.* **264**, 7085–7088
- 25 Stock, J. B., Ninfa, A. J. and Stock, A. M. (1989) Protein phosphorylation and regulation of adaptive responses in bacteria. *Rev. Microbiol.* **53**, 450–490
- 26 Lim, W. A. (2002) The modular logic of signaling proteins: building allosteric switches from simple binding domains. *Curr. Opin. Struct. Biol.* **12**, 61–68
- 27 Chatton, E. (1973) *Titres et Travaux Scientifiques*. Sete, Sottano, Italy
- 28 de Duve, C. (1996) The birth of complex cells. *Sci. Am.* **274**, 50–57
- 29 Olsen, G. J. and Woese, C. W. (1993) Ribosomal RNA: a key to phylogeny. *FASEB J.* **7**, 113–123
- 30 Woese, C. R. and Wolfe, R. S. (1985) Archaeobacteria: the urkingdom. *The Bacteria* **VIII**, 561–564
- 31 Doolittle, R. F. (1995) The origins and evolution of eukaryotic proteins. *Philos. Trans. R. Soc. London Ser. B* **349**, 235–240
- 32 Kandler, O. (1994) Cell wall biochemistry and the three-domain concept of life. *Syst. Appl. Microbiol.* **16**, 501–509
- 33 Doolittle, W. F. (1999) Lateral genomics. *Trends Cell Biol.* **9**, M5–M8
- 34 Ludwig, W. and Schliefer, K.-H. (1999) Phylogeny of Bacteria beyond the 16S rRNA standard. *ASM News* **65**, 752–757
- 35 Spudich, J. L. and Stoekenius, W. (1980) Light-regulated retinal-dependent reversible phosphorylation of *Halobacterium* proteins. *J. Biol. Chem.* **255**, 5501–5503
- 36 Wang, J. Y. J. and Koshland, Jr, D. E. (1978) Evidence for protein kinase activity in the prokaryote *Salmonella typhimurium*. *J. Biol. Chem.* **253**, 7605–7608
- 37 Spudich, E. A. and Spudich, J. L. (1981) Photosensitive phosphoproteins in halobacteria: regulatory coupling of transmembrane proton flux and protein dephosphorylation. *J. Cell Biol.* **91**, 895–900
- 38 Skorko, R. (1984) Protein phosphorylation in the archaeobacterium *Sulfolobus acidocaldarius*. *Eur. J. Biochem.* **145**, 617–622
- 39 Skorko, R. (1989) Polyphosphate as a source of phosphoryl group in protein modification in the archaeobacterium *Sulfolobus acidocaldarius*. *Biochimie* **71**, 1089–1093
- 40 Osorio, G. and Jerez, C. A. (1996) Adaptive response of the archaeon *Sulfolobus acidocaldarius* BC65 to phosphate starvation. *Microbiology* **142**, 1531–1536
- 41 Smith, S. C., McCartney, B., Kennelly, P. J. and Potts, M. (1997) Protein-tyrosine phosphorylation in the *Archaea*. *J. Bacteriol.* **179**, 2418–2420
- 42 Jeon, S.-J., Fujiwara, S., Takagi, M., Tanaka, T. and Imanaka, T. (2002) *Tk*-PTP, protein tyrosine/serine phosphatase from hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1: enzymatic characteristics and identification of its substrate proteins. *Biochem. Biophys. Res. Commun.* **295**, 508–514
- 43 Levitzki, A. and Gazit, A. (1995) Tyrosine kinase inhibition: an approach to drug development. *Science* **267**, 1782–1788
- 44 Solow, B., Bischoff, K. M., Zylka, M. J. and Kennelly, P. J. (1998) Archaeal phosphoproteins. Identification of a hexosephosphate mutase and the α -subunit of succinyl-CoA synthetase in the extreme acidothermophile *Sulfolobus solfataricus*. *Protein Sci.* **7**, 105–111
- 45 Daas, P. J. H., Wassenaar, R. W., Willemsen, P., Theunissen, R. J., Keltjen, J. T., van der Drift, C. and Vogels, G. D. (1996) Purification and properties of an enzyme involved in the ATP-dependent activation of the methanol:2-mercaptoethanesulfonic acid methyltransferase reaction from *Methanosarcina barkeri*. *J. Biol. Chem.* **271**, 22339–22345
