Posttranslational Protein Modification in *Archaea*

Jerry Eichler¹* and Michael W. W. Adams²

Department of Life Sciences, Ben Gurion University, Beersheva, 84105 Israel,¹ and Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602²

INTRODUCTION ............................................. 394

PROTEIN GLYCOXYLATION ................................................ 395
Glycosylated Archaeal Proteins ................................................. 395
S-layer glycoproteins ............................................................... 395
(i) S-layer glycoproteins reveal unique aspects of archaeal protein glycosylation .............................................. 395
Flagellins ........................................................................ 397
(ii) Evidence for flagellin glycosylation ................................. 397
Other proteins .................................................................. 397
Process of Protein N-Glycosylation in *Archaea* ..................... 398
Dolichol carrier .................................................................. 399
(i) Antibiotics that affect dolichol processing interfere with archaeal protein glycosylation ................................. 399
(ii) Analysis of dolichol-bound glycans .................................. 399
Enzymes of N-glycosylation ................................................. 399
(i) Genomic studies ......................................................... 399
(ii) Biochemical studies .................................................... 400
Subcellular localization of glycosylation ................................. 401
Role of Protein Glycosylation in *Archaea* ......................... 401
Structural roles ................................................................ 401
Functional roles ................................................................ 401
Glycosylation as an environmental adaptation ....................... 401
LIPID MODIFICATION ........................................... 402
Membrane Lipids of *Archaea* ............................................ 402
Lipid-Modified Archaeal Proteins ........................................ 402
Lipoproteins ................................................................ 402
Isoprenylated proteins ...................................................... 404
Acylated proteins ............................................................ 404
GPI-anchored proteins ...................................................... 404
PROTEIN PHOSPHORYLATION ...................................... 404
Targets and Functions of Protein Phosphorylation in *Archaea* 405
Phosphorylation of components involved in signal transduction 405
Phosphorylation of components involved in DNA replication, cell cycle regulation, and translation .................. 405
Phosphorylation of other proteins ....................................... 406
Archaeal Protein Kinases and Phosphatases ......................... 406
Eucaryal protein kinases .................................................. 406
Histidine kinases ........................................................... 406
Protein serine/threonine phosphatases ............................... 407
Protein tyrosine phosphatases ......................................... 407
GPI-anchored protein phosphatases of *Thermoplasma acidophilum* ................................................................. 407
PROTEIN METHYLATION ........................................ 407
Protein Methylation in Response to External Stimuli ................ 407
Methylation of Methyl-Coenzyme M Reductase .................... 408
Methylated Proteins in Thermophilic *Archaea* .................... 408
Methylation of Archaeal DNA-Binding Proteins ................... 409
Methylation of Archaeal Ribosomal Proteins ....................... 409
DISULFIDE BONDS IN PROTEINS ................................. 409
Disulfide Bonds in Cytoplasmic Archaeal Proteins ............ 409
Disulfide Bonds in Extracellular Archaeal Proteins ............. 410
Enzymes Involved in Disulfide Bond Formation in *Archaea* 410
PROTEOLYTICALLY PROCESSED PROTEINS ............ 411
Archaeal Signal Sequences .............................................. 411
Protein translocation in *Archaea* ................................ 411

* Corresponding author. Mailing address: Dept. of Life Sciences, Ben Gurion University, P.O. Box 653, Beersheva 84105, Israel. Phone: 972 8646 1343. Fax: 972 8647 9175. E-mail: jeichler@bgu.ac.il.
INTRODUCTION

With complete genome sequences appearing at an ever more rapid rate, attention is becoming increasingly directed towards describing the protein complement of a given organism, i.e., the proteome. Studies of proteins conducted both at the level of the individual polypeptide and cellwide have revealed that the repertoire of expressed proteins can expand beyond what is predicted by direct translation of the complement of open reading frames contained within a genome. For example, the proteome can assume additional levels of complexity with differential expression of individual polypeptides or members of protein families as a function of developmental stage or in response to environmental cues. The various permutations of protein-protein interactions possible further expand the complexity of the proteome. However, one of the most important and fundamental aspects of proteomic complexity comes from the various processing events that many proteins experience following their synthesis, i.e., posttranslational modification.

Proteins can be modified posttranslationally by covalent attachment of one or more of several classes of molecules, by the formation of intra- or intermolecular linkages, by proteolytic processing of the newly synthesized polypeptide chain, or by any combination of these events. By chemically linking various modifying groups either permanently or temporarily and by allowing for changes in the molecular composition of the modifying moieties, covalent modifications can endow proteins with properties that are very different from those that are predicted by the encoding genes. Examples of such covalent modifications include glycosylation, lipid attachment, phosphorylation, and methylation.

The covalent bonding of pairs of Cys residues to form disulfide bridges not only modulates the three-dimensional conformation of a polypeptide chain but can also be used to maintain proteins in multisubunit complexes. Controlled reduction and reoxidation of protein disulfide bonds is also employed in electron transfer reactions fundamental to many cellular processes. Proteolytic processing of newly synthesized polypeptide chains similarly allows the cell to control the folding and function of a protein. By removing specific targeting sequences or other stretches of amino acid residues, the cell is able to control where, when, and how a protein will act. As such, posttranslational modifications can significantly modulate the physicochemical and biological properties of a protein through effects on protein function, subcellular localization, oligomerization, folding, or turnover. The distribution of posttranslational modifications and their effects on protein chemistry and cell biology become even broader when one also considers the effects of additional, secondary posttranslational modification steps such as the addition of organic (e.g., flavins) or inorganic (e.g., metal groups) cofactors. Such modifications, however, lie beyond the scope of this review.

Long-known to be widespread in Eucarya and Bacteria, it is becoming clear that posttranslational modification of proteins also takes place in Archaea. Best known in their capacities as extremophiles, i.e., microorganisms able to thrive in the harshest environmental conditions on this planet, Archaea express proteins that enable them to succeed in such habitats. Indeed, archaeal proteins are able to remain properly folded and functional in the face of extremes of salinity, temperature, and other adverse physical conditions that would normally lead to protein denaturation, loss of solubility, and aggregation. Although posttranslational modifications may help archaeal proteins overcome the challenges presented by their surroundings, in most cases, the reason for posttranslational modification of a particular archaeal protein remains unclear. Table 1 lists the posttranslational modifications that archaeal proteins may experience.

Analysis of the various posttranslational modifications experienced by archaeal proteins has served to reveal not only novel protein modifications not previously observed in Eucarya or Bacteria but also variations of previously characterized posttranslational modifications. By and large, however, archaeal
posttranslational modifications often resemble their eucaryal or bacterial counterparts. Hence, elucidating such similarities provides insight into evolutionary relationships across the three domains of life. Moreover, the mosaic profile of eucaryal, bacterial, and archaean traits that describes posttranslational protein modification in *Archaea* also holds true when one examines the enzymes and mechanistic steps involved in archaean protein modification processes. Here too, examination of archaean systems has served to expand our understanding of natural pathways or to underscore the similarities between archaean, eucaryal, and/or bacterial biology. Nonetheless, numerous aspects of archaean posttranslational processing remain poorly described. In the following review, what is currently known of posttranslational protein modification in *Archaea* is considered.

**PROTEIN GLYCOSYLATION**

One of the more prevalent posttranslational modifications experienced by eucaryal proteins is glycosylation. Indeed, protein glycosylation, which begins in the lumen of the endoplasmic reticulum and continues in the Golgi apparatus, is thought to be experienced by more than half of all eucaryal proteins (12). Upon translocation into the endoplasmic reticulum, proteins can be N-glycosylated, when branched oligosaccharide trees of 14 subunits are initially added to selected Asn residues. O-glycosylation of Ser or Thr residues usually takes place in the Golgi. In *Eucarya*, the glycan moieties of glycosylated proteins fulfill a multitude of roles related to protein solubility, folding, stability and turnover, and subcellular localization as well as participating in numerous recognition events (46, 157, 333, 409, 448). Long believed to be an exclusively eucaryal trait, it is now clear that both *Bacteria* and *Archaea* are also capable of attaching glycan moieties to selected proteins (285, 292, 381, 425, 445, 456). A list of those archaean strains reported to contain glycosylated proteins is provided in Table 2.

**Glycosylated Archaean Proteins**

*S-layer glycoproteins.* The surface (S)-layer glycoprotein of the halophilic archaean *Halobacterium salinarum* was the first prokaryotic glycoprotein to be described in detail (246, 283). Subsequently, S-layer glycoproteins have been studied in numerous prokaryotes (292, 381–383). Serving as the main, if not sole, component of the protein layer surrounding many archaean cells (101, 382, 383) (Fig. 1), S-layer glycoproteins remain among the best-characterized archaean glycoproteins. Indeed, examination of the processes used for glycosylation of archaean S-layer glycoproteins has not only served to enhance our understanding of prokaryotic cell surface biogenesis but has also provided insight into the general phenomenon of protein glycosylation in *Archaea*.

While the glycosylated nature of S-layer proteins has been proposed in many archaean species, experimental proof for this posttranslational modification is limited to the S-layer glycoproteins of *Halobacterium salinarum* (246), *Halofexax volcanii* (421), *Halocarcna japonica* (299), *Methanothermus fervidus* (204), *Methanothermus sociabilis* (41), *Sulfolobus* spp. (146), and components of the S-layer of *Staphylothermus marinus* (345). Although the experimental evidence for glycosylation, ranging from chemical characterization of the bound glycan moieties to glycol staining, is stronger in some cases than others, it has been calculated that these S-layer glycoproteins experience an overall degree of glycosylation of up to 15% (292, 382).

Like eucaryal glycoproteins, archaean S-layer glycoproteins can undergo both N- and O-glycosylation. In contrast, bacterial S-layer glycoproteins contain only O-linked glycans (285, 445), although examples of N-glycosylation of other bacterial proteins have been shown (107, 425, 456). Analysis of the composition of the N-linked glycan moieties of archaean S-layer glycoproteins has revealed the wide variety of saccharides available for protein glycosylation in *Archaea*, including galactofuranose, galactouronic acid, glucose, glucuronic acid, iduronic acid, mannose, *N*-acyethylglucosamine, *N*-acyethylglucosamine, and rhamnose (204, 280, 335, 421, 456). In many cases, these sugar subunits may themselves be modified by methylation or sulfation. Such diversity in the range of saccharides used in archaean S-layer glycoprotein N-glycosylation exceeds that seen in the bacterial and eucaryal N-glycosylation processes (425, 456).

(i) *S-layer glycoproteins reveal unique aspects of archaean protein glycosylation.* Despite the proposed evolution of the eucaryal N-glycosylation system from a precursor process in *Archaea* (46, 157), studies of archaean S-layer glycoprotein glycosylation, and in particular glycosylation of the *Halobacterium salinarum* S-layer glycoprotein, have revealed differences in

---

**TABLE 2. Archeaal reported to contain glycoproteins**

<table>
<thead>
<tr>
<th>Species</th>
<th>Evidence for glycosylation</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Halocarcna japonica</em></td>
<td>G</td>
<td>299</td>
</tr>
<tr>
<td><em>Halocarcna marismortui</em></td>
<td>C</td>
<td>136</td>
</tr>
<tr>
<td><em>Halobacterium saccharovorum</em></td>
<td>E</td>
<td>389</td>
</tr>
<tr>
<td><em>Halobacterium salinarum</em></td>
<td>A, B</td>
<td>246, 280</td>
</tr>
<tr>
<td><em>Halofexax mediterranei</em></td>
<td>E</td>
<td>232</td>
</tr>
<tr>
<td><em>Halofexax volcanii</em></td>
<td>A, B, D</td>
<td>98, 421</td>
</tr>
<tr>
<td><em>Methanothermus fervidus</em></td>
<td>D, F</td>
<td>219</td>
</tr>
<tr>
<td><em>Methanococcus delais</em></td>
<td>F, H</td>
<td>27</td>
</tr>
<tr>
<td><em>Methanococcus mazei</em></td>
<td>D</td>
<td>481</td>
</tr>
<tr>
<td><em>Methanococcus voltae</em></td>
<td>A, B</td>
<td>453</td>
</tr>
<tr>
<td><em>Methanosetae soehngenii</em></td>
<td>A</td>
<td>340</td>
</tr>
<tr>
<td><em>Methanosipillum hungatii</em></td>
<td>E, F</td>
<td>406</td>
</tr>
<tr>
<td><em>Methanothermalus fervidus</em></td>
<td>A, B, E</td>
<td>196, 204</td>
</tr>
<tr>
<td><em>Methanothermalus sociabilis</em></td>
<td>A, B</td>
<td>41</td>
</tr>
<tr>
<td><em>Natrálba magadíi</em></td>
<td>E</td>
<td>197</td>
</tr>
<tr>
<td><em>Pyrococcus furiosus</em></td>
<td>C, D, E</td>
<td>44, 230, 231, 455, 464</td>
</tr>
<tr>
<td><em>Sulfolobus acidocaldarius</em></td>
<td>A, B, E, G, H</td>
<td>146, 147, 161, 258, 286</td>
</tr>
<tr>
<td><em>Sulfolobus shibutae</em></td>
<td>E</td>
<td>112</td>
</tr>
<tr>
<td><em>Sulfolobus sofutaricus</em></td>
<td>D, E</td>
<td>106, 146, 262</td>
</tr>
<tr>
<td><em>Staphylothermus marinus</em></td>
<td>C, F</td>
<td>345</td>
</tr>
<tr>
<td><em>Thermococcus litoralis</em></td>
<td>E</td>
<td>44, 145</td>
</tr>
<tr>
<td><em>Thermococcus stetteri</em></td>
<td>E</td>
<td>184</td>
</tr>
<tr>
<td><em>Thermoplasma acidophilum</em></td>
<td>A, B</td>
<td>478</td>
</tr>
<tr>
<td><em>Thermoplasma volcanium</em></td>
<td>E</td>
<td>112</td>
</tr>
</tbody>
</table>

* A, saccharide-amino acid linkage determined; B, glycan structure determined; C, saccharide content determined; D, lectin binding; E, glycan staining; F, deglycosylation; G, inhibition of glycosylation; H, aberrant SDS-PAGE migration. Where A and/or B is true for a given glycoprotein, other evidence for glycosylation is not listed for that case. See text for details.
N-glycosylation in the two domains. Such differences are reflected, for example, in the failure thus far to detect antennary structures in *Archaea* similar to those employed in eucaryal protein N-glycosylation (46, 157, 235, 333, 409, 442), or in the identified amino acid sequence motifs recognized by the archaean N-glycosylation machinery.

