ABSTRACT Proteasomes are ATP-dependent, barrel-shaped proteases found in all three domains of life. In eukaryotes, proteins are typically targeted for degradation by post-translational modification with the small protein ubiquitin. In 2008, the first bacterial protein modifier, Pup (prokaryotic ubiquitin-like protein), was identified in Mycobacterium tuberculosis. Functionally analogous to ubiquitin, conjugation with Pup serves as a signal for degradation by the mycobacterial proteasome. Proteolysis-dependent and -independent functions of the M. tuberculosis proteasome are essential for virulence of this successful pathogen. In this article we describe the discovery of the proteasome as a key player in tuberculosis pathogenesis and the biology and biochemistry of the Pup-proteasome system.

Murine models of tuberculosis have implicated the production of nitric oxide (NO) by activated macrophages as a pivotal part of the immune response, because mice lacking inducible nitric oxide synthase (iNOS−/−) readily succumb to infection with M. tuberculosis (1). Formation of reactive nitrogen and oxygen intermediates (RNIs and ROIs) is toxic to a variety of microbes (reviewed in reference 2). The free radical NO is neutral and hydrophobic, allowing it to pass cellular and bacterial membranes. Reaction with superoxide generated by NADPH phagocyte oxidase results in the formation of the particularly destructive product peroxynitrite. The cytotoxic effects of RNI and ROI include DNA strand breakage, lipid peroxidation, and protein damage (reviewed in references 3, 4, and 5). Although the importance of NO has not yet been irrevocably demonstrated in humans, several in vitro studies suggest a role of host-derived RNI in control of tuberculosis (reviewed in references 2 and 6).

In an effort to deepen the understanding of the bacterial mechanisms in place to avoid elimination by the host immune system, Darwin et al. performed a genetic screen to discover genes that confer resistance to RNI in vitro. After screening 10,100 transposon mutants of H37Rv for hypersensitivity to acidified nitrite, the authors identified five mutants with independent insertions in Rv2097c and Rv2115c, two genes encoding putative accessory factors of a protease called the proteasome (2). Rv2115c was predicted to encode a AAA ATPase related to those associated with the proteasome in eukaryotes. On the basis of its genomic location and a high degree of sequence similarity (81%) to ARC (AAA ATPase forming ring-shaped complexes) from Rhodococcus erythropolis, the authors named Rv2115c mpa (mycobacterial proteasome ATPase). ARC was the first characterized bacterial ATPase with suggested proteasomal function (8). In contrast, Rv2097c was annotated as a hypothetical protein (9) and shared no homology with any protein of known function. Mutations in Rv2097c and mpa caused similar phenotypes in vivo and in vitro, suggesting that both gene products participate in the same pathway. As a result of this observation,
as well as its genomic association with the proteasome core protease-encoding prcBA genes, Rv2097c was named pafA for proteasome accessory factor A (7).

The identification of proteasomal components as potential antituberculosis targets that would sensitize M. tuberculosis to host immunity immediately attracted the interest of the scientific community. Previously, no function had been assigned to the bacterial proteasome, and attempts to reconstitute protein degradation by a bacterial proteasome in vitro had proven unsuccessful (8), likely due to the lack of additional required cofactors. Over the course of the last decade, seminal experiments demonstrated the existence of a bacterial posttranslational protein modifier called Pup (prokaryotic ubiquitin-like protein), which serves as a signal for degradation by the mycobacterial proteasome. Here we summarize what is currently known about the genetics and function of the M. tuberculosis Pup-proteasome system (PPS) and its role in pathogenesis.

**PROKARYOTIC PROTEASOMES**

Proteasomes are self-compartmentalizing, ATP-dependent multisubunit proteases that are responsible for the majority of nonlysosomal protein degradation in eukaryotes. The barrel-shaped complexes have a critical function in protein quality control, stress response, cell cycle control, transcription, metabolism, signal transduction, immune response, and many other biological processes (reviewed in references 10, 11). The “26S proteasome” is composed of two general substructures: the catalytically active 20S core particle (CP) and a 19S regulatory particle (RP) involved in substrate recognition, unfolding, and translocation into the inner core. The CP is assembled of 14 structurally related alpha (α1 to α7) and beta (β1 to β7) subunits, which associate as four stacked heteroheptameric rings. The outer α-rings provide an entry gate to the inner proteolytic chamber composed of β-rings. Of the seven different β-subunits, three have been shown to have distinct catalytic activities as threonine proteases, cleaving peptides after hydrophobic (“chymotrypsin-like activity”), basic (“trypsin-like activity”), or acidic (“peptidylglutamyl-hydrolyzing activity”) residues (reviewed in reference 12). Proteasomes degrade proteins in a highly processive fashion, releasing peptides with a median length of five residues (13). All cleavage sites are sequestered on the inside of the CP, and access is controlled by the RP attached to the α-rings on one or both ends (reviewed in reference 14). The RP is made up of 19 subunits, which form 2 subcomplexes referred to as lid and base. Six AAA+ (ATPase associated with various activities) ATPases that assemble into a hexameric ring at the base of the RP function as the driving force in substrate unfolding and translocation. Other non-ATPase subunits are involved in various aspects of substrate recognition and processing.