- 46 Condo, I., Ruggero, D., Reinhardt, R. and Londei, P. (1998) A novel aminopeptidase associated with the 60 kDa chaperonin in the thermophilic archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* **29**, 775–785
- 47 Grabowski, B. and Kelman, Z. (2001) Autophosphorylation of archaeal Cdc6 homologues is regulated by DNA. *J. Bacteriol.* **183**, 5459–5464
- 48 Lower, B. H., Bischoff, K. M. and Kennelly, P. J. (2000) The archaeon *Sulfolobus solfataricus* contains a membrane-associated protein kinase activity that preferentially phosphorylates threonine residues. *J. Bacteriol.* **182**, 3452–3459
- 49 Kawarabayasi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A. et al. (1999) Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res.* **6**, 83–101
- 50 Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D. et al. (1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature (London)* **390**, 364–379
- 51 Ng, W. V., Kennedy, S. P., Mahairas, G. G., Berquist, B., Pan, M., Shukla, H. D., Lasky, S. R., Baliga, N. S., Thorsson, V., Sbrogna, J. et al. (2000) Genome sequence of *Halobacterium* species NRC-1. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12176–12181
- 52 Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K. et al. (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* Δ H: functional analysis and comparative genomics. *J. Bacteriol.* **179**, 7135–7155
- 53 Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., Fitzgerald, L. M., Clayton, R. A., Gocayne, J. D. et al. (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058–1073
- 54 Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, S., Baba, S.-I., Kosugi, H., Hosomaya, A. et al. (1998) Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3. *DNA Res.* **5**, 55–76
- 55 She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., De Moors, A. et al. (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7835–7840
- 56 Ruepp, A., Graml, W., Santos-Martinez, M.-L., Koretke, K., Volker, K., Mewes, H. W., Frishman, D., Stocker, S., Lupas, A. N. and Baumeister, W. (2000) The complete genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature (London)* **407**, 508–513
- 57 Kawashima, T., Amano, N., Koike, H., Makino, S.-I., Higuchi, S., Kawashima-Ohya, Y., Watanabe, K., Yamazaki, M., Kanehori, K., Kawamoto, T. et al. (2000) Archaeal adaptation to higher temperatures revealed by genomic sequence of *Thermoplasma volcanium*. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14257–14262
- 58 Koretke, K. K., Lupas, A. N., Warren, P. V., Rosenberg, M. and Brown, J. P. (2000) Evolution of two-component signal transduction. *Mol. Biol. Evol.* **17**, 1956–1970
- 59 Kim, D.-J. and Forst, S. (2001) Genomic analysis of the histidine kinase family in bacteria and archaea. *Microbiology* **147**, 1197–1212
- 60 Saito, H. (2001) Histidine phosphorylation and two-component signaling in eukaryotic cells. *Chem. Rev.* **101**, 2497–2509
- 61 Harris, R. A., Popov, K. M., Zhao, Y., Kedishvili, N. Y., Shimomura, Y. and Crabb, D. W. (1995) A new family of protein kinases – the mitochondrial protein kinases. *Adv. Enzyme Regul.* **35**, 147–162
- 62 Yeh, K.-C. and Lagarias, J. C. (1998) Eukaryotic phytochromes: light-regulated serine/threonine kinases with histidine kinase ancestry. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13976–13981
- 63 Min, K.-T., Hilditch, C. M., Diederich, B., Errington, J. and Yudkin, M. D. (1993) Sigma F, the first compartment-specific transcription factor of *B. subtilis*, is regulated by an anti-sigma factor that is also a protein kinase. *Cell* **74**, 735–742