It was observed that replacement of the Ser residue of the Asn-2-Ala-3-Ser-4 sequence of the *Halobacterium salinarum* S-layer glycoprotein with Val, Leu, or Asn did not prevent N-glycosylation at the Asn-2 position (486). By contrast, the eucaryal system almost invariably recognizes the Asn-X-Ser/Thr sequence motif, where X is any residue apart from Pro (46, 157, 235, 333, 409, 442), although a rare exception of N-glycosylation at an Asn-Gly-Gly-Thr motif has been reported (211). The ability of *Archaea* to glycosylate proteins at Asn residues that are not part of the consensus Asn-X-Ser/Thr motif suggests that predictions of the glycosylation status of archaean proteins may have overlooked similar or novel N-glycosylation sites. Moreover, the finding that the repeating sulfated pentasaccharide moiety attached at the Asn-2 position of the *Halobacterium salinarum* S-layer glycoprotein through an N-acetylgalactosamine link is chemically distinct from the sulfated polysaccharide unit attached via glucose subunits found at the other ten N-glycosylation sites in the S-layer glycoprotein (247) implies the existence of two different N-saccharyl-transferases in this species. At present, it remains unclear how the cell would distinguish between the different N-glycosylation sites.

Finally, the linkage of glycan moieties to the *Halobacterium salinarum* S-layer glycoprotein at selected Asn residues through either N-acetylgalactosamine or glucose subunits (335) is in contrast to the N-acetylglucosamine linkage largely employed in eucaryal N-glycosylation (46, 157, 235, 333, 409, 442). In the case of the eucaryal protein laminin, however, N-glycosylation involves a β-glucosyl-Asn protein linkage (385). It is of note that laminin is a component of the extracellular basement membrane, a structural layer surrounding mammalian cells in a manner reminiscent of the archaean S-layer.

In addition to N-glycosylation, archaean S-layer glycoproteins can also be modified by O-glycosylation of selected Ser or Thr residues. In both *Halobacterium salinarum* and *Haloferax volcanii*, Thr-rich regions adjacent to the membrane-spanning domain of the protein are decorated at numerous positions with galactose-glucose disaccharides (283, 421). Interestingly, a glycoprotein isolated from a eucaryal basement membrane contains a similar disaccharide (254). Presently, little is known of the steps involved in archaean O-glycosylation or the relation of such steps to the parallel eucaryal or bacterial processes.

![FIG. 1. Schematic depiction of the glycosylation of the *Halobacterium salinarum* S-layer glycoprotein. The topology of the S-layer glycoprotein, the positions of the 11 Asn residues that undergo N-glycosylation, and the heavily O-glycosylated Thr-rich region between Thr-755 and Thr-779 are indicated (246). The inset shows the composition of the three oligosaccharide moieties bound to the protein (247). Abbreviations used: G, glucose; GA, glucuronic acid; Gal, galactose; GalA, galacturonic acid; Galf, galactofuranose; GalN, N-acetylgalactosamine; GN, N-acetylglucosamine; OMe, O-methyl; SO₄, sulfate. Approximately a third of the glucuronic acid residues may be replaced by iduronic acid.](http://mmbr.asm.org/)
Flagellins. In Archaea, cell motility mediated by flagella has been reported for representatives of the major phenotypic groups, i.e., the halophiles, the methanogens, the thermophiles, and the hyperthermophiles, largely based on microscopic investigation (20, 184, 436). Although fulfilling similar roles, archaeal flagella bear little resemblance to their better-characterized bacterial counterparts (7, 265) in terms of structure or assembly. Such differences become evident when one considers the flagellar filament in the two domains. Ultrastructural studies have shown that, unlike bacterial filaments, archaeal flagellar filaments are not hollow structures (72) and that the archaeal structures are generally thinner than their bacterial counterparts (79, 185, 190, 406).

Archaeal and bacterial flagella also differ at the level of flagellin, the major structural component of the flagellar filament. Whereas bacterial flagella are, for the most part, composed of a single type of flagellin, archaeal flagellar filaments are made up of several types of flagellins (with the possible exception of Sulfolobus solfataricus, where genome annotation efforts have reported the existence of only a single flagellin-encoding gene) (20, 184, 436). Indeed, archaeal and bacterial flagellins do not share sequence similarity (19). Moreover, many archaeal flagellins are glycosylated (184), a posttranslational modification that is considered rare for bacterial flagellins (95, 139, 291, 384, 435).

(i) Evidence for flagellin glycosylation. Glycosylation has been reported for flagellins of numerous archaeal strains (112, 184, 196, 197, 389, 436), including Halobacterium salinarum (470), Methanococcus deltae (27), Methanococcus voltae (453), and Methanosparllum hungatei (406). In most of these examples, the evidence for glycosylation comes from studies employing glycans, such as thymol-sulfuric acid or periodic acid-Schiff reagent. Such techniques, however, may not always accurately reflect the glycosylated nature of a protein (222). Hence, additional evidence for glycosylation is desirable.

This has been achieved for the flagellins of Halobacterium salinarum and Methanococcus voltae, for which the chemical compositions of the covalently linked glycan moieties have been elucidated. The Halobacterium salinarum flagellin contains a sulfated glycoconjugate, N-linked through a glucose bridge and based on glucuronic or iduronic acid, similar to the glycan moiety found on the S-layer glycoprotein (420, 468). More recently, Methanococcus voltae flagellins have been shown to contain a novel N-linked trisaccharide (453), despite the fact that earlier glycoprotein-staining-based studies had failed to detect flagellar glycosylation in this species (195). Analysis of trypsin-generated peptides derived from the Methanococcus voltae S-layer glycoprotein also revealed modification by the same novel trisaccharide (453), suggesting a common glycosylation process. Support for the glycosylation of Methanosaprum hungatei flagella by glycan staining was presented by chemical deglycosylation with trifluoromethansulfonic acid, a treatment that decreased molecular mass, as estimated by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) (406). The same was noted for Halobacterium salinarum flagellins upon similar treatment (247).

The glycosylated nature of Methanococcus deltae flagellins was indicated upon incubation of cultures with bacitracin, an antibiotic that interferes with protein glycosylation (see below) (247). Such treatment resulted in more rapid migration of the protein as reflected by SDS-PAGE analysis (27). Similar bacitracin treatment, however, had no effect on the glycosylation of Halobacterium salinarum flagellins, as gauged by migration in SDS-PAGE, although incubation with EDTA, thought to specifically inhibit an externally oriented Mg$^{2+}$-dependent oligosacchoidetransferase (420), successfully modified flagellin migration. By contrast, treating cells with EDTA did lead to the appearance of Methanococcus deltae flagellins of lower apparent molecular weight (27). Together, these observations point to differences in the glycosylation machineries of the two species.

Other proteins. While the bulk of attention on archaeal protein glycosylation has focused on S-layer glycoproteins and flagellins, other archaeal glycoproteins have been identified. Of those additional glycoproteins whose identities are known, the majority are membrane associated. In many instances, these are binding proteins involved in nutrient uptake (see below), such as the maltose/trehalose-binding proteins of Thermococcus litoralis, shown to react with glyco-stain (145) and of Pyrococcus furiosus, shown to contain glucose-containing glycan moieties by lectin binding and molecular analysis (231), or the Pyrococcus furiosus cellobiose-binding protein, which reacts with lectins and glyco-stain (230). Glyco-staining also indicated the glycosylated nature of Pyrococcus furiosus CipA and CipB, two ABC transporter binding proteins whose expression is up-regulated in response to cold shock in this hyperthermophile (464). Glycosylation of pyrulosin, a thermo-stable serine-protease also associated with Pyrococcus furiosus membranes, was proposed on the basis of sequence analysis that revealed the presence of numerous potential N-glycosylation sites and supported by glyco-staining of the protein (455).

Based on lectin binding, a series of glycosylated sugar-binding proteins, apparently containing mannose, glucose, galactose, and N-acetylglucosamine, was detected in Sulfolobus solfataricus membranes (106). Sulfolobus acidocaldarius cytochrome b$\text{558/560}$ was shown to be a highly glycosylated integral membrane protein, containing both O-linked mannosyl hexosamines, and N-linked hexosaccharides (161). Analysis of the composition of the latter glycan moiety revealed the presence of glucose, mannose, and N-acetylglucosamine in addition to 6-sulfoquinovose (484). 6-Sulfoquinovose (or 6-deoxy-6-sulfoglucose) is a rare acidic sugar, commonly found in the glycolipids of chloroplasts and photosynthetic bacteria (177), but not previously found in a glycoprotein. The glycosylated character of a membrane-associated Sulfolobus solfataricus protein serine/threonine kinase was confirmed through precipitation of a protein with kinase activity using lectin-conjugated agarose beads and by the decreased apparent molecular mass of the protein and resistance to glyco-staining following treatment with chemical deglycosylation agents (262).

In addition to membrane proteins, secreted archaeal glycoproteins have also been detected. Lectin binding and chemical deglycosylation confirmed the glycosylated nature of the copper response extracellular proteins secreted by the copper-resistant methanogen Methanobacterium bryantii BKYH (219). Indeed, differential glycosylation is responsible for the appearance of multiple isoforms of the copper response protein. A secreted, inducible alkaline phosphatase purified from Haloar-
**Process of Protein N-Glycosylation in Archaea**

In **Eucarya**, N-glycosylation begins on the cytoplasmic face of the endoplasmic reticulum membrane, where nucleotide-activated monosaccharides are sequentially added by membrane-embedded monosaccharyltransferases to the saturated polyisoprenol-based lipid carrier dolichyl pyrophosphate. This generates the heptasaccharide core of the glycan structure initially found on all eucaryal N-glycosylated proteins (46, 157, 235, 333, 409, 442). Once assembled, the glycan-charged lipid translocates (or “flips”) across the plane of the endoplasmic reticulum membrane bilayer so that the oligosaccharide is now oriented within the endoplasmic reticulum lumen. The translocation of the glycan-charged dolichyl pyrophosphate across the membrane is catalyzed by an ATP-independent flipase (165), identified as the RTF1 protein in *Saccharomyces cerevisiae* (159), with homologues reported in other *Eucarya* (158). Additional sugar subunits are then added to the lipid-bound polysaccharide, transferred from flipped, lumen-facing dolichyl phosphate glucose or mannose carriers (158). The completed oligosaccharide is next transferred to appropriate Asn residues of a nascent polypeptide chain entering the endoplasmic reticulum (46, 157, 235, 333, 409, 442). This is mediated by oligosaccharide transferase, a multisubunit complex associated with the translocon, the membrane protein complex responsible for protein translocation across the endoplasmic reticulum membrane (392).

If, as proposed (46, 157), the elaborate process responsible for protein N-glycosylation in *Eucarya* originated from a simpler archael system, then many of the fundamental steps and central components involved in eucaryal protein N-glycosylation should also be present in *Archaea*. As summarized in Table 3 and discussed in the following section, available evidence suggests that this is indeed the case.

### Table 3. N-glycosylation of proteins across the three domains of life*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Eukarya</th>
<th>Archaea</th>
<th>Bacteria (Campylobacter jejuni)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saccharide donors</td>
<td>ER (Golgi)</td>
<td>Plasma membrane</td>
<td>Plasma membrane</td>
</tr>
<tr>
<td>Lipid carrier</td>
<td>UDP-GlcNAc, GDP-Man, dolicholphosphate-Man/Glc</td>
<td>Dolicholphosphate</td>
<td>Undecaprenolpyrophosphate</td>
</tr>
<tr>
<td>Addition of saccharides following lipid flipping</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Modification of lipid-bound oligosaccharide</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Final oligosaccharide composition</td>
<td>GlcNAc, Man, Glc</td>
<td>Variable</td>
<td>GalNAc, Glc, GalNAc, Bac</td>
</tr>
<tr>
<td>Protein glycosylation motif</td>
<td>Asp-X-Ser/Thr</td>
<td>Asp-X-Ser/Thr/Val/Leu/Asp</td>
<td>Asp-X-Ser/Thr</td>
</tr>
<tr>
<td>Linking sugar</td>
<td>GlcNAc</td>
<td>Variable</td>
<td>GalNAc</td>
</tr>
<tr>
<td>Oligosaccharide-transferring enzyme</td>
<td>Oligosaccharide transferase</td>
<td>STT3 (isoforms?), additional proteins?</td>
<td>PgtB</td>
</tr>
<tr>
<td>Oligosaccharide modification following protein transfer</td>
<td>Yes</td>
<td>?</td>
<td>No</td>
</tr>
</tbody>
</table>

* Abbreviations used: Bac, bacillosamine; Glc, glucose; GalNAc, N-acetylgalactosamine; GlkNAc, N-acetylglucosamine; Man, mannose; ER, endoplasmic reticulum.
Dolichol carrier. Across evolution, isoprene-based lipids play essential roles in the glycosylation process by delivering their bound glycans to selected protein targets (46, 362). In *Archaea*, glucose-, mannose-, N-acetylgalactosamine-, and sulfate tetrasaccharyl-containing phospho- and pyrophosphopolysaccharide (containing 11 to 12 isoprene units) were first observed in *Halobacterium salinarum* by ion exchange and thin-layer chromatography (281). Later studies (248) confirmed the lipid identity of these compounds is C60 dodecaprenol. This is similar to the dolichol used in eucaryal protein N-glycosylation (46) but distinct from undecaprenol, which is composed of 11 unsaturated isoprene units and used by *Bacteria* for protein glycosylation and peptidoglycan synthesis (362, 425). Mass spectrometry and nuclear magnetic resonance-based approaches revealed the presence of *Eucarya*-like sugar carriers in *Halofexa volcanii*, including mannosyl-galactosyl-phosphodolichol, lesser quantities of a dihexosyl-phosphodolichol and a tetrasaccharyl-phosphodolichol containing mannose, galactose, and rhamnose, all linked to a dolichol containing 11 or 12 isoprene units (242).