Prokaryotic proteasomes were first discovered in the thermoacidophilic archaean Thermoplasma acidophilus, where electron microscopy revealed the existence of proteolytically active particles similar in shape and size to eukaryotic proteasomes (15). In comparison to the eukaryotic 20S CP, the subunit composition of the archaeal proteasome is less complex, which facilitated the determination of the first crystal structure of a proteasome from any of the three domains of life (16). The first bacterial proteasome was identified in the actinomycete R. erythropolis (17). Later, 20S proteasomes were biochemically or genetically characterized in other members of this order, namely Mycobacterium smegmatis (18), Streptomyces coelicolor (19), and Frankia (20). Additionally, sequence information suggested the presence of proteasomes in M. tuberculosis (2) and Mycobacterium leprae (reviewed in reference 21). The notion that bacterial proteasomes are limited to actinomycetes was challenged when the Banfield group discovered peptides encoded by proteasomal genes in biofilm samples of the uncultivated, Gram-negative bacterium Leptospirillum using a mass-spectrometry-based approach (22). Thus far, actinomycetes and Leptospirillum remain the only known bacterial lineages with a proteasome system. It is hypothesized that both have independently acquired the gene clusters encoding this complex protease via lateral gene transfer events (23, 24).

Overall, the architecture of proteasomes from all three domains of life is remarkably similar: cylinder-shaped particles 15 nm in height and 11 nm in diameter are assembled in the α-β-α-β-α-β fashion. While cognate ATPases have been identified and genetically linked to 20S CPs in both archaea and bacteria, robust physical interactions have not been detected, which is suggestive of other factors required for proteolysis (25). In contrast, the eukaryotic 26S holocomplex can be purified. Most prokaryotic CPs consist of only one α (PrCA) and one β (PrCB) subunit, both of which form homoheptameric rings. This arrangement generally limits prokaryotic proteasomes to chymotryptic activity. The M. tuberculosis proteasome, which shares modest sequence similarity with the Thermoplasma CP (32% for both α- and β-subunits) and a high degree of identity (65%) with the Rhodococcus proteasome, is compliant with these features. However, several
observations distinguish the *M. tuberculosis* proteasome from other prokaryotic 20S CPs. All β-subunits are synthesized with N-terminal pro-peptides, which undergo autocatalytic processing to reveal the catalytic threonine residue. Unlike the *Rhodococcus* pro-peptide, which acts as an assembly-promoting factor, the *M. tuberculosis* pro-peptides represent a thermodynamic barrier to this process. Additionally, the *M. tuberculosis* proteasome shows broad substrate specificity despite the presence of a single type of catalytic β-chain. Topological features of the substrate-binding pocket that mimic that of corresponding surfaces in *Saccharomyces* cause this unusual characteristic (26, 27). It is notable that although proteasomes share certain properties with the four classes of energy-dependent proteases commonly found in bacteria, they are structurally and enzymatically distinct from ClpXP, HslUV, FtsH, and Lon (reviewed in reference 28).

**PUPYLATION**

Almost all proteins targeted for proteasomal degradation in eukaryotes are tagged with ubiquitin (Ub). Ub is a small protein of 76 amino acids that can be covalently linked to other proteins via its C-terminal di-glycine motif. Produced as an inactive precursor, Ub is proteolytically processed by deubiquitylases (DUBs) to expose the site of substrate attachment. “Ubiquitylation” (also referred to as ubiquitination) then proceeds in a multistep cascade (Fig. 1). Initially, Ub is activated through adenylation at the C-terminus. A high-energy thioester bond to the active site cysteine of a Ub activating enzyme (E1) quickly replaces the mixed anhydride bond. Subsequently, Ub is transferred to the catalytic cysteine of a Ub conjugating enzyme (E2). Ub can then be transferred to its substrates by a Ub ligase (E3), resulting in the formation of an isopeptide bond between the C-terminal glycine of Ub and the ε-amino group of substrate lysines (reviewed in references 29, 30). Substrate specificity is

**FIGURE 1** Eukaryotic ubiquitin-proteasome system. Ubiquitin (Ub) precursors are processed to expose a C-terminal di-glycine motif. The conjugation-competent Ub is adenyylated and subsequently bound in a high-energy thioester bond by the E1-activating enzyme. Ub is then transferred to the catalytic cysteine of an E2-conjugating enzyme. Ub can be ligated to substrates with the help of E3-ligases. Typically, tetra-Ub chains linked at lysine 48 are recognized by the 26S proteasome. Deubiquitylases can remove Ub from substrates.