- 64 Wu, J., Ohta, N., Zhao, J.-L. and Newton, A. (1999) A novel bacterial tyrosine kinase essential for cell division and differentiation. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13068–13073
- 65 Shi, L., Bischoff, K. M. and Kennelly, P. J. (1999) The *icfG* gene cluster of *Synechocystis* sp. PCC6803 encodes an Rsb/Spo-like protein kinase, protein phosphatase, and two phosphoproteins. *J. Bacteriol.* **181**, 4761–4767
- 66 Hess, J. F., Oosawa, K., Kaplan, N. and Simon, M. I. (1988) Phosphorylation of three proteins involved in bacterial chemotaxis. *Cell* **53**, 79–87
- 67 Keener, J. and Kustu, S. (1988) Protein kinase and phosphoprotein phosphatase activities of nitrogen regulatory proteins NTRB and NTRC of enteric bacteria: roles of conserved amino-terminal domain of NTRC. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4976–4980
- 68 Weiss, V. and Magasanik, B. (1988) Phosphorylation of nitrogen regulator I (NR_I) of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8919–8923
- 69 Igo, M. M., Ninfa, A. J., Stock, J. B. and Silhavy, T. J. (1989) Phosphorylation and dephosphorylation of a bacterial transcriptional activator by a transmembrane receptor. *Genes Dev.* **3**, 1725–1734
- 70 Perego, M. (1998) Kinase-phosphatase competition regulates *Bacillus subtilis* development. *Trends Microbiol.* **6**, 366–370
- 71 Reizer, J., Reizer, A., Perego, A. and Saier, Jr, M. H. (1997) Characterization of a family of bacterial response regulator aspartyl-phosphate (RAP) phosphatases. *Microb. Comp. Genomics* **2**, 103–111
- 72 Taylor, S. S., Knighton, D. R., Zheng, J., Sowadski, J. M., Gibbs, C. S. and Zoller, M. J. (1993) A template for the protein kinase family. *Trends Biochem. Sci.* **18**, 84–89
- 73 Hanks, S. K. and Hunter, T. (1995) The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* **9**, 576–596
- 74 Johnson, M. (1996) Genome sequencing: the complete code for a eukaryotic cell. *Curr. Biol.* **6**, 500–503
- 75 Plowman, G. D., Sudarsanam, S., Bingham, J. and Hunter, T. (1999) The protein kinases of *Caenorhabditis elegans*: a model for signal transduction in multicellular organisms. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13603–13610
- 76 Rubin, G. M., Yandell, M. D., Wortman, J. R., Miklos, G. L. G., Nelson, C. R., Hariharan, I. K., Fortini, M. E., Li, P. W., Apweiler, R., Fleischmann, W. et al. (2000) Comparative genomics of the eukaryotes. *Science* **287**, 2204–2215
- 77 Kennelly, P. J. (2002) Protein kinases and protein phosphatases in prokaryotes: a genomic perspective. *FEMS Microbiol. Lett.* **206**, 1–8
- 78 Hanks, S. K. and Lindberg, R. A. (1991) Use of degenerate oligonucleotide probes to identify clones that encode protein kinases. *Methods Enzymol.* **200**, 525–532
- 79 Facchin, S., Sarno, S., Marin, O., Lopreiato, R., Satori, G. and Pinna, L. A. (2002) Acidophilic character of yeast PID261/BJD32, a putative ancestor of eukaryotic protein kinases. *Biochem. Biophys. Res. Commun.* **296**, 1366–1371
- 80 Angermayr, M., Roidl, A. and Bandlow, W. (2002) Yeast Rio1p is the founding member of a novel subfamily of protein serine kinases involved in the control of cell cycle progression. *Mol. Microbiol.* **44**, 309–324
- 81 Zhang, W., Inouye, M. and Inouye, S. (1996) Reciprocal regulation of the differentiation of *Myxococcus xanthus* by Pkn5 and Pkn6, eukaryotic-like Ser/Thr protein kinases. *Mol. Microbiol.* **20**, 435–447