(i) Antibiotics that affect dolichol processing interfere with archaeal protein glycosylation. The use of various antibiotics and other compounds known to prevent protein glycosylation by interfering with the processing of dolichol carriers has provided insight into the role of this lipid in archaeal protein N-glycosylation. Tunicamycin hinders transfer of UDP-N-acetylgalactosamine to polysaccharide-loaded dolichol carriers (105). Treatment with this antibiotic interferes with *Sulfolobus acidocaldarius* S-layer glycoprotein glycosylation (147). In contrast, tunicamycin has no effect on the biosynthesis of the *Halofexa volcanii* S-layer glycoprotein (99) and accordingly, the glycan moiety of the *Halofexa volcanii* S-layer glycoprotein does not include N-acetylgalactosamine (242, 280). Bacitracin is another drug that interferes with protein glycosylation via an interruption of the recycling of pyrophosphate-containing dolichol species (420). Accordingly, in *Halobacterium salinarum*, bacitracin interferes with the attachment of the repeating sulfated pentasaccharide found at the Asn-2 position of the S-layer glycoprotein (284, 469), although not with the attachment of the sulfated polysaccharide found at the other N-glycosylation sites of the protein (469).

Bacitracin also inhibits glycosylation of flagellins in *Methanococcus deltae* (27) and slowed *Sulfolobus acidocaldarius* growth, possibly through interference with the protein N-glycosylation pathway (286). In contrast, bacitracin had no effect on the glycosylation of the S-layer glycoprotein or a second 98-kDa glycoprotein in *Halofexa volcanii* (99, 232). The failure of the antibiotic to prevent *Halofexa volcanii* glycoprotein biogenesis is likely related to the fact that, unlike *Halobacterium salinarum*, in which both monophosphate- and pyrophosphate-containing dolichol oligosaccharide carriers are present (247), only bacitracin-insensitive monophosphate-containing oligosaccharide-dolichol intermediates are found in *Halofexa volcanii* (242). Incorporation of glucose from UDP-glucose into *Halofexa volcanii* glycoproteins was, however, inhibited by amphotericin and two sugar nucleotide analogs, PP36 and PP55 (489), compounds reported to block transfer of nucleotide-conjugated sugars to phosphopolysaccharins in *Eucarya* (201, 202, 336).

(ii) Analysis of dolichol-bound glycans. Evidence for the involvement of dolichol phosphate-linked oligosaccharides in archaeal protein N-glycosylation also comes from examination of the carrier-bound glycan moieties. The transfer of radio-labeled glucose from UDP-[14C]glucose to *Halofexa volcanii* glycopolysaccharides proceeds through a glucose-containing phosphopolyisoprenol intermediate (489). The dolichol-linked sulfated polysaccharide moiety found in *Halobacterium salinarum* is identical to glycan moieties found on the S-layer glycoprotein and flagellin in this species (248, 470). On the other hand, the sulfated polysaccharide is methylated at the dolichol-linked stage, whereas no 3-O-methylglucose is detected in the protein-linked polysaccharide (249).

The importance of this transient methylation is illustrated by the detrimental effect of inhibiting S-adenosylmethionine-dependent methylation. Such treatment interfered with glycoprotein biosynthesis but did not affect either general protein biogenesis or the biosynthesis of sulfated phosphodolichol-bound oligosaccharides. It thus appears that methylation is an essential step in the biosynthesis of the sulfated oligosaccharide moiety prior to being transferred to its nascent polypeptide target. By contrast, the hexasaccharide moiety attached to the *Methanothermus fervidus* S-layer glycoprotein retains its methylation (204). It is not clear whether such methylation is involved in the translocation of the sulfated oligosaccharide phosphodolichol across the membrane or the subsequent transfer of the glycan moiety to the nascent polypeptide chain. In *Eucarya*, chemical modification of glycoprotein glycan moieties occurs only after the oligosaccharide has been transferred to the nascent polypeptide (449).

Enzymes of N-glycosylation. Just as archaeal N-glycosylation relies on the dolichol carriers implicated in eucaryal protein glycosylation, *Archaea* also contain homologues of many of the enzymes involved in eucaryal N-glycosylation. These include those involved in oligosaccharide charging of the lipid carrier, translocation of the dolichol carrier across the membrane, and transfer of the oligosaccharide entity to the nascent polypeptide chain (Fig. 2).

(i) Genomic studies. Analysis of the NCBI protein database (www.ncbi.nlm.nih.gov) reveals the presence of genes encoding homologues of the stauorosporine- and temperature-sensitive yeast protein 3 (Stt3p) (425), an essential protein thought to contain the active site of the multisubunit yeast oligosaccharide transferase complex (309, 493), in 18 archaeal strains. In *Bacteria*, such as *Campylobacter jejuni*, it is believed that the Stt3p homologue PglB is the only component needed for transfer of glycans to Asn residues during protein N-glycosylation (425).

A close examination of the *Archaeoglobus fulgidus* genome sequence revealed genes encoding STT3-like proteins within two gene clusters encoding putative homologues of other enzymes involved in yeast protein glycosylation (Fig. 3) (46). One of these clusters contains three adjacent open reading frames (ORFs), one of which encodes a polypeptide that appears to contain a motif present in the yeast Alg1p and Alg2p glycosyltransferase proteins. In the yeast proteins, this motif is involved in the transfer of nucleotide sugars to the phosphodolichol carrier (46). The other two ORFs putatively encode a dolichyl-phosphoglucose synthase homologue and a homologue of Stt3p. Other ORFs in this cluster show high sequence
similarity to RfbA and RfbB, components of a transporter family presumably involved in the flipping of bacterial O-antigen (467) and lipopolysaccharides (364) across the plasma membrane. While the functions of these putative gene products remain to be shown, it has been postulated that this *Archaeoglobus fulgidus* gene cluster encodes a functional unit involved in the assembly, translocation, and transfer of dolicholphosphate-linked oligosaccharides to protein targets (46).

The second gene cluster in *Archaeoglobus fulgidus* includes ORFs also encoding putative glycosyltransferase, dolichyl-phosphoglucose synthase, and STT3 proteins, and lies near six ORFs bearing similarity to genes encoding proteins involved in bacterial lipopolysaccharide biosynthesis (46).

(ii) Biochemical studies. In addition to such gene-based predictions, enzymatic activity has also been demonstrated for some archaeal glycosylation-related proteins. Biochemical characterization of *Pyrococcus furiosus* UDP-α-D-glucose pyrophosphorylase, responsible for UDP-glucose synthesis, rep-

---

**FIG. 2.** Schematic depiction of archaeal N-glycosylation. Step 1. A dolichol pyrophosphate (or monophosphate) species is glycosylated by transfer of saccharide subunits from nucleotide sugars (or possibly from lipid-bound sugar precursors). Step 2. Glycosylated phosphodolichol "flips" across the plasma membrane, likely in an enzyme-mediated process. Step 3. The oligosaccharide structure is transferred to selected Asn residues of a newly translocated polypeptide. The figure does not consider the relationship between protein translation and protein translocation or the relationship between protein translocation and protein glycosylation. Step 4. Following transfer of the oligosaccharide moiety to a protein target, the phosphorylated dolichol carrier is recycled to its original topology. See references 247, 420, and 468, the text, and Table 3 for additional information.

**FIG. 3.** Schematic depiction of two *Archaeoglobus fulgidus* gene clusters putatively involved in protein glycosylation. Putative gene products are given above each ORF. For further details, see reference 46.
resents the first analysis of an archaeal sugar nucleotidyltransferase (290). An N-acetylg glucosamine transferase was also partially characterized from membranes of *Halobacterium salinarum* (281). Dolichylphosphate mannose synthase, which is able to transfer GDP-mannose to a dolichol phosphate carrier, was purified from *Thermoplasma acidophilum* (490). Amphotericin, an inhibitor of dolicholphosphate mannose synthases (202), blocked the activity of the enzyme (490). Using 5-azido-[^32P]UDP-glucose in a photoaffinity approach, a single 45-kDa species was identified in *Halofexx volcanii* homogenates that is thought to correspond to dolichylphosphate synthase (489).

Pyrophosphatases with their active site oriented towards the cell exterior have been purified from the membranes of two different *Sulfolobus acidocaldarius* strains (8, 286). The pyrophosphate-hydrolyzing activity of the enzymes, proposed to participate in the hydrolysis of dolicholpyrophosphate-linked oligosaccharides during protein glycosylation, was stimulated in the presence of *Sulfolobus* membrane lipids. Sequence analysis of one of these pyrophosphatases has led to the identification of putative homologues in the genome sequences of *Sulfolobus tokodaii* and *Sulfolobus solfataricus* as well as in *Methanobacterium thermotrophicum* (294). This study also revealed the presence of a strongly conserved phosphatase tripartite sequence motif, Lys–XXXXX-Arg-Pro-X12-54-Pro–Ser–Gly–His–X31,44-Ser–Arg–XXXXX-His–XXX-Asp, also detected in Lpp1p and Dpp1p, *Saccharomyces cerevisiae* proteins showing hydrolytic activity towards dolichylphosphate, dolichylpyrophosphate, and other isoprenoid phosphates/pyrophosphates (116).

**Subcellular localization of glycosylation.** Several lines of evidence suggest that archaeal glycosylation occurs at the outer cell surface, the topological equivalent of the luminal-facing leaflet of the endoplasmic reticulum membrane bilayer, the site of N-glycosylation in *Eucarya* (46, 157, 235, 333, 409, 442). Despite its inability to cross the plasma membrane of haloarchaea (284), bacitracin is nonetheless able to interfere with *Halobacterium salinarum* protein glycosylation by preventing transfer of sulfated oligosaccharides to the S-layer glycoprotein (284, 469). The external orientation of the archaeal glycosylation apparatus is further supported by the decoration of exogenously added, soluble cell-impermeable hexapeptides containing the Asn-based N-glycosylation motif with sulfated oligosaccharides by living *Halobacterium salinarum* cells (248). Other observations also favor the assignment of archaeological protein glycosylation to the cell’s outer surface. These include the enol-enzyme nature of a *Sulfolobus acidocaldarius* pyrophosphatase (8, 286), the proposed specific inhibition of an externally oriented Mg²⁺-dependent oligosaccharidetranferase by EDTA, a non-cell-permeant reagent, and subsequent interference with *Halobacterium salinarum* flagellin glycosylation (420), as well as studies supporting the cotranslational mode of membrane protein insertion in *Archaetia* (360).

**Role of Protein Glycosylation in Archaea**

**Structural roles.** Given the seemingly routine glycosylation of archaeal proteins, one can ask what role is played by this posttranslational modification in *Archaetia*. The observation that bacitracin treatment transformed rod-shaped *Halobacterium salinarum* cells into spheres led to the proposed structural function of archaeal protein glycosylation (282). In fitting with a role for the sulfated S-layer glycoprotein oligosaccharide chains in maintaining the rod shape of *Halobacterium salinarum* cells, it was noted that similarities exist in the overall structures of the S-layer glycoprotein and proteoglycans, components of the extracellular matrix of animal cells (30, 468). For example, iduronic acid, a major component of proteoglycans (296), is found in the glycans decorating the *Halobacterium salinarum* S-layer glycoprotein. Similarly, the O-glycosylation cluster situated near the membrane-spanning base of the *Haloflux volcanii* S-layer glycoprotein has also been assigned a structural support role in the formation of a periplasmic-like space (217). In *Thermoplasma acidophilum*, an organism that lacks a cell wall, it has been suggested that the glycan moieties attached to the major glycosylated membrane-bound protein species coating the cell surface act to either trap water molecules or allow the cell surface proteins to interact with each other. In either scenario, glycosylation would contribute to the rigidity of the cell surface (478).

**Functional roles.** The glycosylation of archaeal proteins has also been implicated in protein assembly and function. In archaeal flagellins, glycosylation is associated with proper flagellar assembly, since upon bacitracin-mediated interference with flagellin glycosylation, a loss of Methanococcus delicae flagella was observed microscopically (196). In a mutant *Halobacterium salinarum* strain in which underglycosylated flagellins are overproduced, increased levels of flagella were detected in the growth medium, suggesting proper flagellin glycosylation to be important for correct flagellar incorporation into the plasma membrane (470). This explanation is, however, inconsistent with the apparent nonglycosylated nature of other archaeal flagellins (184) or the glycosylation of *Methanospirillum hungatei* flagellins, which only occurs in low-phosphate media (406). Similarly, evidence against glycosylation’s playing a role in protein function comes from bacterial expression of archaeal binding proteins. Normally glycosylated in their native hosts, nonglycosylated heterologously expressed versions of these proteins were also capable of substrate binding (170, 230, 231). Nevertheless, glycosylation could play a role in stabilization against proteolysis or could affect the interaction of binding proteins with the cell membrane or envelope (4).

**Glycosylation as an environmental adaptation.** Coping with the often harsh environmental conditions encountered by *Archaetia* serves as the basis for yet another hypothesized role for archaeal protein glycosylation. In a comparison of the glycosylation profiles of S-layer glycoproteins from the moderate halophile *Haloflux volcanii* and the extreme halophile *Halobacterium salinarum*, it was noted that the latter experiences a higher degree of glycosylation than the former (280). Moreover, the glycan moieties of the extreme halophile were enriched in sulfated glucuronic acid subunits as opposed to the neutral sugars found in the moderate halophile. These properties endow the *Halobacterium salinarum* S-layer glycoprotein with a drastically increased surface charge density relative to its *Haloflux volcanii* counterpart.