doi:10.1128/microbiolspec.MGM2-0008-2013.f1
largely determined by the presence of more than 600 E3 ligases (31) that are subdivided into the HECT (homologous to the E6-AP C-terminus) and RING (really interesting new gene) classes (reviewed in references 32, 33). Ub has seven lysine (K) residues, which allows for the generation of branched poly-ubiquitin chains. Classically, tetra-Ub chains linked at K48 are regarded as the minimal signal for degradation (34), although an increasing number of studies report monoubiquitylation to be sufficient for recognition (35). Polyubiquitylated substrates can be recognized by specific Ub receptors located at the proteasome. Before substrates are fed into the CP, proteasome-associated DUBs can cleave and recycle the poly-Ub chain (reviewed in reference 11).

Although there are striking structural and biochemical similarities between eukaryotic and prokaryotic CPs, it was unclear how proteins are targeted for proteasomal degradation in bacteria, which do not have Ub. Unsuccessful attempts to reconstitute the ATPase-dependent degradation of proteasomal substrates in vitro suggested the involvement of additional cofactors. An important first step was the identification of endogenous mycobacterial proteasome substrates (36). Two proteins, malonyl CoA-acyl carrier protein transacylase (FabD) and ketopantoate hydroxymethyltransferase (PanB), accumulate in an mpa null strain or in wild-type M. tuberculosis treated with the mammalian proteasome inhibitor epoxomicin. Transcriptional analysis showed no difference in fabD or panB transcript levels, suggesting that increased protein abundance is not caused by induced gene expression. To further test these results, epitope-tagged fabD and panB were placed under the control of a heterologous Mycobacterium bovis hsp60 promoter and an Escherichia coli ribosome binding site to control for changes in transcription and translation initiation. Heterologously produced FabD and PanB accumulate in mpa and pafA mutant strains, as well as in wild-type M. tuberculosis treated with proteasome inhibitor (36), thus further strengthening the functional association between Mpa, PafA, and the proteasome. Interestingly, the same study also demonstrated that Mpa levels are regulated by the concerted efforts of Mpa, PafA, and the proteasome. Mpa accumulates in strains producing mutant Mpa lacking its ATPase activity (37), as well as in the pafA null strain or in wild-type M. tuberculosis after chemical inhibition of the proteasome (36). Thus, Mpa is autoregulated via proteasomal degradation.

In order to determine additional components of the M. tuberculosis proteasome system, Darwin and coworkers performed a genetic screen for interaction partners of Mpa, utilizing an E. coli two-hybrid system (38). Screening an M. tuberculosis genomic library of ~100,000 fragments for positive interactions with an Mpa bait fusion, the authors identified Rv2111c, which is the gene located immediately upstream of prcBA and predicted to form an operon with the CP genes. However, addition of Rv2111c to CP and Mpa did not stimulate degradation of FabD in vitro, suggesting the requirement of additional factors for proteolysis. The authors hypothesized that mycobacteria-specific cofactors were required to promote proteolysis. Consequently, they employed a mycobacterial two-hybrid system (39), which enables the study of protein-protein associations in M. smegmatis. Surprisingly, Rv2111c and a proteasome substrate (FabD) showed a strong interaction in this system, which was further tested by copurification of the recombinant proteins from M. smegmatis. Unexpectedly, these experiments yielded a stable covalent complex between the two proteins. Mass spectrometry revealed the formation of an isopeptide bond between the C-terminal residue of Rv2111c and a particular FabD lysine residue (K173). Unlike ubiquitin, Rv2111c does not encode a C-terminal di-glycine motif, but instead contains this motif at the penultimate position, followed by glutamine. Moreover, the C-terminal residue is deamidated to glutamate prior to conjugation. Mutation of the pupylated lysine of FabD (K173A) results in stabilization of the protein in M. smegmatis, indicating that modification of proteins with Rv2111c is a signal for degradation. Based on its functional, albeit not biochemical, analogy to ubiquitin, Rv2111c was named Pup. Interestingly, while numerous pupylated proteins can be observed by immunoblot in total lysates from wild-type and mpa M. tuberculosis strains, no anti-Pup reactive bands are detectable in a pafA null strain, indicating the involvement of PafA in substrate conjugation (40).

In an independent study, Burns et al. showed that a homologue of Rv2111c (MSMEG_3896) acts as a protein modifier in M. smegmatis. Production of epitope-tagged Pup in M. smegmatis was sufficient to modify numerous proteins, demonstrating that the signal for proteasomal degradation is conserved in both pathogenic and nonpathogenic mycobacteria (41). Taken together, these studies showed for the first time that some bacteria are able to utilize a posttranslational tagging system reminiscent of ubiquitylation in eukaryotes, thereby marking a turning point in the understanding of prokaryotic proteasome biology.