- 82 Shi, L., Potts, M. and Kennelly, P. J. (1998) Protein kinases and protein phosphatases in prokaryotes: a family portrait. *FEMS Microbiol. Rev.* **22**, 229–253
- 83 Hon, W.-C., McKay, G. A., Thompson, P. R., Sweet, R. M., Yang, D. S., Wright, G. D. and Berghuis, A. M. (1997) Structure of an enzyme required for aminoglycoside antibiotic resistance reveals homology to eukaryotic protein kinases. *Cell* **89**, 887–895
- 84 Krupa, A. and Srinivasan, N. (2002) Lipopolysaccharide phosphorylating enzymes encoded in the genomes of Gram-positive bacteria are related to the eukaryotic protein kinases. *Protein Sci.* **11**, 1580–1584
- 85 Shaw, K. J., Rather, P. N., Hare, R. S. and Miller, G. H. (1993) Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes. *Microbiol. Rev.* **57**, 138–163
- 86 Leonard, C. J., Aravind, L. and Koonin, E. V. (1998) Novel families of putative protein kinases in *Bacteria* and *Archaea*: Evolution of the “eukaryotic” protein kinase superfamily. *Genome Res.* **8**, 1038–1047
- 87 Ponting, C. P., Aravind, L., Schultz, J., Bork, P. and Koonin, E. V. (1999) Eukaryotic signalling domain homologues in *Archaea* and *Bacteria*. Ancient ancestry and horizontal gene transfer. *J. Mol. Biol.* **289**, 729–745
- 88 Barford, D. (1996) Molecular mechanisms of the protein serine/threonine phosphatases. *Trends Biochem. Sci.* **21**, 407–412
- 89 Barton, G. J., Cohen, P. T. W. and Barford, D. (1994) Conservation analysis and structure prediction of protein serine/threonine phosphatases. Sequence similarity with diadenosine tetraphosphate from *Escherichia coli* suggests homology to protein phosphatases. *Eur. J. Biochem.* **220**, 225–237
- 90 Cohen, P. T. W. (1999) Novel protein serine/threonine phosphatases: variety is the spice of life. *Trends Biochem. Sci.* **22**, 245–251
- 91 Missiakis, D. and Raina, S. (1997) Signal transduction pathways in response to protein misfolding in the extracytoplasmic compartments of *E. coli*: Role of two new phosphoprotein phosphatases. *EMBO J.* **16**, 1670–1685
- 92 Shi, L., Carmichael, W. W. and Kennelly, P. J. (1999) Cyanobacterial PPP-family protein phosphatases possess multifunctional capabilities and are resistant to microcystin-LR. *J. Biol. Chem.* **274**, 10039–10046
- 93 Cohen, P. T. W. and Cohen, P. (1989) Discovery of a protein phosphatase activity encoded in the genome of bacteriophage λ . Probable identity with open reading frame 221. *Biochem. J.* **260**, 931–934
- 94 Bork, P., Brown, N. P., Hegyi, H. and Schultz, J. (1996) The protein phosphatase 2C (PP2C) superfamily: detection of bacterial homologues. *Protein Sci.* **5**, 1421–1425
- 95 Yang, X., Kang, C. M., Brody, M. S. and Price, C. W. (1996) Opposing pairs of serine protein kinases and phosphatases transmit signals of environmental stress to activate a bacterial transcription factor. *Genes Dev.* **10**, 2265–2275
- 96 Adler, E., Donella-Deana, A., Arigoni, F., Pinna, L. A. and Stragier, P. (1997) Structural relationship between a bacterial developmental protein and eukaryotic PP2C protein phosphatases. *Mol. Microbiol.* **23**, 57–62
- 97 Treuner-Lange, A., Ward, M. J. and Zusman, D. R. (2001) Pph1 from *Myxococcus xanthus* is a protein phosphatase involved in vegetative growth and development. *Mol. Microbiol.* **40**, 126–140