The enhanced negative surface charges are thought to contribute to the stability of haloarchaeal proteins in the face of molar salt concentrations (266). Accordingly, the *Halobacterium salinarum* S-layer glycoprotein also contains 20% more...
acidic amino acid residues than does the corresponding protein in *Haloflexa volcanii* (246, 421). The enhanced negative surface charge associated with protein glycosylation and the resulting protection that this would afford in the face of acidic conditions has been suggested as the role of *Sulfolobus acidocaldarius* cytochrome *b* 558/560, glycosylation (161, 484). It has also been suggested that a significant amount of the protein surface is shielded from the ~pH 2 environment by the high degree of glycosylation (484). Finally, glycosylation has also been implicated in the stabilization of thermophilic archaeal glycoproteins (4, 258, 455).

**LIPID MODIFICATION**

Lipid modification, defined herein as the permanent or temporary covalent attachment of lipid-based groups at various positions within a polypeptide chain, is a common modification experienced by both eucaryal and bacterial proteins. An examination of known lipid modifications reveals that a wide variety of lipid moieties can be directly or indirectly linked to a protein at any of numerous attachment sites through the use of any of several linkages (414). For instance, lipid modification can involve myristoyl or palmitoyl acyl groups (358), isoprenyl polymers of various lengths (393), or aminoglycan-linked phospholipids (103). These can be added at the amino terminus, the carboxy terminus, or at internal residues via ester, thioester, thioether, or amide bonds, or through mediating elements, such as the phosphopantethene group of the acyl carrier protein (267).

Lipid modification of proteins is largely a posttranslational event (115). It serves a variety of roles, the most obvious being to enhance the membrane affinity of the modified protein. Accordingly, amino-terminal acylation leads to the localization of numerous proteins to the outer membrane of gram-negative *Bacteria* (156, 379), as exemplified by Braun’s lipoprotein in *Escherichia coli* (40). Similarly, otherwise soluble eucaryal proteins also become membrane associated upon the covalent attachment of one or more lipid moieties (102, 153, 194, 462). Lipid modification can also modulate protein-protein interactions in *Eucarya*, as shown by the effects of myristylation or prenylation upon trimeric G protein subunit affinity (124, 178, 462), and in viruses, exemplified by the involvement of myristylation of the capsid proteins of human immunodeficiency virus type 1 and picornavirus in virion particle assembly and secretion (65, 142).

Lipid modifications of eucaryal proteins has also been implicated in a variety of other cellular events. These include signal transduction (287), embryogenesis and pattern formation (271), protein trafficking through the secretory pathway (297), and evasion of the immune response by infectious parasites (369, 461). Yet another role for lipid modification is exemplified by the bacterial toxin hemolysin A, which requires fatty acid acylation on an internal Lys residue for its activation (414).

Given the ubiquitous distribution and numerous functions of lipid modifications in eucaryal and bacterial proteins, it is not surprising that lipid-modified proteins have also been identified in *Archaea*.

**Membrane Lipids of Archaea**

One of the defining traits of *Archaea* that distinguish them from *Eucarya* and *Bacteria* is the chemical composition of their membrane phospholipids (206, 208). First, unlike eucaryal and bacterial phospholipids which are built on a glycerol-3-phosphate backbone, archaeal phospholipids are based on the opposite stereoisomer, glycerol-1-phosphate. Second, archaeal phospholipids contain polyisoprenyl side chains rather than the acyl groups employed by eucaryal and bacterial phospholipids. Third, archaeal phospholipids rely on ether bonds to link the isoprenyl side chains to the glycerol-1-phosphate backbone. In *Eucarya* and *Bacteria*, ester bonds link acyl side chains to the glycerol-3-phosphate backbone. Of these three traits, the use of glycerol-1-phosphate is considered the most defining, since examples of ether-linked lipids have been observed in *Eucarya* and *Bacteria* (172, 328) and non-ester-linked phospholipid fatty acids and genes encoding components involved in the metabolism of fatty acids have been reported in *Archaea* (127, 342). Indeed, free fatty acids have been observed in the lipid phase of *Methanosphaera stadtmanae* and *Pyrococcus furiosus* (51, 191). Finally, archaeal phospholipids are generally organized into the bilayer structure that is also present in eucaryal and bacterial cells, although tetraether lipid-based monolayers can be found in thermophilic and hyperthermophilic *Archaea* (92, 226).

Whereas phospholipids and other polar lipids (phosphoglycolipids, glycolipids, and sulfolipids) account for the vast majority of archaeal membrane lipids, archaeal membranes also contain acetate-soluble nonpolar lipid species, primarily neutral squalenes and other isoprenoid-based polymers (206, 207, 334, 439, 440). In halophilic *Archaea*, in which membrane lipid composition has been best studied, pigmented carotenoids, in particular bacterioruberins, are major components of the nonpolar lipid pool (243, 438). These have been implicated in affording protection from UV-induced damage (390). In addition, many halophilic *Archaea* also contain retinal as part of bacteriorhodopsin, the purple retinal-containing protein complex that functions as a light-driven proton pump (244).

**Lipid-Modified Archaeal Proteins**

In *Archaea*, lipid-modified proteins have been reported from a wide range of species. In many cases, modification involves uncharacterized lipid entities, whereas in others, direct proof for the presence of attached lipid groups remains lacking. Table 4 summarizes the various lipid-based modifications shown or presumed to exist in *Archaea*, while Fig. 4 offers a schematic presentation of representative archaeal lipid-modified proteins.

**Lipoproteins.** In the haloalkaliphile *Natronobacterium pharaonis*, halocyanin, a small blue copper protein, was proposed to undergo amino-terminal lipid modification based on the presence of the so-called lipobox sequence motif near the start of predicted amino acid sequence (274). In *Bacteri*, the Leu-Ala-Gly-Cys lipobox sequence motif (156) lies at the end of the signal sequence, the short N-terminal extension not found in the mature, lipid-modified protein (see below). At the membrane, the bacterial lipobox motif is sequentially recognized and processed by three enzymes. The sulphydryl group of
the Cys residue is first modified with a diacylglyceride by pro-
lipoprotein diacylglycerol transferase, after which the upstream
Gly-Cys bond is cleaved by signal peptidase II. The newly
exposed N-terminal Cys residue of the protein then undergoes
additional acylation by apolipoprotein N-acyltransferase to
yield the mature, lipid-modified lipoprotein (379). Direct proof
for such modification of halocyanin has not been provided
since the amino-terminal sequence of the protein could not be
determined, possibly due to modification of the amino-termi-
nal residue. Support for lipid modification of
*Natronobacterium pharaonis*
halocyanin, however, extends beyond the presence
of the lipobox motif. Halocyanin is predicted to contain a
β-turn after the lipobox, a structural feature that is character-
istic of bacterial lipoproteins (130). Furthermore, mass spec-
troscopic analysis of halocyanin was consistent with the pres-
ence of two C20 phytanyl groups ether linked to a glyceryl
group (274).

In gram-positive bacteria, it is accepted that substrate-bind-
ing proteins, components of multisubunit ABC transporters
responsible for cellular uptake of substrates, are lipoproteins
(131, 422, 430). The same may well be true in
*Archaea*. The
trehalose/maltose-binding protein of the hyperthermophile
*Thermococcus litoralis*
contains a lipobox-like sequence motif
and requires detergent for its solubilization (170). Similar mo-
tifs have been identified in other ABC sugar transporter bind-
ing proteins identified in
*Archaea*, suggesting that amino-ter-
ninal lipid modification of binding proteins takes place in
other species (4, 228).

Lipid modification is not, however, the sole mode of mem-
brane association for archaeal sugar-binding proteins. For ex-
ample, a membrane-spanning domain is predicted to anchor
the glucose-binding protein of
*Sulfolobus solfataricus* (3). It
should be noted, however, that binding proteins in this organ-
ism differ from those in other
*Archaea* in terms of amino-
terminal sequence and subsequent posttranslational processing
(see below). In *Halobacterium salinarum*, BasB and CosB, the
first examples of binding proteins involved in chemotaxis in
*Archaea*, are also thought to be lipoproteins due to their mem-
brane localization and bearing of the lipobox sequence motif
(228). Indeed, sequence analysis of putative substrate-binding
proteins in *Halobacterium salinarum*, be they involved in nu-
trient uptake or chemotaxis, suggests that all are lipoproteins (228). Finally, in the case of Pyrococcus species peptide-binding proteins, a conserved Gly-Cys motif reminiscent of the lipobox sequence located near the carboxy terminus may also be a target for lipid modification (4).

Despite the proposed presence of lipoproteins in Archaea, no archaeal homologue of signal peptidase II, one of the enzymes involved in lipoprotein precursor maturation, has been observed. Whether this is because there is no such enzyme in Archaea or because its sequence differs beyond recognition from that of its bacterial homologues, possibly in adaptation to the ether-based phospholipids of the archaeal membrane, remains unknown.

Isoprenylated proteins. Growth of Halobacterium cutirubrum, Halobacterium salinarum, and Haloferax volcanii in the presence of radiolabeled mevalonate, a precursor of the isoprene building block used to synthesize archaeal lipids (38, 398), led to the appearance of several proteins radiolabeled through the covalent attachment of a lipid entity (233, 376). Subsequent chemical analysis of the modifying lipid moiety in Halobacterium salinarum revealed a novel diphantanylglycerol methyl unit, linked to Cys residues of the modified proteins by a thioetheric bond (376). Further analysis of isoprenoid-modified proteins in Halobacterium salinarum using other radiolabeled isoprenyl derivatives revealed that the S-layer glycoprotein is modified by a second novel group, diphantanylglycerol phosphate, which is attached through an as yet uncharacterized linkage (218). Amino acid sequencing places the modification near an O-glycosylated Thr-rich stretch found in the C-terminal region of the protein, upstream of the single transmembrane domain (218). In Haloferax volcanii, lipid modification of the S-layer glycoprotein was also shown, although the chemical composition of the attached lipid is unknown, as is the site of attachment (99, 233).

Since haloarchaeal S-layer glycoproteins include a membrane-spanning domain (246, 421, 457), it is unclear why an additional membrane anchor in the form of a lipid would be required. Nonetheless, the attachment of the lipid moiety that takes place on the external surface of Haloferax volcanii and Halobacterium salinarum cells is responsible for the posttranslational, posttranslational maturation of the S-layer glycoprotein in these strains, as detected through pulse-chase radiolabeling studies (99, 233). Furthermore, since other haloarchaeal S-layer glycoproteins also contain a sequence similar to that modified in Halobacterium salinarum (246, 421, 457), it would appear that such isoprenoid-based lipid modification of S-layer glycoproteins is a general trait of halophilic Archaea (218).

Acylated proteins. Since some Archaea contain significant amounts of fatty acids (51, 127, 191) and completed archaeal genome sequences reveal the presence of genes involved in fatty acid biosynthesis and β-oxidation (342), it should not come as a surprise that the acylation of archaeal proteins has been reported. In Halobacterium cutirubrum and Methanobacterium thermoaotrophicum, subcellular fractionation and analytical chemical techniques were employed to show the acylation of several proteins (350). Chromatographic analyses identified palmitic and stearic acids as the main modifying agents, although lower levels of modification by myristic acid and other fatty acids were also observed. These acyl groups are thought to be linked to the protein via amide or ester bonds.

GPI-anchored proteins. Glycosylphosphatidylinositol (GPI) anchors represent a carboxy-terminal posttranslational lipid-based modification used to tether eucaryal proteins to various membranes (176). The GPI anchor is added to target proteins using a preformed GPI-anchoring moiety which consists of a molecule of phosphatidylinositol linked at its myoinositol headgroup to ethanolamine phosphate through an aminoglycan bridge. This lipid is transferred to the newly exposed carboxy terminus of a nascent polypeptide. The modified protein is first synthesized as a membrane-anchored precursor that undergoes proteolytic processing upstream of its carboxy-terminal transmembrane domain. The cleaved protein is thus attached to the ethanolamine end of the preassembled GPI moiety.

Although widespread in the eucaryal domain, GPI-anchored proteins have not been observed in Bacteria (103). They have, however, been detected in Archaea. In Sulfolobus acidocaldarius, three proteins were identified that incorporate radiolabeled precursors of the GPI anchor moiety (224). One of these, a 185-kDa species, was also solubilized by the actions of a bacterial phosphatidylinositol-specific phospholipase C, characteristic of GPI-anchored proteins (175). Although the other two Sulfolobus proteins were not released by the phospholipase, this is not inconsistent with GPI anchoring as phosphatidylinositol-specific phospholipase C-resistant GPI-anchored proteins have been reported (122, 365). Similarly, a typical archaeal ether-based phospholipid bearing the identical GPI anchor moiety head group as found in Eucarya was identified in Methanosarcina barkeri (310). Incubation of this lipid species with phosphatidylinositol-specific phospholipase C led to the release of the polar head group.

In addition to these biochemical studies, a bioinformatic analysis of available archaeal genome sequences predicts the presence of GPI-anchored proteins in other archaeal species (103). Moreover, many of the 19 enzymes known to participate in the biosynthesis of GPI anchors have been detected in archaeal genome sequences (104).

PROTEIN PHOSPHORYLATION

Like other forms of posttranslational modification considered in this review, the covalent attachment of phosphate groups to protein targets at any of a number of surface Asp, His, Ser, Thr, or Tyr residues can profoundly affect protein behavior. However, in contrast to N-glycosylation and, in most cases, lipid modification, covalent modification of proteins by phosphorylation is a reversible event. This property, combined with the major perturbation in protein structure that results from phosphorylation (189), has made this versatile form of posttranslational modification widely used when rapid and profound changes in protein behavior are called for (214, 215). As such, protein phosphorylation and dephosphorylation are most commonly exploited by the cell in adaptive pathways designed to present appropriate responses to various cues associated with a multitude of external and internal stimuli (173).