Significant progress has been made in elucidating the biochemistry of this posttranslational modifier (Fig. 2).
Pup is an intrinsically disordered protein of 64 amino acids in length (42–44), which is in stark contrast to Ub and other related eukaryotic modifiers that have a conserved β-grasp fold (reviewed in reference 45). Intrinsically disordered proteins lack a compact structure under physiological conditions, although binding-induced folding of the C-terminal half of Pup aids its recognition by Mpa (see below).

At the time of its discovery, pafA did not share any homology with proteins of known function (7, 46). Later, sensitive sequence analysis classified PafA as a member of the carboxylate-amine/ammonia ligase superfamily. Proteins of this class catalyze a two-step ligase reaction that results in an amide linkage between an amino and a carboxylate group. Aravind and colleagues proposed a model in which PafA phosphorylates the γ-carboxylate of the C-terminal glutamate of Pup (PupGlu), priming it for nucleophilic attack by the ε-sidechain amino group of a target lysine, which results in formation of an isopeptide bond (47). In vitro, PafA is sufficient to conjugate PupGlu, to mycobacterial proteasome substrates in an ATP-dependent fashion (48). Site-directed mutagenesis of any residues predicted to coordinate Mg²⁺ or ATP abrogated pupylation in vivo, supporting the predicted model (49). A report by Weber-Ban and coworkers tested the Aravind lab’s hypothesis and found that PafA indeed catalyzes ligation of Pup to substrates via a two-step reaction mechanism. The authors were able to detect a phosphorylated Pup intermediate bound to PafA following ATP hydrolysis (50). Nuclear magnetic resonance spectroscopy revealed that coupling to substrates occurs via the side chain carboxylate of Pup, ruling out any possible involvement of the α-carboxylate of its C-terminus (51).

Purification of epitope-tagged Pup from M. smegmatis yields both PupGlu and PupGln. The observation that the majority of Pup is PupGlu when purified from M. smegmatis but not from E. coli suggested that an enzymatic activity is responsible for this conversion (40). Aravind and colleagues noticed that the M. tuberculosis genome encodes two paralogues of PafA and thus proposed that the ligase assembles into a heterodimer (47). This hypothesis was refuted in a series of experiments that demonstrated that the paralogue of PafA has a distinct function: Dop (deamidase of Pup) converts the C-terminal glutamine residue of Pup to glutamate. Unlike PafA, Dop appeared to utilize ATP only as a cofactor, because no hydrolysis and accompanying release of ADP or AMP was detected. Consolidating this finding, deamidation also proceeds in the presence of a nonhydrolyzable nucleotide analogue, such as ATPγS, albeit at a lower rate (48).

Sequential action of Dop and PafA constitutes a two-step pupylation pathway. However, in vitro, the requirement for Dop can be bypassed, because PafA is sufficient to conjugate PupGlu to target substrates. Actinomycetes encode pup either with a C-terminal glutamine residue, like mycobacteria, or with glutamate (e.g., Corynebacterium spp.), thus eliminating the need for a deamidation step (reviewed in reference 52). This observation hinted at another function for Dop.

**DEPUPYLATION**

In humans, over 70 DUBs are responsible for removing Ub chains from modified substrates. Their action prevents the turnover of Ub together with proteins targeted to the 26S proteasome, so that Ub can be recycled for subsequent conjugation reactions. Under
certain circumstances, removal of Ub rescues a protein from degradation (reviewed in reference 53).

It was hypothesized that mycobacterial cells have the ability to resolve the bond between Pup and its substrate. A series of experiments reported in two independent studies identified Dop, the deamidase of Pup, as a bifunctional enzyme with depupylase (DPUP) activity. Darwin and coworkers noted that numerous pupylated substrates (the “pupylome”) disappeared over time in the presence of ATP but absence of proteasomes. In addition, specific recombinant pupylated proteasome substrates (Pup~FabD and Pup~Ino1 [myo-inositol-1-phosphate synthase]) were “depupylated” in mycobacterial lysates. However, M. tuberculosis lysates of a dop mutant lack the ability to remove Pup from substrates. The authors reasoned that a putative depupylation reaction is chemically identical to the previously described deamidation reaction: both activities involve hydrolysis of an amide bond at the C-terminal residue of Pup. Indeed, purified Dop is able to depupylate different proteasomal substrates in vitro, releasing PupGlu. As for deamidation, ATP binding appears to be critical for functionality of the enzyme (54). Similarly, Weber-Ban and coworkers demonstrated that Dop functions as a depupylase in M. smegmatis. The proteasome substrate Pup~PanB is stable in lysates of an M. smegmatis dop mutant, while depupylation is restored in a complemented strain. Mass spectrometry analysis of the products released after incubation of Dop with Pup~PanB identifies Pup with a C-terminal glutamate residue, demonstrating that the enzyme breaks the isopeptide bond between Pup and a substrate (55). Dop is a strict isopeptidase: unlike certain DUBs, which can cleave linear Ub chains, Dop does not show any reactivity against linear Pup-substrate fusions (54–56). Going forward, it will be critical to determine if depupylation of substrates can rescue them from degradation by the proteasome or if this step facilitates their delivery into the catalytic core of the protease.