- 98 Irmeler, A. and Forchhammer, K. (2001) A PP2C-type phosphatase dephosphorylates the PII signaling protein in the cyanobacterium *Synechocystis* PCC6803. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12978–12983
- 99 Lohse, D. L., Denu, J. M. and Dixon, J. E. (1995) Insights derived from the structures of the Ser/Thr phosphatases calcineurin and protein phosphatase 1. *Structure* **3**, 987–990
- 100 Solow, B., Young, J. C., White, R. H. and Kennelly, P. J. (1997) Gene cloning and expression and characterization of a toxin-sensitive protein phosphatase from the methanogenic archaeon *Methanosarcina thermophila* TM-1. *J. Bacteriol.* **179**, 5072–5075
- 101 Mai, B., Frey, G., Swanson, R. V., Mathur, E. J. and Stetter, K. O. (1998) Molecular cloning and functional expression of a protein-serine/threonine phosphatase from the hyperthermophilic archaeon *Pyrodicticum abyssi* TAG11. *J. Bacteriol.* **180**, 4030–4035
- 102 Zhang, Z.-Y. (1997) Structure, mechanism and specificity of protein-tyrosine phosphatases. *Curr. Top. Cell Regul.* **35**, 21–68
- 103 Denu, J. M. and Dixon, J. E. (1998) Protein tyrosine phosphatases: mechanisms of catalysis and regulation. *Curr. Biol.* **2**, 633–641
- 104 Barford, D., Jia, Z. and Tonks, N. K. (1995) Protein tyrosine phosphatases take off. *Nat. Struct. Biol.* **2**, 1043–1053
- 105 van Huijsduijnen, R. H. (1998) Protein tyrosine phosphatases: counting trees in the forest. *Gene* **225**, 1–8
- 106 Martell, K. J., Angelotti, T. and Ullrich, A. (1998) The “VH1-like” protein tyrosine phosphatases. *Mol. Cell* **8**, 2–11
- 107 Maehama, T. and Dixon, J. E. (1998) The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. *J. Biol. Chem.* **273**, 13375–13378
- 108 Taylor, G. S., Maehama, T. and Dixon, J. E. (2000) Myotubularin, a protein tyrosine phosphatase mutated in myotubular myopathy, dephosphorylates the lipid second messenger, phosphatidylinositol 3-phosphate. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8910–8915
- 109 Takagi, T., Moore, C. R., Diehn, F. and Buratowski, S. (1997) An RNA 5′triphosphate related to the protein tyrosine phosphatases. *Cell* **89**, 867–873
- 110 Ramponi, G. and Stefani, M. (1997) Structure, catalytic and functional properties of low M_r phosphotyrosine protein phosphatases. Evidence for a long evolutionary history. *Int. J. Biochem. Cell. Biol.* **29**, 270–292
- 111 Li, Y. and Strohl, W. R. (1996) Cloning, purification and properties of a phosphotyrosine phosphatase from *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **178**, 136–142
- 112 Guan, K.-L. and Dixon, J. E. (1990) Protein tyrosine phosphatase activity of an essential virulence determinant in *Yersinia*. *Science* **249**, 553–556
- 113 Potts, M., Sun, H., Mockaitis, K., Kennelly, P. J. and Tonks, N. K. (1993) A protein-tyrosine/serine phosphatase encoded by the genome of the cyanobacterium *Nostoc commune* UTEX 584. *J. Biol. Chem.* **268**, 7632–7635
- 114 Liu, J. and Rosen, B. P. (1997) Ligand interactions of the ArsC arsenate reductase. *J. Biol. Chem.* **272**, 21084–21089
- 115 Zegers, I., Martins, J. C., Willem, R., Wyns, L. and Messens, J. (2001) Arsenate reductase from *S. aureus* plasmid pI258 is a phosphatase drafted for redox duty. *Nat. Struct. Biol.* **8**, 843–847
- 116 Nilsson, I. and Hoffmann, I. (2000) Cell cycle regulation by the Cdc25 phosphatase family. *Prog. Cell Cycle Res.* **4**, 107–114