Although discovered in the 1950s (240), it took approximately 25 years for the first reports of phosphorylated proteins in Bacteria to appear (126, 459). Shortly thereafter, in 1980, the presence of phosphorylated proteins in Halobacterium salinarum was reported (413), confirming that Archaea too are ca-
pable of performing this posttranslational modification. With the subsequent availability of genome sequences, it became clear that Archaea also contain numerous kinases and phosphatases, enzymes responsible for protein phosphorylation and dephosphorylation, respectively (214, 215, 253).

**Targets and Functions of Protein Phosphorylation in Archaea**

The first examples of archaeal protein phosphorylation were reported when *Halobacterium salinarum* grown in the presence of $^{32}$P-labeled orthophosphate was shown to phosphorylate Ser and Thr residues of several protein species (413). The radio-labeling of 100- and 80-kDa proteins and, as shown later, an additional 62-kDa species (411) was, however, greatly diminished upon exposure to light. Moreover, the light-dependent dephosphorylation of these proteins could be linked to the proton motive force generated by the light-driven proton pump bacteriorhodopsin. In related studies (395), it was shown that growth in $^{32}$P-labeled orthophosphate-containing growth medium led to the appearance of serine- and threonine-phosphorylated medium of proteins 71, 52, 42, and 31.5 kDa in *Sulfolobus acidocaldarius*, in a growth-phase-dependent manner. Further examination revealed the existence of an additional 40-kDa *Sulfolobus acidocaldarius* phosphoprotein that was threonine-phosphorylated in the presence of $[^3P]$polyphosphate (396). The first phosphoprotein with a known function to be identified in *Archaea*, however, was the methyltransferase activation protein from *Methanosarcina barkeri*, a key enzyme involved in the metabolic transformation of carbon dioxide to methane (81).

Although other phosphorylated proteins have been identified in *Archaea* (475), the observed phosphorylation cannot usually be attributed to a regulated protein kinase (see below), but rather reflects phosphorylated intermediates that appear during an enzyme’s catalytic cycle. Such enzymes apparently include the alpha subunit of succinyl-coenzyme A synthase in *Sulfolobus solfataricus* (403) and *Sulfolobus acidocaldarius* glycogen synthase (52, 397). Nevertheless, examples of regulated protein phosphorylation in *Archaea* have been reported (Table 5) and are discussed below.

**Phosphorylation of components involved in signal transduction.** Protein phosphorylation as part of an archaeal two-component signal transduction pathway was first shown for *Halobacterium salinarum* (373, 374). In *Bacteria* and a very limited number of *Eucarya*, two-component signal transduction response pathways are responsible for the appropriate response of the cell to a wide range of environmental conditions (234, 332, 423). The conformational changes that result upon ligand binding to the extracellular portion of a transmembrane receptor are transduced into the cell, where they lead to the modulation of sensor (histidine kinase, see below) and response regulator proteins. Such modulations ultimately activate the transcription of genes encoding compensatory proteins or affect the motion of the microorganism via motility structures. Transduction of the ligand binding event to sensor and response regulator proteins is achieved via a cascade of phosphorylation reactions. Hence, the detection of phosphorylated *Halobacterium salinarum* CheA and CheY, well-characterized sensor and response regulator proteins, respectively (114, 423), pointed to the presence of a two-component system in *Archaea*, charged with responding to various chemotactic and phototactic stimuli (373, 374).

Protein phosphorylation in response to environmental change has also been observed in other archaeal species. Growth of *Sulfolobus acidocaldarius* in the presence of radio-labeled phosphate under limited-phosphate conditions revealed the existence of numerous phosphoproteins (319). In particular, the phosphorylation of a 36-kDa protein was augmented under phosphate starvation, hinting at a regulatory role in a cellular response pathway for this protein. In *Halofexa volcanii*, growth at elevated salt concentrations may lead to the appearance of several serine-phosphorylated proteins not detected during growth under optimal salt conditions (32). A threonine-phosphorylated 67-kDa membrane protein displaying serine kinase activity has been found in *Sulfolobus solfataricus*, although the pathway in which this protein participates remains to be defined (261, 264).

**Phosphorylation of components involved in DNA replication, cell cycle regulation, and translation.** In addition to playing a role in signal transduction, protein phosphorylation has also been implicated in eucaryal DNA replication, cell cycle regulation, and protein translation (313, 314, 349). Similar roles for protein phosphorylation have also been observed in *Archaea*. In *Methanobacterium thermoautotrophicum*, *Pyrococcus aerophilum*, and *Sulfolobus solfataricus* (89, 144), DNA-dependent serine autophosphorylation has been reported for the Cdc6 protein, an initiator protein that fulfills an essential role in DNA replication and is known to be phos-
phosphorylated in Eucarya (182, 212). The autophosphorylation of Cdc6 proteins reveals similarities between the archaean and eucaryal replication processes, even though domain-specific differences in Cdc6 autophosphorylation have been noted (144). Protein phosphorylation also takes place during both eucaryal and archaean protein translation. In vitro studies addressing the heterotrimeric archaean initiation factor 2 complex (aIF2) from Pyrococcus horikoshii showed that the aIF2 α subunit could be phosphorylated (426), as is the case for the parallel eucaryal eIF2 α subunit (93, 251).

**Phosphorylation of other proteins.** In other instances, archaean phosphoproteins have been identified in which the role of this posttranslational modification remains obscure. In Sulfolobus solfataricus, for example, a novel zinc-dependent aminopeptidase, originally isolated from cell lysates in complex with a chaperonin, was shown to be phosphorylated (73).

Finally, whereas the bulk of phosphorylated archaean proteins experience modification of Asp, His, Ser, or Thr residues, it is now known that archaean proteins can also undergo phosphorylation at Tyr residues. Using antiphosphotyrosine antibodies, tyrosine-phosphorylated proteins were first identified in cell extracts of Haloferax volcanii, Methanosarcina thermophila, and Sulfolobus solfataricus (401). In Thermococcus kodakarensis KOD1, tyrosine-phosphorylated proteins recognized by antiphosphotyrosine antibodies were subsequently identified by N-terminal sequencing as RtcB, which is involved in RNA processing (128), the phenylalanyl-tRNA synthetase identified by N-terminal sequencing as KOD1, tyrosine-phosphorylated proteins recognized by antiphosphotyrosine antibodies.

**Archaeal Protein Kinases and Phosphatases**

In general, phosphorylated proteins do not contain readily recognizable sequence regions that allow their assignment as candidates for this posttranslational modification. In contrast, protein kinases and phosphatases, the enzymes responsible for the addition and removal, respectively, of orthophosphate groups from target proteins, contain conserved sequence motifs (213). Based on such motifs, protein kinases and phosphatases can be divided into several functional families (213). Thus, the availability of several archaean genome sequences has allowed a catalogue of the potential protein kinases and phosphatases to be assembled (214, 215). A better understanding of the archaean proteins should also provide insight into the relationship between eucaryal and bacterial kinases and phosphatases, which were once thought to be distinct (234, 253). For a more detailed examination of archaean kinases and phosphatases, the reader is directed to a recent review of the subject (215).

**Eucaryal protein kinases.** Members of the eucaryal protein kinase superfamily, an evolutionarily conserved group of proteins sharing a common core, serve as the major providers of protein serine/threonine/tyrosine kinase activity in Eucarya (154). Long considered to be restricted to Eucarya, homologues of eucaryal protein kinases were subsequently reported in Bacteria and more recently detected in Archaea (214). Initially, searches of the then-available archaean gene sequences identified ORFs in Methanococcus thermolithothrophicus, Methanococcus vannielii, and Methanococcus voltae encoding proteins whose carboxy-terminal regions contain 9 of 11 subdomains associated with eucaryal protein kinases (400). In a later study (215), analysis of nine completed archaean genomes revealed the presence of ORFs encoding polypeptides containing sequence motifs essential for eucaryal protein kinase activity in seven.

Gene-based studies of individual strains have also revealed the existence of eucaryal protein kinases in other Archaea, such as in Haloferax volcanii cells exposed to elevated salt levels, in which a salt-regulated gene putatively encoding a protein similar to the eucaryal serine/threonine kinase was detected (32). Subsequent studies employing complete archaean genome sequences, moreover, have expanded our knowledge of eucaryal protein kinases. In a comprehensive search based on a large number of completed genome sequences, including those of four Archaea, archaeal representatives of four novel putative protein kinase families were reported (253), such as the Riol family, comprising only archaean and eucaryal members, or the ABC1 family, including only a single archaean representative (from Methanobacterium thermoautotrophicum). Furthermore, the recent solution of the crystal structure of Archaeoglobus fulgidus Rio2 suggests that this protein defines a new family of protein kinases (245).

In addition to sequence-based analyses, archaean homologues of eucaryal protein kinases have been examined at the protein level. Analysis of threonine-modified phosphoproteins in Sulfolobus solfataricus membranes following incubation with ^32P-ATP led to the identification of the protein encoded by ORF sso0469 (264). Sequence analysis revealed the presence of eukaryotic protein kinase motifs, while biochemical characterization of a recombinant version of the encoded protein revealed its ability to phosphorylate Ser residues of exogenous polypeptides in vitro. Similarly, SsoPK2, the product of Sulfolobus solfataricus ORF sso2387, also contains sequence motifs found in eucaryal protein kinases (263). Moreover, a recombinant form of the protein was able to phosphorylate itself as well as various exogenous targets, relying on that part of the protein homologous to eucaryal protein kinases, as revealed by mutagenesis approaches (263).

**Histidine kinases.** Histidine kinases are elements of the two-component signal transduction pathway described above. In response to conformational changes experienced by upstream receptor-transducer teams, histidine kinase sensors use ATP to autophosphorylate His residues before transferring the phosphoryl group to Asp residues of downstream response regulators. The first example of an archaean histidine kinase as part of a two-component system identified was Halobacterium salinarum CheA (373). A recombinant version of the haloarchaeal CheA histidine kinase was autophosphorylated upon addition of radiolabeled ATP and was subsequently able to transfer its phosphoryl group to an Asp residue of the Halobacterium salinarum CheY response regulator (374).

In later homology-based searches of nine completed archaean genome sequences, histidine kinases were identified in four: Archaeoglobus fulgidus, Halobacterium sp. strain NRC-1, Methanobacterium thermoautotrophicum, and Pyrococcus horikoshii (215, 220, 234). Of these, Methanobacterium thermoautotrophicum and Archaeoglobus fulgidus contain the most histidine kinases (16 and 14, respectively) and response regulators...
(10 and 11, respectively). At the other extreme, *Pyrococcus horikoshii* contains only a single histidine kinase and two response regulators (corresponding to CheA and to CheY and CheB, respectively), while *Aeropyrum pernix*, *Methanococcus jannaschii*, and *Thermoplasma acidophilum* are not predicted to encode such proteins. The absence of Che proteins in *Methanococcus jannaschii* is noteworthy, given that this species is both flagellated and motile (436).

**Protein serine/threonine phosphatases.** Protein serine/threonine phosphatases can be structurally and functionally grouped into the protein serine/threonine phosphatase (PPP) family and the Mg\(^{2+}\) and Mn\(^{2+}\) protein phosphatase (PPM) families (21). PPP family members are mainly responsible for serine/threonine dephosphorylation in *Eucarya* and have also been reported in *Bacteria* (71, 213, 214). In contrast, members of the PPM family are the primary mediators of dephosphorylation in *Bacteria*, although this family encompasses several eucaryal protein phosphatase classes as well (37, 214). In *Archaea*, members of both protein serine/threonine phosphatase families have been identified in completed genome sequences and some have been studied at the protein level (213–215).

To date, three PPM family protein serine/threonine phosphatases have been characterized from *Archaea*. The genes encoding PP1-arch1, PP1-arch2, and Py-PP1 were cloned from *Sulfolobus solfataricus* (216, 252), *Methanosarcina thermophila* TM-1 (321, 403), and *Pyrodiculum abyssi* TAG11 (268), respectively. In addition, other archaeal PPP family phosphorylases have been predicted following analysis of genome sequences, relying on the presence of conserved sequence motifs (24, 215). Such sequence comparisons revealed the archaeal enzymes to be more closely related to their eucalyar than their bacterial homologues (24). However, despite their sequence similarities to eucalyar PPP family members, archaeal PPP family protein serine/threonine phosphatases display a combination of eucalyar and bacterial features (215). Like their eucalyar counterparts, the archaeal enzymes specifically act upon protein-bound phosphoserine and phosphothreonine residues and, in the cases of PP1-arch2 and Py-PP1, are inhibited by toxic secondary metabolites such as okadaic acid (268, 321, 403). In contrast, the three archaeal PPP family members require the addition of metal ions such as Mn\(^{2+}\) for activity, as is the case for bacterial PPP family protein serine/threonine phosphatases (391). Finally, protein serine phosphatase activity has also been detected in extracts of *Halobacterium salinarum* (36) and *Haloferax volcanii* (320), but the enzymes responsible have not been identified.

A single ORF encoding a potential PPM family protein serine/threonine phosphatase was identified in the genome sequence of *Thermoplasma volcanium*. The putative protein includes all of the conserved sequence elements of PPM family members (209).

**Protein tyrosine phosphatases.** While ORFs thought to encode protein tyrosine phosphatases have been detected in *Archaeoglobus fulgidus*, *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Pyrococcus abyssi*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Sulfolobus solfataricus*, and *Thermococcus kodakaraensis* KOD1 (215, 418), only the *Thermococcus kodakaraensis* KOD1 enzyme has been examined biochemically (188). A recombinant version acted on both free phosphotyrosine and phosphoserine, suggesting that it had dual specificity. Moreover, a mutant form of the enzyme was used to capture putative native substrates from a cell extract (188). In addition, studies performed with *Halobacterium salinarum* extracts detected protein serine/threonine phosphatase activity also able to hydrolyze phosphotyrosine, suggesting the responsible enzyme similarly had dual specificity (36).