Assessment of Dop’s DPUP activity in vivo was complicated by its requirement for deamidation of Pup. In M. smegmatis, ectopic expression of pupGlu in a dop mutant is sufficient to restore pupylation to wild-type levels (57). Interestingly, the pupylation defect of an M. tuberculosis dop mutant cannot fully be restored by production of PupGlu. However, treatment of an M. tuberculosis dop strain expressing pupGlu with the proteasome-inhibitor epoxomicin restores a robust pupylome. This indicates that the DPUP activity of Dop is essential to maintain steady-state levels of pupylation through the recycling of Pup (49).

Interestingly, Mpa contributes to depupylation. Overproduction of Pup results in the rapid degradation of Ino1 in M. smegmatis. In an M. smegmatis mpa mutant, Pup~Ino1 accumulates robustly but is virtually undetectable in wild-type or proteasome deletion strains (54, 58). This suggested Mpa helps facilitate depupylation. Along these lines, Weber-Ban and colleagues demonstrated that Mpa enhances depupylation in vitro. Substrate unfolding by Mpa appears to increase the accessibility of the isopeptide bond for subsequent cleavage (54, 55). This result implies that Dop interacts with pupylated proteins subsequent to or concurrent with Mpa.

The crystal structures of Dop and PafA homologues from Acidothermus cellulolyticus and Corynebacterium glutamicum, respectively, have been solved (59). Withstanding moderate conservation on a primary sequence level (38% sequence identity), the two proteins show high structural similarity. It appears that PafA and Dop arose from a gene duplication event that results in the production of enzymes with opposing activities. One might assume that the shared structural elements of PafA and Dop were conserved to accommodate Pup binding. As previously predicted, the large N-terminal domain of both PafA and Dop adopts a carboxylate-amine ligase fold. A smaller C-terminal domain is unique to these proteins. Even though the active site of both enzymes is located in a particular structural element (referred to as β-sheet cradle) present in the N-terminal domain, mutation analysis has demonstrated that the C-termini of PafA and Dop are critical for function (49, 59, 60).

While multiple differences in the molecular architecture of the enzymes have been identified, site-directed mutagenesis and structural analysis did not identify a singular determinant of Dop activity (59). In another study, Burns et al. utilized an electrophilic trap consisting of Pup modified with a C-terminal glutamine mimic to identify a nucleophilic residue in Dop. Labeling of aspartate 95, an atypical nucleophile for a protease, classified Dop as an unusual amidase whereby the deamidation or depupylation reaction proceeds via an anhydride intermediate (61). Although this model is supported by the crystal structure, further experimentation is necessary to definitively prove this unusual protease activity.

**GENOMIC ORGANIZATION OF THE PUP-PROTEASOME LOCUS**

All PPS-associated genes identified to date are localized in direct proximity to the core proteasome (Fig. 3). In mycobacteria, pup is encoded immediately upstream of prcBA, and the overlap of the stop and start codons of pup and prcB, as well as prcB and prcA, suggest
translational coupling. Usually, dop precedes this operon, while paFA is encoded further downstream, separated by several presumably unrelated genes. The mpa gene is found several open reading frames (ORFs) upstream of dop.

It is notable that corynebacteria encode homologues of all genes associated with (de)pupylation, despite lacking the proteasomal core genes (reviewed in reference 62). The presence of mpa in these organisms is particularly intriguing and may suggest that pupylation can serve as a degradation tag in the absence of the proteasome. It is tempting to speculate that in these organisms, Mpa interacts with other degradation machines to catalyze turnover of pupylated substrates.

Perhaps somewhat surprisingly, little to nothing is known about the expression of proteasome-associated genes. Future studies will need to more closely examine if the PPS is differentially regulated under certain conditions.

SUBSTRATE RECOGNITION BY THE MYCOBACTERIAL PROTEASOME

The RP of the eukaryotic proteasome contains two subunits designated as intrinsic ubiquitin receptors (63). Although a similar arrangement is missing from the M. tuberculosis 20S proteasome, Mpa interacts strongly with Pup (40). Optimal degradation of substrates by the 26S proteasome requires both a proteasome binding signal (e.g., modification with Ub) and an unstructured region to initiate degradation (reviewed in reference 14). Nuclear magnetic resonance studies determined that the C-terminal half of Pup interacts with Mpa (42, 44). Cocrystallization experiments of Mpa and Pup further elucidated the mechanistic details of recognition: Pup binds the N-terminal coiled coil domains of an Mpa hexamer through hydrophobic interactions. This contact induces the formation of an α-helix in the central part of Pup (64). Burns et al. demonstrated that this region of Pup also facilitates interaction with enzymes involved in pupylation, while the N-terminal half is required to initiate unfolding and degradation. Thus, Pup itself provides the two components of the degron necessary for proteasomal degradation (58).