- 117 Fauman, E. B., Cogswell, J. P., Lovejoy, B., Rocque, W. J., Holmes, W., Montana, V. G., Piwnica-Worms, H., Rink, M. J. and Saper, M. A. (1998) Crystal structure of the catalytic domain of the human cell cycle control phosphatase Cdc25A. *Cell* **93**, 617–625
- 118 Galinier, A., Kravajna, M., Engelmann, R., Hengstenberg, W., Kilhoffer, M.-C., Deutscher, M. and Haiech, J. (1998) New protein kinase and protein phosphatase families mediate signal transduction in bacterial catabolite repression. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1823–1828
- 119 Reizer, J., Hoischen, C., Titgemeyer, F., Rivolta, C., Rabus, R., Stulke, J., Karamata, D., Saier, Jr, M. H. and Hillen, W. (1998) A novel protein kinase that controls carbon catabolite repression in bacteria. *Mol. Microbiol.* **27**, 1157–1169
- 120 Clancy, C. E., Mendoza, M. G., Naismith, T. V., Kolman, M. F. and Egelhoff, T. T. (1997) Identification of a protein kinase from *Dictyostelium* with homology to the novel catalytic domain of myosin heavy chain kinase A. *J. Biol. Chem.* **272**, 11812–11815
- 121 Andersson, S. G. E. and Kurland, C. G. (1998) Reductive evolution of resident genomes. *Trends Microbiol.* **6**, 263–268
- 122 Rudolph, J. and Oesterhelt, D. (1995) Chemotaxis and phototaxis require a CheA histidine kinase in the archaeon *Halobacterium salinarum*. *EMBO J.* **14**, 667–673
- 123 Rudolph, J. and Oesterhelt, D. (1996) Deletion analysis of the *che* operon in the archaeon *Halobium salinarum*. *J. Mol. Biol.* **258**, 548–554
- 124 Rudolph, J., Tolliday, N., Schmitt, C., Schuster, S. C. and Oesterhelt, D. (1999) Phosphorylation in halobacterial signal transduction. *EMBO J.* **14**, 4249–4257
- 125 Jung, K.-H., Spudich, E. N., Trivedi, V. D. and Spudich, J. L. (2001) An archaeal photosignal-transducing module mediates phototaxis in *Escherichia coli*. *J. Bacteriol.* **183**, 6365–6371
- 126 Bischoff, K. M. and Kennelly, P. J. (1999) An 'in-gel' assay for alternative nucleotide donors for protein kinases. *Anal. Biochem.* **271**, 199–202
- 127 Pinna, L. A. (1990) Casein kinase 2: an 'eminence grise' in cellular regulation? *Biochim. Biophys. Acta* **1054**, 267–284
- 128 Lower, B. H. and Kennelly, P. J. (2002) The membrane-associated protein-serine/threonine kinase from *Sulfolobus solfataricus* is a glycoprotein. *J. Bacteriol.* **184**, 2614–2619
- 129 Schaffer, C. and Messner, P. (2001) Glycobiology of surface layer proteins. *Biochimie* **83**, 591–599
- 130 Burda, P. and Aebi, M. (1999) The dolichol pathway of N-linked glycosylation. *Biochim. Biophys. Acta* **1426**, 239–257
- 131 Goldman, S., Hecht, K., Eisenberg, H. and Mevarech, M. (1990) Extracellular Ca^{2+} -dependent inducible alkaline phosphatase from the extremely halophilic archaeobacterium *Haloarcula marismortui*. *J. Bacteriol.* **172**, 7065–7070
- 132 Eichler, J. (2000) Novel glycoproteins of the halophilic archaeon *Haloferax volcanii*. *Arch. Microbiol.* **173**, 445–448
- 133 Elfernick, M. G. L., Albers, S.-V., Konings, W. N. and Driessen, A. J. M. (2001) Sugar transport in *Sulfolobus solfataricus* is mediated by two families of binding protein-dependent ABC transporters. *Mol. Microbiol.* **39**, 1494–1503
- 134 Moens, S. and Vanderleyden, J. (1997) Glycoproteins in prokaryotes. *Arch. Microbiol.* **168**, 169–175
- 135 Lubben, M. and Schafer, G. (1989) Chemiosmotic energy conversion of the archaeobacterial thermoacidophile *Sulfolobus acidocaldarius*: oxidative phosphorylation and the presence of an F_0 -related N,N' -dicyclohexylcarbodiimide-binding proteolipid. *J. Bacteriol.* **171**, 6106–6116