**Protein kinases and phosphatases of *Thermoplasma acidophilum*.** It should be noted that analysis of the genome of *Thermoplasma acidophilum*, using tools available today, has failed to detect the presence of any protein kinase or phosphatase (214). While it remains to be seen whether the current inability to recognize such proteins will be remedied in future with the development of more powerful bioinformatic prediction tools, it is also possible that *Thermoplasma acidophilum* contains novel archaee-specific kinases or phosphatases, or does not perform protein phosphorylation. Interestingly, genome analysis of two bacterial strains, *Buchnera* sp. and *Ricksettia prowazekii*, also failed to detect ORFs encoding putative protein kinases or phosphatases (214), although the implications of these studies are at present unknown.

## PROTEIN METHYLATION

Although methylation of nucleic acids is well known, in part due to a role in disease states such as cancer (86, 256, 424), a wide variety of proteins have also been reported to experience posttranslational modification. This modification affects the amino group in the side chains of Ala, Arg, Glu, His, Lys, and Pro residues, the hydroxyl group in the side chains of Glu and Asp, and the thiol group of Cys residues (327). Enzyme-catalyzed addition of methyl groups from S-adenosylmethionine can either occur reversibly, as in O-methylation of carboxyl groups, or irreversibly, as in the N-methylation of amino-terminal or side chain nitrogen atoms (70).

As is the case with other posttranslational modification events considered in this review, analysis of protein methylation in *Archaea* has revealed novel forms of protein methylation as well as providing new insights into the biological role served by this posttranslational modification.

**Protein Methylation in Response to External Stimuli**

As described above, various external stimuli that modulate the motility of archaeeal cells rely on phosphorylation of elements of the two-component signal transduction response pathway. Phosphorylation is not, however, the sole posttranslational modification experienced by proteins involved in taxis responses to environmental cues. As in *Bacteria* (90, 229, 236, 423), numerous proteins involved in the archaeeal response to growth conditions also undergo methylation (see below). Methylation of taxis receptor or transducer proteins is thought to be responsible for adaptation, a form of cellular memory necessary for cells to be able to sense and move towards ever higher attractant concentrations or to recognize when motion is occurring in the wrong direction, i.e., away from elevated attractant concentrations (423).

Three methylation-dependent taxis responses, phototaxis, chemotaxis, and aerotaxis, have been detected in *Halobacte-
In Halobacterium salinarum, the phototactic response is initiated by the excitation of the two retinal-containing photoreceptors, sensory rhodopsin I and sensory rhodopsin II (121, 239, 341, 412, 482, 488). These subsequently relay the excitatory signal to their respective transducer proteins, HtrI and HtrIII. During phototaxis, these proteins undergo methylation, a posttranslational modification previously shown to modulate the life span of phototactic signals in Halobacterium salinarum, i.e., to play an adaptive role (164). Methylation of HtrII is also involved in the transducer role assumed by the protein during serine chemotaxis (171). The cytoplasmic transducer HtrXI undergoes methylation/demethylation in response to changes in extracellular histidine, aspartate and glutamate concentrations (43).

Arginine taxis in Halobacterium salinarum involves the methylatable soluble transducer Car, which monitors intracellular levels of the amino acid (417), while the methylation status of the membrane-bound transducer BasT affects chemotactic behavior towards leucine, isoleucine, valine, methionine, and cysteine (227). HtpIV, or CosT, the transducer for the haloarchaeal chemotaxis response towards trimethylamine, also experiences methylation (228). The aerotactic (oxygen gradient-sensing) response of Halobacterium salinarum was also shown to rely on methylation, in this case of the membrane-bound transducer HtrVIII (259). In contrast, aerotaxis in Bacteria such as Escherichia coli and Salmonella enterica serovar Typhimurium does not require transducer methylation (259). Most recently, MpcT, the transducer of membrane potential changes in Halobacterium salinarum (formerly known as HtrXIV) was shown to experience differential degrees of methylation (225).

**Methylation of Methyl-Coenzyme M Reductase**

In methanarchaeae, the final reaction in the release of methane is catalyzed by the enzyme methyl-coenzyme M reductase (434). Analysis of the crystal structure of the enzyme from Methanobacterium thermoautotrophicum revealed the presence of five modified amino acid residues in the α subunit of the hexameric enzyme, all situated near the active-site region (108). In addition to a thioglycine residue, the enzyme contains 1-N-methylhistidine, 5-(S)-methylarginine, 2-(S)-methylglutamine, and an S-methylcysteine residue (Fig. 5). Whereas 1-N-methylhistidine and S-methylcysteine have been detected in other proteins (70, 326) and a thiol-modified glycine residue has been identified in ThiS, one of the enzymes involved in thiamine biosynthesis in Escherichia coli (432), Methanobacterium thermoautotrophicum methyl-coenzyme M reductase is the first example of a 2-(S)-methylglutamine and 5-(S)-methylarginine. Previously, only N-methylglutamine and N-methylarginine had been reported (162, 492).

The posttranslational modifications leading to the appearance of the four methylated amino acids in methyl-coenzyme M reductase involve the transfer of the methyl group of methionine, most likely in the S-adenosylmethylionine form (387). The modifications are thought to occur before methyl-coenzyme M reductase assumes its quaternary structure, since the modified residues are buried deep inside the native enzyme, where they would be inaccessible to S-adenosylmethionine or methyltransferases, which catalyze protein methylation (108, 143). Furthermore, considering the differences in amino acid composition in the vicinities of the four methylated residues (387), it is probable that four different S-adenosylmethionine-dependent methyltransferases are involved in the modification reactions (308). Accordingly, multiple methyltransferases appear to be present in the genome sequence of Methanobacterium thermoautotrophicum (399).

In terms of function, methylation of His-257, which is involved in substrate binding, likely affects the substrate affinity of the enzyme (108). The thiglycine residue has been proposed to serve as a one-electron relay in the catalytic mechanism (434). The functional significance of the methylation of the other three modified residues, i.e., 5-(S)-methylarginine-271, 2-(S)-methylglutamine-400, and S-methylcysteine-452, remains unknown. However, analysis of methyl-coenzyme M reductase sequences in a wide range of methanarchaeal species reveals the absolute conservation of the five amino acid residues modified in the Methanobacterium thermoautotrophicum enzyme (311, 410). Moreover, the crystal structure of methylcoenzyme M reductase from Methanosarcina barkeri also revealed the presence of thioglycine, S-methylcysteine, 1-N-methylhistidine, and S-methylarginine residues, i.e., four of the five posttranslational modifications found in the Methanobacterium thermoautotrophicum enzyme, suggesting that such modifications are important for catalysis (143).

**Methylated Proteins in Thermophilic Archaea**

Methylated Lys residues have been detected in several thermophilic archael proteins, such as Sulfolobus acidocaldarius.
ferredoxin (289) and *Sulfolobus solfataricus* glutamate dehydrogenase (272), aspartate aminotransferase (485), and β-glycerophosphate dehydrogenase (117). In the case of *Sulfolobus solfataricus* β-glycerophosphate dehydrogenase, N-terminal methylation of specific Lys residues was associated with increased thermal stability as well as with a lower susceptibility to denaturation and aggregation, in comparison to the nonmethylated recombinant version of the enzyme produced in *Escherichia coli* (117). The methylated Lys residues found in the *Sulfolobus solfataricus* enzyme are not conserved in other mesophilic glycosidases belonging to glycosyl hydrolase family 1, again pointing to a thermostabilizing role for this posttranslational modification.

**Methylation of Archaeal DNA-Binding Proteins**

Although grouped with *Bacteria* as prokaryotes (472), *Archaea* resemble *Eucarya* in many aspects, including that members of both domains contain histones, proteins involved in DNA packaging (355, 466). First demonstrated in *Methanothermus fervidus* (378, 415), over 30 archaeal histone sequences have since been identified (355). Archaeal histones are, however, apparently restricted to *Euryarchaea*, an archaeal subdomain, in which several different histone-encoding genes have been detected (355, 456). No archaeal histones have been observed in *Crenarchaeota*, the other major archaeal subdomain. Instead, crenarchaeal species contain small, basic DNA-binding proteins thought to fulfill the same functions as histones, based on their physical properties (63, 64, 149, 355, 367, 466). In *Sulfolobus*, these can be grouped into 7-, 8-, and 10-kDa classes, with the 7-kDa proteins, referred to as the Sul7 family (466), predominating. Members of the Sul7 family in both *Sulfolobus acidocaldarius* and *Sulfolobus solfataricus* are modified by monomethylation of selected Lys residues to different extents in a strain-dependent manner (26, 63, 64, 97, 278, 317).

Given the modulation of eucaryal histone function that results from methylation (59, 238), it is likely that methylation of archaeal Sul7 proteins also affects their behavior. Indeed, the observation that methylation of Sul7 proteins increased during heat shock suggests that such posttranslational modification is of functional, although as yet undefined, significance (26). Sul7 methylation does not, however, affect DNA binding affinity, consistent with the positioning of methylated Lys residues on the surface of the Sul7d-DNA crystal rather than at the protein-DNA interface (25). Finally, it is somewhat ironic that while eucaryal Sul7 proteins are methylated, no evidence for methylation of archaeal histones has appeared, in contrast to their eucaryal counterparts (355). This is due to the fact that archaeal histones lack the amino- and carboxy-terminal extensions that undergo this posttranslational modification in eucaryal histones (59). Indeed, analysis of archaeal genome sequences reveals homologues of only one of the components involved in the eucaryal histone modification event, i.e., the histone acetyltransferase Elp3 (355).

**Methylation of Archaeal Ribosomal Proteins**

Several bacterial ribosomal proteins, mainly found in the large 50S subunit, undergo methylation (55, 56). Of these, L11 is the major methylated ribosomal component. Analysis of *Halobacterium cutirubrum* and *Sulfolobus solfataricus* L11 proteins from cells grown in the presence of radiolabeled methionine and/or methylmethionine revealed that they are also methylated, albeit in a pattern distinct from that of the bacterial protein (9, 353, 354). Accordingly, genome searches have failed to identify an archaeal homologue of the bacterial L11 methyltransferase PrmA (45). The role of L11 methylation in both *Bacteria* and *Archaea* remains unknown.

**Disulfide Bonds in Proteins**

In both *Eucarya* and *Bacteria*, secreted and extracellularly oriented membrane proteins are often stabilized by disulfide bonds, i.e., covalent links between the sulfhydryl groups of Cys residues in the same or different polypeptide chains. These can serve two roles. First, they can stabilize proteins by entropic destabilization of the unfolded conformation (78, 322, 463, 465). Second, they serve to limit damage to a protein resulting from oxidative or proteolytic agents, thereby enhancing protein lifetime. Accordingly, disulfide bonds are routinely employed by secretory and plasma membrane proteins in numerous organisms (315, 451).

The various compartments of the cell greatly differ in terms of redox potential and hence in their ability to catalyze disulfide bond formation. Accordingly, disulfide bond formation takes place in the endoplasmic reticulum of eucaryal cells (444) and in the periplasmic/extracellular compartment of bacterial cells (193, 351). In both locations, oxidative conditions favor disulfide bond formation and enzymes implicated in this posttranslational modification are found. Conversely, it had been generally accepted that proteins found in the reducing environment of the cytosol do not contain disulfide bonds, although it has recently become clear that a number of cytosolic proteins can contain specific and reversible disulfide bonds (see below). In such cases, the cyclic oxidation/reduction of a disulfide bond can control the activation/deactivation or otherwise modulate the activity of a protein (62, 80, 181, 325, 363). Indeed, controlled reduction of disulfide bonds has also been adopted by certain disulfide-containing secreted proteins and cell surface receptors (166). Nevertheless, the number of cytosolic proteins in *Eucarya* and *Bacteria* experimentally shown to contain disulfide bonds is limited. *Archaea*, however, do not follow this trend (269).

**Disulfide Bonds in Cytoplasmic Archaeal Proteins**

Unexpectedly, biochemical and structural characterization of many cytoplasmic archaeal proteins has revealed the presence of disulfide bonds. A disulfide bond was detected in *Pyrococcus furiosus* ferredoxin, in which it plays a role in the redox cycle of the protein (141). The crystal structures of DNA polymerases from *Thermococcus gordonii* and *Thermococcus* sp. strain 9’N-7 revealed the presence of two disulfide bridges in each case (169, 368). The recombinant form of *Aeropyrum pernix* alcohol dehydrogenase was shown to contain a disulfide bond (152) as was *Sulfolobus solfataricus* glyceraldehyde-3-phosphate dehydrogenase (180). The three-dimensional structure of the TATA box-binding protein from the hyperthermophile *Pyrococcus woesei* revealed the presence of a disulfide bond not found in mesophilic versions of the protein (88). Indeed, in many of these examples, the presence of di-
sulfide bonds is believed to contribute to the enhanced thermostability of the modified protein.

Disulfide bonds are also used by cytosolic archaeal proteins for the generation of higher-order structures. As revealed by X-ray crystallography and site-directed mutagenesis, a single intersubunit disulfide bridge is responsible for the dimeric nature of Sulfolobus solfataricus glycansyltransferase trehalohydrolase (118) and pyrrolidone carboxylase from Thermococcus litoralis (394) and Pyrococcus furiosus (316). Similarly, ferric reductase from Archaeoglobus fulgidus, with a single disulfide bond serving to link the two subunits of the protein (60). In Pyrococcus horikoshii, oligomerization of isopropylmalate isomerase relies on intersubunit disulfide bridges (483). The homotetrameric structure of the subunit of dissimilatory sulfite reductase is also due to disulfide bonding (370). The nuclear magnetic resonance structure of Pyrococcus abyssi tRNA (m1A) methyltransferase is also due to disulfide bond formation (370). The nuclear magnetic resonance structure of Pyrobaculum aerophilum DsrC, the archaeal homologue of the γ subunit of dissimilatory sulfate reductase, responsible for the reduction of sulfate in sulfate-reducing bacteria, was also shown to contain two disulfide bonds (76).