LACK OF A PUPYLASTION MOTIF

Several proteomics studies identified targets of pupylation in M. tuberculosis and M. smegmati.s. Purification of proteins conjugated to Pup coupled with mass spectrometry identified 604 M. tuberculosis proteins, representing ~15% of the predicted proteome. A Pup attachment site was identified for 55 of these proteins (65). Two separate M. smegmati.s pupylome analyses identified 52 (66) and 41 (67) proteins, respectively, with confirmed pupylation sites. In all cases Pup was attached to substrates via a lysine residue, corroborating initial observations. Furthermore, Pup does not appear to form polymeric chains as seen in ubiquitylation. Determining the functional assignment of proteins identified in the M. tuberculosis pupylome, three categories appeared to be overrepresented when compared to their presence in the genome: intermediary metabolism, lipid metabolism, and detoxification/virulence (65). This functional clustering of Pup targets is preserved in M. smegmati.s (67).

The pupylome studies did not reveal a consensus primary sequence recognition motif at the site of modification. Generally, the attachment sites for posttranslational protein modifiers are too diverse to define a consensus sequence. SUMO (small ubiquitin-related modifier) is the only notable exception, because many of
its target proteins contain an acceptor lysine within a ψKxD/E motif, where ψ is a large, hydrophobic residue (reviewed in reference 68). Although Dorrestein and colleagues mapped the pupylation sites on available crystal structures to the periphery of the target subunits (66), it remains a mystery how a single ligase PafA can act on potentially all pupylation substrates and discriminate them from nontargets in the cell.

Darwin and coworkers reconstituted the pupylation pathway in E. coli, a surrogate host that does not natively encode the Pup machinery. Coexpression of pafA and pupGlu is sufficient to pupylate mycobacterial substrates in E. coli. Surprisingly, anti-Pup immunoblots revealed that numerous E. coli proteins were also modified in this system, suggesting that no additional mycobacteria-specific cofactors are required in the process of pupylation. The E. coli puplyome consists of 51 proteins with identified Pup attachment sites. The E. coli substrate phosphoenolpyruvate protein phosphotransferase I (PtsI), which does not have a homologue in mycobacteria, could be pupylated in M. smegmatis, demonstrating that a completely foreign protein can be recognized by a native pupylation system. However, pupylation is not an arbitrary event, because not every protein can be pupylated. For example, dihydrolipoamide acyltransferase (DlaT) is pupylated neither in mycobacteria nor in the E. coli system, despite having 29 lysines. Furthermore, there is a distinct preference for the modified lysine residue in mycobacteria, which is particularly striking in FabD, where all eight lysines are surface-exposed but only one is targeted for pupylation in the native system. The authors speculate that the unmodified lysines are shielded from PafA by other components of the multienzyme fatty acid synthesis pathway in which FabD participates (60). It remains to be determined how protein-protein interactions shape pupylation in general, because the current knowledge cannot rule out involvement of additional factors in target selection or delivery of substrates to the proteasome.

**CHARACTERIZATION OF PUP-PROTEASOME MUTANTS**

*M. tuberculosis* mutants in *mpa* and *pafA* are highly attenuated in a mouse model of infection as measured by reduced bacterial load, gross pathology, and histologically alleviated pneumonitis after a low-dose aerosol challenge (~200 bacilli/mouse) (7). Time-to-death experiments showed that unlike mice infected with wild-type *M. tuberculosis*, which show a median survival of ~1.5 years after infection, mice infected with *mpa* or *pafA* remained alive and healthy while maintaining robust body weight and exhibiting no symptoms beyond normal aging (36, 37). As explained earlier, Nathan and coworkers observed increased *in vitro* sensitivity to acidified nitrite for *mpa* and *pafA* mutants, raising the question if higher susceptibility to RNI *in vivo* was the reason for their attenuation in mouse infections. In mammals, NO is produced by three isoforms of nitric oxide synthase: endothelial (eNOS), neuronal (nNOS), and inducible/immune (iNOS) (reviewed in reference 69). Phagocytes are prolific generators of NO, contributing to the control of various infections, such as *Leishmania major* and Cox sackie B3 virus (reviewed in reference 2). Mice genetically inactivated for iNOS (iNOS−/−) or treated with a chemical inhibitor of iNOS are extremely susceptible to *M. tuberculosis* infection compared to wild-type mice (1). Low-dose aerosol infection of iNOS−/− mice with wild-type *M. tuberculosis* results in death after two to three months (7, 36, 37). In contrast, iNOS−/− mice infected with *mpa* or *pafA* mutant strains live significantly longer than iNOS−/− mice infected with wild-type *M. tuberculosis*, but nonetheless eventually all die (36, 37). Hence, genetic inactivation of iNOS partially restores susceptibility to *mpa* and *pafA* mutants, indicating the importance of these genes to resisting NO.