- 136 Brock, T. D., Brock, K. M., Belly, R. T. and Weiss, R. L. (1972) *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. Microbiol.* **84**, 54–68
- 137 Johnson, L. N., Noble, M. E. M. and Owen, D. J. (1996) Active and inactive protein kinases: structural basis for regulation. *Cell* **85**, 149–158
- 138 Adams, J. A. (2001) Kinetic and catalytic mechanisms of protein kinases. *Chem. Rev.* **101**, 2271–2290
- 139 Huse, M. and Kuriyan, J. (2002) The conformational plasticity of protein kinases. *Cell* **109**, 275–282
- 140 Kennelly, P. J., Oxenrider, K. A., Leng, J., Cantwell, J. S. and Zhao, N. (1993) Identification of a serine/threonine-specific protein phosphatase from the archaeobacterium *Sulfolobus solfataricus*. *J. Biol. Chem.* **268**, 6505–6510
- 141 Leng, J., Cameron, A. J., Buckel, S. and Kennelly, P. J. (1995) Isolation and cloning of a protein-serine/threonine phosphatase from an archaeon. *J. Bacteriol.* **177**, 6510–6517
- 142 Oxenrider, K. A., Rasche, M. E., Thorsteinsson, M. V. and Kennelly, P. J. (1993) An okadaic acid-sensitive protein phosphatase from the archaeon *Methanosarcina thermophila* TM-1. *FEBS Lett.* **331**, 291–295
- 143 Cohen, P. (1991) Classification of protein-serine/threonine phosphatases: identification and quantitation in cell extracts. *Methods Enzymol.* **201**, 389–398
- 144 MacKintosh, C. and MacKintosh, R. W. (1994) Inhibitors of protein kinases and phosphatases. *Trends Biochem. Sci.* **19**, 444–448
- 145 Oxenrider, K. A. and Kennelly, P. J. (1993) A protein-serine phosphatase from the halophilic archaeon *Haloferax volcanii*. *Biochem. Biophys. Res. Commun.* **194**, 1330–1335
- 146 Bonet, M. L., Llorca, F. I. and Cadenas, E. (1991) Purification and some properties of an atypical alkaline p-nitrophenylphosphate phosphatase activity from *Halobacterium halobium*. *Int. J. Biochem.* **23**, 1445–1551
- 147 Millward, T. A., Zolnierowicz, S. and Hemmings, B. A. (1999) Regulation of protein kinase cascades by protein phosphatase 2A. *Trends Biochem. Sci.* **24**, 186–191
- 148 Aggen, J. B., Nairn, A. C. and Chamberlin, R. (2000) Regulation of protein phosphatase-1. *Chem. Biol.* **7**, R13–R23
- 149 Virshup, D. M. (2000) Protein phosphatase 2A: a panoply of enzymes. *Curr. Opin. Cell Biol.* **12**, 180–185
- 150 Miller, S. L. (1953) A production of amino acids under possible primitive earth conditions. *Science* **117**, 528–529
- 151 Cech, T. R. (1993) The efficiency and versatility of catalytic RNA: implications for an RNA world. *Gene* **135**, 33–36
- 152 Szathmari, E. (1999) The origin of the genetic code: amino acids as cofactors in an RNA world. *Trends Genet.* **15**, 223–229
- 153 Peterson, J. D., Umayam, L. A., Dickinson, T. M., Hickey, E. K. and White, O. (2001) The comprehensive microbial resource. *Nucleic Acids Res.* **29**, 123–125
- 154 Fauman, E. B. and Saper, M. A. (1996) Structure and function of protein tyrosine phosphatases. *Trends Biochem. Sci.* **21**, 413–417
- 155 Woese, C. R. (2002) On the evolution of cells. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8742–8747
- 156 Das, A. K., Helps, N. R., Cohen, P. T. W. and Barford, D. (1996) Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 Å resolution. *EMBO J.* **15**, 6798–6809

Received 4 October 2002/20 November 2002; accepted 22 November 2002

Published as BJ Immediate Publication 22 November 2002, DOI 10.1042/BJ20021547