Disulfide bond formation is also responsible for the hexameric states of t-isopropyl-3-O-methyltransferase from Sulfolobus tokodaii (431) and of S'-deoxyS'-methylthioadenosine phosphorlyase from Sulfolobus solfataricus (11, 47).

Despite the seemingly widespread presence of disulfide bonds in cytoplasmic archaeal proteins, it is only in the detection of three disulfide bonds in the crystal structure of Pyrobaculum aerophilum adenyllyosuccinate lyase (441) that the concept of the general use of disulfide bonds in cytoplasmic proteins in this and possibly other hyperthermophiles Archaea was proposed (269). Accordingly, computational analysis of completed archaeal genomic sequences, involving sequence-structure mapping approaches with subsequent analysis of the proximity of pairs of Cys residues, indicated that disulfide bonds are indeed prevalent in thermophilic and hyperthermophilic Archaea. For instance, Thermotoga maritima, which includes thioredoxins, glutaredoxins, disulfide bond formation (Dsb) proteins, and protein disulfide isomerasers (PDI), have active sites containing the Cys-X-X-Cys sequence motif and the thioredoxin fold structural motif (273). DsbA is found in the bacterial periplasmic space and is involved in protein disulfide bond formation (193, 351), while PDI catalyzes protein disulfide bond formation, reduction, and rearrangement in the eucaryal endoplasmic reticulum (444, 471). Acting as strong reductants in various cellular processes (120, 348), both the thioredoxin system, involving two thioredoxins and thioredoxin reductase, and the glutaredoxin system, including three glutaredoxins and glutathione reductase, maintain intracellular disulfide bonds in the reduced state through NADPH-dependent pathways (168, 363).

To date, few archaeal protein disulfide oxidoreductases have been described (see below) and, considering the limited information available, it is too early to assign any of them a physiological role. What is known, however, points to the unique character of the archaeal proteins. For instance, Methanobacterium thermoautotrophicum contains a small protein (Mt08077) with a thioredoxin/glutaredoxin-like fold that exhibits sequence similarity to glutaredoxins, including the characteristic Cys-Pro-Tyr-Cys active-site motif (279). While its function was initially tentative, subsequent structural analysis and sensitive enzyme assays (10) revealed it to be a true thioredoxin. Nuclear magnetic resonance-based structural studies of another Methanobacterium thermoautotrophicum protein (Mt08959) revealed that it too contains a thioredoxin/glutaredoxin-like fold. This protein was originally annotated as a conserved hypothet-
Archaeal Signal Sequences

In any cell, a subset of proteins must cross one or more membranes to realize their ultimate localization and fulfill their designated roles. Across evolution, such proteins are generally synthesized with a cleavable amino-terminal extension referred to as the signal sequence that is enzymatically removed once such proteins have traversed the membrane. Analysis of signal sequence composition in Archaea as well as their posttranslational removal reveals a mosaic of archaeal, eukaryal, and bacterial traits.

Protein translocation in Archaea. Translocation of extracytoplasmic proteins begins with their delivery to translocation sites in the membrane (42, 119, 298). Examples of both post- and cotranslational translocation have been found in Archaea. Chimeric signal sequence-bearing reporter proteins are secreted posttranslationally from transformed Haloferax volcanii cells (179). In addition, Haloferax volcanii has been reported to posttranslationally insert a chimeric protein containing the multi-spanning membrane protein bacterio-opsin (318). In contrast, cotranslational translocation, shown to be the general mode of membrane protein insertion in Haloferax volcanii (360), likely involves the archaeal signal recognition particle pathway (293, 494), as first reported for Halobacterium salinarum bacterio-opsin (83, 84, 148).

In Archaea, as across evolution, the Sec translocon is the major site for protein export (94). The SecY, SecE, and Sec61β proteins that form the core of the translocation apparatus are closer to their eukaryal than their bacterial homologues (49, 155, 223, 347, 357, 361). The recent solution of the three-dimensional structure of the Methanococcus jannaschii SecYEβ translocon has provided major insight into the translocation event across evolution, including the mode of translocon gating and mechanism of membrane protein insertion (446). The Sec translocon may also be involved in the translocation of archaeal flagellins, despite their distinct signal sequence composition (see below) (184, 436). In contrast to their bacterial counterparts, which cross the plasma membrane through the hollow core at the center of the growing flagellum (7, 265), archaeal flagellins are likely translocated across the membrane and only then added to the base of the growing motility structure, as gauged by the presence of unique cleavable N-terminal signal sequences in the archaeal proteins (20, 184, 436).

Archaea also use the twin-arginine targeting (Tat) pathway, a second protein export pathway (446). The Tat pathway can be distinguished from the Sec pathway by the unique composition of substrate signal sequences (see below) and by the ability of the Tat pathway to translocate folded or cofactor-incorporating proteins (29, 366). Although the Tat pathway is proposed to predominate in halophilic Archaea (35, 371), little is presently known of the workings of the Tat system in these or other organisms.

Genomic surveys of archaeal signal sequences. Descriptions of archaeal signal sequences have largely relied on analysis of genome sequences, using computer-based tools originally designed to detect eucaryal or bacterial signal sequences (2, 16, 35, 94, 306, 371, 377). At best, these algorithms should be able to identify only those archaeal signal sequences bearing sufficient similarity to their eucaryal and bacterial counterparts. Archaeal signal sequences possessing domain-specific traits would, therefore, likely be overlooked in such searches. Thus, true characterization of archaeal signal sequences will have to wait for the number of experimentally verified targets to be extended well beyond the few experimentally verified sequences presently available. Nonetheless, such efforts have identified signal sequences recognized by the Sec and Tat pathways, archaeal flagellin-like signal sequences on both flagellin and nonflagellin proteins, as well as lipoprotein signal se-
quences (Fig. 6).

While the signal sequences of Sec pathway substrates can differ widely, they share common structural traits, such as a positively charged amino-terminal region leading to a hydrophobic core region that continues into an uncharged polar region terminating in the signal peptidase cleavage site (454). From examination of 10 genome sequences, it was concluded that predicted archaeal Sec signal sequences are more similar to their bacterial than their eucaryal counterparts (16, 307). The findings of this multigenome study (16) are in agreement with earlier studies addressing predicted signal sequences in *Methanococcus jannaschii* (306) and *Solfolobus solfataricus* (2), although differences exist. Nonetheless, as discussed below, apparent similarities in the mechanism of archaeal and eucaryal signal peptidase action (18, 100, 437) suggest similarities between the cleavage site regions of signal sequences in these two domains.

While sharing the same tripartite organization, Tat pathway signal sequences differ from those recognized by the Sec pathway in that the former include an extended amino-terminal region containing a highly conserved motif based on two Arg residues and a less hydrophobic core region (29, 366). Analysis of genome sequences predicts limited-use presence of Tat signal sequences in *Archaea* (2, 16, 94), with the apparent exception of halophilic *Archaea* (35, 371). Here, proteins bearing Tat signal sequences are predicted to greatly outnumber those synthesized with Sec signal sequences. As first noted in *Methanococcus voltae* (113), archaeal flagellins are made as precursors bearing atypical short, positively charged signal sequences, reminiscent of signal sequences found in bacterial type IV prepilins (20, 184, 436). Unexpectedly, genome analysis predicted the presence of the same signal sequence in a set of 10 extracellular *Sulfolobus solfataricus* proteins, including six putative solute-binding proteins (2). Archaeal flagellin signal sequences have also been predicted to exist in other types of protein, including those assigned solute-binding roles, in *Methanococcus jannaschii*, *Pyrococcus horikoshii*, *Sulfolobus shibatae*, and *Thermococcus litoralis* (5). In contrast to Sec and Tat pathway signal sequences, cleavage of archaeal flagellin signal sequences by the appropriate signal peptidase (see below) occurs upstream, rather than downstream, of the hydrophobic core region.

As discussed above, sequence analysis studies have also predicted the existence of proteins synthesized with lipoprotein signal sequences in *Archaea* (4, 170, 228, 274), although experimental support for these predictions has yet to be presented.

**Removal of archaeal signal sequences.** The signal sequences of proteins translocated by either the Sec or Tat pathway are removed by the actions of type I signal peptidases (82, 324). In *Archaea*, type I signal peptidases incorporate traits of both their eucaryal and bacterial counterparts. As in *Eucarya*, the archaeal signal peptidase does not rely on the catalytic Ser-Lys dyad employed by the bacterial enzyme and has replaced the conserved bacterial Lys with a His residue (100, 324, 437). At
present, the catalytic mechanisms of both archaeal and euca-
ryal signal peptidases remain largely undefined (18, 447). On
the other hand, in contrast to the eucaryal enzyme, which
functions as part of a larger signal peptidase complex (477),
both bacterial and archaeal signal peptidases apparently func-
tion independently (100, 324). Furthermore, certain archaeal
signal peptidases incorporate a sequence domain of unknown
function, referred to as domain II (323). This domain is found
in the bacterial but not the eucaryal enzyme (100).

The limited experimental analysis of archaeal signal pepti-
dase activity available comes from studies in which the gene
encoding the enzyme from *Escherichia coli* was heterolo-
gously expressed in *Escherichia coli* (302). Isolated bacterial
membranes then served as the source of the archaeal enzyme
in an in vitro signal peptidase assay, using a truncated, poly-
His-tagged version of the *Methanococcus voltae* S-layer protein
as the substrate. Site-directed mutagenesis of the *Methanococ-
cus voltae* enzyme identified three conserved residues, Ser-52,
His-122, and Asp-148, essential for activity (18). The finding
that a second conserved Asp residue (Asp-142) was not crucial
for catalytic function suggests differences in the mechanisms of
the archaeal and eucaryal signal peptidases, since Asp residues
found at both corresponding positions are essential for activity
of the *Saccharomyces cerevisiae* enzyme (324).

Type II signal peptidases are involved in the removal of
signal sequences from lipoproteins (156). However, as noted
above, no archaeal type II signal peptidase has been described,
developing the apparent existence of archaeal lipoproteins (4, 170,
228, 274).

The unique signal sequences of archaeal flagellins are re-
moved by the actions of a signal peptidase reminiscent of
bacterial type IV prepilin peptidases (17, 75), exemplified by
*Pseudomonas aeruginosa* PilD (419). Those translocated non-
flagellar *Sulfolobus solfataricus* proteins bearing the archaeal
flagellin signal sequence also rely on an archaeal version of the
bacterial type IV prepilin peptidase, termed PilD (peptidase
involved in biogenesis of prepilin-like proteins), for their pro-
cessing (2, 6). Site-directed mutagenesis studies have begun to
provide insight into the catalytic mechanism of the archaeal
enzyme (6, 17).

**Amino-Terminal Methionine Removal**

In many instances, the initiator Met residue of a nascent
polypeptide chain (or N-formyl-Met residue in *Bacteria*) is
cleaved by the actions of methionine aminopeptidases (39,
137). While the reason for such processing remains unclear,
several explanations, including facilitation of additional amino-
termal processing (13) and modulation of protein lifetime
(39, 450), have been suggested. Indeed, methionine aminopep-
tidases are essential for the survival of *Bacteria* and yeasts (57,
257).

Methionine aminopeptidases are cobalt-dependent enzymes
that can be divided into two groups, based on sequence com-
parison (14, 210). Type I methionine aminopeptidases are found in
*Eucarya* and *Bacteria*, although the eucaryal enzyme
includes an amino-terminal extension not found in its bacterial
counterpart. *Eucarya* also contain a second isoform of the
enzyme, referred to as type II methionine aminopeptidases. The
two enzyme classes can be distinguished by the presence of an
additional ~60-aminoc-acid-residue carboxy-terminal stretch of
unknown function in type II enzymes (14). Genome sequence
analysis has revealed that *Archaea* contain only type II methi-
onine aminopeptidases, although these lack an amino-terminal
extension found in the eucaryal enzyme (427, 443). Examina-
tion of the crystal structure of the *Pyrococcus furiosus* enzyme
confirmed the similarities of the catalytic domains of the two
methionine aminopeptidase isoforms, despite their limited de-
gree of sequence homology (74, 427).

**Inteins in Archaeal Proteins**

Inteins are genetic elements lying within a protein-encoding
ORF that are transcribed and translated together with their
host to yield an immature precursor protein (135, 260, 337).
Self-splicing of inteins occurs at the posttranslational level,
when the inteins excise themselves from the host protein, al-
lowing the intein-bordering residues of the flanking segments
of the host polypeptide to join through a peptide bond to yield
the mature protein, which is now able to fold and function
properly. Although first discovered in a yeast vacuolar ATPase
(200) and found in proteins across evolution, inteins are most
commonly observed in archaeal proteins; by spring 2005, ap-
proximately 200 inteins had been reported, with almost half
being found in *Archaea* (references 343 and 346 and databases
cited therein).

Inteins are most often found in enzymes involved in DNA
replication and repair. Indeed, the first archaeal intein was
found in a *Thermococcus litoralis* DNA polymerase (344). In-
teins have subsequently been detected in numerous other ar-
chaean DNA polymerases (48, 305, 408, 429, 476) but also in
other proteins (68, 110, 359, 388) from various hyperthermo-
ophilic *Archaea*. This list includes *Methanobacterium thermou-
totrophicum* ribonucleoside diphosphate reductase, which con-
tains the smallest known intein to date (110). Examination of
intein databases reveals that archaeal inteins are not restricted
to hyperthermophilic proteins but are also predicted to exist in
proteins from acidophiles such as *Ferroplasma acidarmanus*,
*Ferroplasma acidiphilum*, and *Picrophilus torridus*, from the
haloarchaea *Halobacterium* sp. strain NRC-1, *Halofex volca-
nii*, and *Haloarcula marismortui* and from the Antarctica-de-
derived methanogen *Methanococcoides burtonii* (343, 346).