In another study, Bishai and colleagues screened a library of transposon mutants in the CDC1551 background for mutants displaying morphological changes on agar media. Three independent mutants of *mpa* (MT2175) consistently formed smaller colonies than wild-type *M. tuberculosis* on solid media. To facilitate a more in-depth analysis of the consequences caused by absence of this gene, the authors generated a deletion mutation in the H37Rv *mpa* mutant. The authors also observed significantly less pathology in mice infected with the Δmpa::hyg mutant in rich medium (7H9) was impaired: the average *in vitro* doubling time was determined to be 22 h, compared to 18 h for wild-type CDC1551. High-dose aerosol infection of BALB/c mice shows an infection profile similar to that obtained with the H37Rv *mpa* mutant. The authors also observed significantly less pathology in mice infected with the Δmpa::hyg mutant, characterized by reduced levels of granuloma-like immune infiltrates in the lungs and no changes in body weight. Furthermore, mice infected with the Δmpa::hyg mutant did not succumb to infection until the experiment was terminated at 180 days, while the median survival time for wild-type *M. tuberculosis*–infected mice was 25 days (70).
Disruption of \( \text{dop} \) also sensitizes \( M. \, \text{tuberculosis} \) to RNI in vitro and severely attenuates survival and growth in mice. The degree of attenuation, as assessed by histopathology and bacterial load, is similar among \( \text{dop}, \, \text{mpa} \), and \( \text{pafA} \) mutants, supporting the hypothesis that a common pathway functionally links all three genes (7, 49). Mice infected with any of these mutants do not show the severe tissue destruction and infiltration of lymphocytes characteristic of wild-type \( M. \, \text{tuberculosis} \). Bishai and coworkers dubbed their \( \Delta \text{mpa::hyg} \) strain as an “immunopathology mutant” due to its ability to persist in the lungs without causing significant immune-induced disease. Interferon gamma production is not induced in mice infected with a \( \Delta \text{mpa::hyg} \) mutant compared to wild-type \( M. \, \text{tuberculosis} \) at 3 weeks postinfection. Interestingly, their \( \Delta \text{mpa::hyg} \) strain offered some immunological protection against virulent \( M. \, \text{tuberculosis} \) challenge in mice (70).

The reduced pathology caused by infections with PPS mutants is likely a direct result of a 100- to 1,000-fold lower bacterial burden during the persistent stage of the infection; however, we cannot exclude the possibility that it is the dysregulation of a specific protein or subset of proteins in proteasomal degradation-deficient mutants that causes the changes in pathology.

The genetic screen for mutants hypersensitive to NO did not yield any mutants with insertions in the core proteasomal genes \( \text{prcBA} \). Employing transposon in situ hybridization (TraSH) of a saturating mutant library, Rubin and coworkers identified \( \text{prcBA} \) but not \( \text{mpa} \) or \( \text{pafA} \) among a set of genes required for cellular growth in vitro (71), suggesting that \( \text{prcBA} \) are essential. A conditional gene silencing system using a tetracycline-responsive promoter facilitated studies to test the essentiality of \( \text{prcBA} \) (72). This study indeed showed that \( \text{prcBA} \) is required for normal growth in vitro. Later, a deletion-disruption strain in which \( \text{prcBA} \) was replaced with a hygromycin resistance cassette (\( \Delta \text{prcBA::hyg} \)) was characterized (73). Although the mutant is viable, it grows very slowly. Not surprisingly, infection with either of the \( \text{prcBA} \) mutants results in fewer bacilli in the lungs of C57BL/6 mice compared to infection with wild-type \( M. \, \text{tuberculosis} \). Importantly, the number of bacteria decreased dramatically in later weeks, suggesting that the proteasome core protease is essential for mycobacterial persistence in mice (72, 73).

A \( \text{pup} \) mutant has never been reported. Like \( \text{prcBA} \), it is predicted to be essential based on TraSH (71). Because \( \text{pup} \) is predicted to be cotranscribed and translated with \( \text{prcBA} \), disruption of \( \text{pup} \) would likely result in polar effects on \( \text{prcBA} \) expression. \( \text{pup} \) is undetectable in either a \( \text{dop} \) or \( \text{pafA} \) mutant, presumably because unconjugated Pup is highly unstable (40, 49). Thus, one would presume that these mutants would be phenotypically similar if not identical to a hypothetical \( \text{pup} \) mutant.