Understanding the mechanics of the self-splicing reaction
associated with intein excision began with experiments employ-
ing hyperthermophilic archaeal DNA polymerases (312, 337,
476). In fitting with the elevated growth temperatures of the
host organism, intein splicing from these proteins occurs ine-
sufficiently at temperatures below 25°C. By inserting the coding
sequence of the intein from Pyrococcus sp. strain GB-D DNA
polymerase between genes encoding two foreign proteins, an
intein-containing chimeric precursor was expressed in *Esche-
richia coli* at low temperatures (476). Subsequent splicing of
the purified precursor could be initiated by raising the tempera-
ture. Such studies revealed that all the information needed
for the splicing process is found within the sequences of the
intein and flanking regions and that the excision reaction is
catalyzed by the intein itself, without need for additional fac-
tors.

Since these pioneering studies, further examination of ar-
chaenal inteins has revealed the diversity of intein biochemistry
and offered additional insight into this posttranslational modification. For instance, during the first of four steps involved in the intein self-splicing reaction, the conserved intein amino-terminal Ser or Cys residue undergoes an acyl rearrangement, resulting in the formation of a (thio)ester bond at the amino-terminal splice junction (312, 337). A Methanococcus jannaschii ATPase provided the first example of an intein bearing an amino-terminal Ala residue (140), leading to the description of an alternative splicing pathway (407). In the third step of the self-splicing reaction, cyclization of the intein carboxy-terminal Asn residue leads to peptide bond cleavage and subsequent excision of the intein (312, 337). This cyclization step is facilitated by the intein’s penultimate His residue, which renders the carboxy-terminal Asn residue’s carboxyl carbon more electrophilic (312, 337). Examination of intein cleavage from Methanococcus jannaschii phosphoenolpyruvate synthase and RNA polymerase subunit A’ has provided insight into how inteins lacking this penultimate His residue are processed (58). Furthermore, the presence of inteins in DNA polymerases from Halobacterium sp. strain NRC-1, Pyrococcus abyssi, and Pyrococcus horikoshii bearing carboxy-terminal Gln rather than Asn residues suggests that inteins may self-excite via mechanisms not involving side chain cyclization (58). Indeed, dissection of the pyrocys pathway of the Pyrococcus abyssi DNA polymerase II DP2 subunit intein failed to detect the formation of an intein intermediate containing a cyclized C-terminal Glu residue (288).

Carboxy-Terminal Maturation of Archaeal [NiFe] Hydrogenases

Examination of the Methanococcus voltae vhuU gene product, a component of a [NiFe] hydrogenase, revealed that the translated polypeptide was shorter than predicted by the gene sequence due to a carboxy-terminal cleavage of the protein (404, 405). Although first demonstrated in Methanococcus voltae, cleavage of a carboxy-terminal region downstream of an Asp-Pro-Cys-X-X-His sequence motif by a dedicated endopeptidase following nickel incorporation has since been shown to be a general feature of prokaryotic [NiFe] hydrogenases (53). Differences in the [NiFe] hydrogenase proteolytic maturation step do exist, however, between the bacterial and archaeal systems. In Escherichia coli and other Bacteria, the hydrogenase cleavage motif is followed by a stretch of 15 or more amino acid residues (53). By contrast, in the proteolytic processing of the Thermococcus kodakaraensis hydrogenase α subunit, only four amino acid residues were released from the carboxy terminus following the conserved cleavage motif (199).

Similarly short sequences are also thought to be released from the large subunit of hydrogenases in other archaeal strains, including Methanobacterium thermautotrophicum (380), Pyrococcus furiosus (352), and Thermococcus litoralis (433). Moreover, in Methanobacterium thermautotrophicum and Pyrococcus furiosus, the mature enzymes are proposed to terminate in an Arg rather than a His residue. Differences between predicted molecular weight and the smaller molecular mass measured by SDS-PAGE migration suggest that proteolytic maturation of the Pyrococcus furiosus enzymes does indeed occur (380, 433).

In EchE, the Methanosarcina barkeri homologue of the Escherichia coli hydrogenase 3 large subunit, the cleavage motif also terminates with an Arg residue, although in this case, proteolytic processing does not occur and the Arg thus corresponds to the terminal residue of the protein (241). In contrast, the homologous proteins in Methanococcus jannaschii and Methanobacterium thermautotrophicum also have Arg rather than His residues at this position and yet contain carboxy-terminal extensions that likely undergo proteolytic processing (241). Finally, although the maturation process experienced by archaeal hydrogenases has been less well characterized than the parallel process in Bacteria, archaeal homologues of bacterial enzymes involved in this posttranslational maturation process have been described (199, 428).

OTHER POSTTRANSLATIONAL MODIFICATIONS IN ARCHAEA

Protein Acetylation

The acetylation of selected Lys residues in a protein was first observed almost 40 years ago with eucaryal histones (129), in which this posttranslational modification acts to modulate transcription (474). Since then, acetylation has been reported to modulate the function of many eucaryal and a limited number of bacterial proteins (237, 416, 479). In Archaea, protein acetylation of so-called Alba proteins has been demonstrated by mass spectrometry. These are a family of small (10 kDa) DNA binding proteins first detected in Sulfolobus species (28, 96). They have since been identified in numerous other euryarchaeotal and crenarchaeotal species as well as in Eucarya (28, 458, 460, 466). Upon acetylation of Sulfolobus solfataricus Alba at the Lys-16 position, the affinity of the protein for DNA was lowered (28). In vitro experiments support a role for the Sulfolobus homologue of the eukaryotic histone deacetylase Sir2 in deacetylating Alba, although other deacetylases may act similarly (466).

The enzyme responsible for Alba acetylation has not been identified, although several possible candidates are evident in archaeal genome sequences (355, 466). In a second case, the amino terminus of halocyanin, the small blue copper protein from the haloalkaliphile Natronobacterium pharaonis, has also been proposed to undergo acetylation, in addition to lipid modification (see above), based on the results of mass spectroscopic studies (274).

Protein Ubiquitination

The proteasome is a multisubunit complex responsible for protein degradation in the cytoplasm of eucaryal (452) and archaeal (91, 276, 277) cells. In Eucarya, proteins are targeted for proteasomal degradation by the posttranslational covalent attachment of ubiquitin, a 76-amino-acid-residue polypeptide (69). At present, the putative posttranslational modification that leads to proteasome-mediated protein degradation in Archaea has not been defined. While some reports suggest the existence of ubiquitin in Archaea (275, 300, 473), no direct demonstration of archaeal ubiquitin has been provided, nor have analyses of archaeal genomes identified ubiquitin-encoding genes or genes encoding ubiquitin-transferring proteins. Structural studies, however, have revealed the existence of ar-
chaetal proteins bearing ubiquitin-like folds (33, 375). Nonetheless, it remains to be shown that these proteins participate in proteasome-mediated protein degradation in Archaea.

**Hypusine-Containing Archaeal Protein**

Hypusine [N-ε-(4-amino-2-hydroxybutyl)-L-lysine] is a non-standard amino acid residue found in all Eucarya in a single protein, eukaryotic translation initiation factor 5A (eIF-5A) (329). Hypusine is irreversibly formed soon after the biogenesis of eIF-5A in a two-step posttranslational reaction (331). In the first step, catalyzed by deoxyhypusine synthase, the 4-aminobutyl moiety of the polyamine spermidine is transferred to the ε-amino group of a single specific Lys residue in the eIF-5A precursor protein to form an intermediate, deoxyhypusine. The 4-aminobutyl group of the intermediate undergoes hydroxylation by deoxyhypusine hydroxylase to yield hypusine. The presence of this hypusine residue is essential for eIF-5A function (50, 330, 331).

Hypusine also exists in Archaea, where it too is found exclusively in αIF-5A, the archaeal homologue of eIF-5A. This has been shown in *Halobacterium cutirubrum*, *Methanococcus jannaschii*, *Pyrococcus horikoshii*, *Pyrococcus abyssi*, *Sulfolobus acidocaldarius*, and *Thermoplasma acidophilum* (22, 221, 338, 386, 480). The ability to synthesize hypusine has been studied in *Acidianus ambivalens*, *Pyrodinium octatum*, and *Thermoproteus tenax* (23). The involvement of αIF-5A in archaeral cell growth and the archaecal cell cycle was shown by the arresting action of Nε-guanyl-1,7-diaminoheptane, a hypusination inhibitor, in *Halobacterium salinarum*, *Halofexa mediterranei*, *Sulfolobus acidocaldarius*, and *Sulfolobus solfataricus* (183).

**PROTEOME-WIDE ANALYSIS OF POSTTRANSLATIONAL MODIFICATIONS IN ARCHEA**

Most studies of posttranslational modification of archaeal proteins have thus far relied on individual genes or proteins, the choice of which has been largely guided by substrate availability. In the future, one can expect that our ever-improving ability to describe the entire genomic, transcriptomic, and proteomic profile of an organism will move the study of posttranslational protein modification from the scale of selected proteins to a cellwide perspective. At present, however, such attempts are limited by various factors, including the possible heterogeneity of posttranslational modifications experienced by a given gene product, the relative abundance of a given posttranslationally modified protein variant, and our ability to discern potential posttranslational modifications not encountered previously. Nevertheless, as better tools become available for the simultaneous analysis of the entire protein complement of a cell (187, 270), proteomewide description of posttranslational modifications will soon become routine.

To date, several Archaea have been the subject of proteomic analysis. Such studies have provided novel insights into the adaptations adopted by extremophilic Archaea or have described technical advances in working with extremophilic proteomes (34, 66, 109, 123, 133, 134, 138, 167, 187, 192, 203, 295, 304, 491). Of these investigations, a number have focused on posttranslational modification of archaeal protein targets. In the first of a series of studies addressing the proteome of *Methanococcus jannaschii*, the appearance of a given gene product in multiple positions in a two-dimensional gel electrophoretic system was taken to reflect the posttranslational modification of that polypeptide (134). Accordingly, the multiple positions of Mj0324, which is annotated as an elongation factor (EF-1α), were assumed to correspond to isoforms modified by various degrees of phosphorylation, as had been observed with the eucaryal version of the protein (160). Mj0822, which is annotated as the S-layer glycoprotein, was also found in multiple positions in two-dimensional gels (134). In fact, protein spots corresponding to Mj0822 are among the most strongly stained by Coomassie blue, although the protein is resistant to silver staining, probably due to the presence of glycan moieties. Such differential staining of glycosylated proteins is well documented (186).

In a subsequent proteomic analysis of *Methanococcus jannaschii* as a function of growth conditions or growth stage, examination of peptide fragments derived from either Mj0891 or Mj0891, which are annotated as flagellin B1 and flagellin B2, respectively, revealed condition-specific changes in isoelectric point and abundance (133). Such modulations were proposed but not shown to result from differential degrees of posttranslational modification of the proteins.

More recent studies relying on advances in mass spectrometry for proteome analysis, which included elimination of intermediate proteolytic steps, resulted in a 100% sequence coverage of a set of 72 *Methanococcus jannaschii* proteins (123). Among these proteins, examples of protein acetylation and methylation, amino-terminal proteolytic processing, and disulfide bonds were detected. The applicability of this approach for the rapid determination of expected posttranslational modifications was shown when it was used to test the validity of histone acetylation in *Methanosarcina acetivorans* (125). Despite the proposed presence of a histone-modifying enzyme in this species, no histone modification was detected.

**CONCLUSIONS**

*Archaea* have proven to be a valuable resource in the search for new information on posttranslational protein modification. In several cases, *Archaea* have provided the first prokaryotic examples of modifications once thought to be restricted to Eucarya. The glycosylation of the *Halobacterium salinarum* S-layer glycoprotein represents one such example. In other cases, *Archaea* present previously unknown variations on a given posttranslational protein modification theme, such as the methylation profile of methyl-coenzyme M reductase or the unique lipid moieties attached to haloarchaeal proteins. Elucidation of the enzymatic steps involved in the archaecal version of a particular posttranslational modification event, such as signal sequence cleavage or intein splicing, has dramatically enhanced our understanding of the mechanics of many posttranslational modifications.

With the advent of the proteomic era, when one can determine the protein profile of individual cells, complete physiological systems, and even entire organisms in response to a myriad of conditions, the protein complexity arising from posttranslational modifications should become even more obvious.
If past contributions are any indication, the study of Archaea will continue to expand understanding of the scope, the roles, and the biogenesis of posttranslational modifications that a protein can experience. Such information could provide insight into the strategies adopted by Archaea in the face of the extreme environments in which they can exist. One immediate benefit of relating posttranslational protein modifications to protein structure, stability, and function, together with enhanced tools for the manipulation of archaeal protein expression and secretion, will be the utilization of enzymes from extremophile Archaea tailored for a broad spectrum of biotechnological and industrial applications.

ACKNOWLEDGMENTS

This work was supported by the Israel Science Foundation-Charles H. Revson Foundation (grant 433/03 to J.E.) and the National Science Foundation (BES-0317911 and MCB 0129841 to M.A.). We thank Frank E. Jenney, Jr., for critical reading of the manuscript and the two anonymous reviewers for valuable comments and suggestions.

REFERENCES


Downloaded from http://mmbr.asm.org/ on July 7, 2021


pusine, a unique amino-acid characteristic for eukaryotic translation initiation.


ERRATUM

Posttranslational Protein Modification in *Archaea*

Jerry Eichler and Michael W. W. Adams

*Department of Life Sciences, Ben Gurion University, Beersheva, 84105 Israel, and Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602*


Page 398, Table 3, column 4: The head should read “Bacteria *Campylobacter jejuni*” instead of “Bacteria *Campylorbacter jejuni*,” and the eighth entry should read “Bac” instead of “GalNAc.”