It is striking that \( \text{prcBA} \) mutants display an in vitro growth defect absent from \( \text{mpa} \), \( \text{pafA} \), and \( \text{dop} \) null strains (7, 49). To discern if the mycobacterial proteasome has functionality beyond its proteolytic activity, Ehrt and colleagues mutated the catalytic threonine residue of PrcB to alanine (T1A). The active-site mutant assembles into a 20S complex with PrcA, but production of this recombinant proteasome in a \( \Delta \text{prcBA} \) mutant does not rescue proteasomal activity. Surprisingly, production of this inactive proteasome complex is sufficient to complement the hypersusceptibility to NO. In contrast, proteasomal proteolysis is essential for persistence in mice (73). It is unclear how proteasomal activity is connected to certain phenotypes but not others. It is possible that an inactive 20S core structure could act as a “trap” to contain proteins that would normally be degraded. Trapping could result in (partial) inactivation of these proteins and may thus explain how an active site mutant of the \( M. \, \text{tuberculosis} \) 20S can still protect against certain stresses but not others.

It has been hypothesized that the mechanism of protection by the proteasome against NO damage involves the degradation of irreversibly oxidized or nitrated proteins to prevent toxicity to the bacterium. However, \( \text{mpa-}, \, \text{pafA-}, \) and \( \text{prcBA-deficient} \) strains are more resistant than wild-type \( M. \, \text{tuberculosis} \) to hydrogen peroxide (7, 22). Although this could be due to compensatory induction of additional antioxidant pathways, these data suggest that mere turnover of accumulating damaged proteins does not explain the hypersensitivity of PPS mutants to NO.

The role of the CP in virulent mycobacteria appears to be distinct from that in nonpathogenic \( M. \, \text{smegmatis} \), in which disruption of the proteasome does not result in any discernible phenotype (18) (Table 1). Interestingly, unlike \( M. \, \text{tuberculosis} \), \( M. \, \text{smegmatis} \) encodes a homologue of the compartmentalized serine protease Lon, which is the primary protease responsible for protein quality control in \( E. \, \text{coli} \) (reviewed in references 74, 75). Hence, it is possible that production of Lon compensates for the effects caused by deletion of the proteasome in \( M. \, \text{smegmatis} \), but no current experimental evidence has tested this assumption.
ROLE OF THE PROTEASOME IN TRANSCRIPTIONAL REGULATION

A microarray study was conducted to investigate a potential role of the proteasome in transcriptional regulation. Self-compartmentalizing proteases in both prokaryotes and eukaryotes can directly or indirectly regulate the stability of transcription factors (reviewed in references 74, 76). Comparison of the early stationary phase transcriptomes from wild-type *M. tuberculosis* to *mpa* and *pafA* mutants revealed that a shared set of genes (less than 2% of all predicted ORFs) is differentially regulated in the strains deficient in proteasome-mediated degradation, suggesting that Mpa and PafA do not have independent roles affecting gene expression.

Among the genes upregulated in the *mpa* and *pafA* mutants are members of the Zur (zinc uptake regulator) regulon. Zur acts as a zinc-sensing transcriptional repressor that directly binds to a conserved palindromic regulatory element found in eight *M. tuberculosis* promoters. Under zinc-limiting conditions, Zur is released from its operators and gene expression is induced (77). Among the Zur-regulated genes identified in the microarray was the *esx-3* (esat-6, region 3) operon. The Esx-3 cluster, which is required for optimal growth of *M. tuberculosis* in vitro, encodes a type VII secretion system involved in iron and zinc uptake (78, 79). Under normal growth conditions, the levels of free zinc in the cytosol are almost negligible, because zinc-coordinating enzymes and ribosomes largely sequester zinc. Several ribosomal genes are controlled by the Zur regulon in *M. tuberculosis*. These genes encode paralogues of zinc-binding proteins, which lack the metal-binding motif (77). In *Bacillus subtilis*, zinc-binding ribosomal proteins are replaced under zinc-limiting conditions, allowing for mobilization of stored zinc, as well as de novo synthesis of ribosomes (80, 81). Transcript levels of *zur* itself are unchanged in *mpa*- and *pafA*-deficient mycobacteria, and it is unclear how the absence of proteasomal degradation results in the deregulation of this regulon. Maintaining cellular metal homeostasis is critical to sustain key metabolic processes without exposing cells to the toxic effects exhibited by excess zinc or free iron (reviewed in reference 82). Hence, altered intracellular zinc levels throughout the course of infection could contribute to the attenuation of *mpa* and *pafA* mutants in vivo.

### TABLE 1 Summary of phenotypes observed for mutants of the PPS

<table>
<thead>
<tr>
<th>Genomic background</th>
<th>Mutation</th>
<th>In vitro growth</th>
<th>NO resistance</th>
<th>H₂O₂ resistance</th>
<th>In vivo phenotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>M. tuberculosis</em> H37Rv</td>
<td><em>mpa</em>::ΦMycMarT7</td>
<td>Comparable to wild type</td>
<td>Hyper-susceptible</td>
<td>Hyper-resistant</td>
<td>Reduced replication in acute phase; bacterial load in lungs declines during chronic phase; mild pathology</td>
<td>7, 36, 37</td>
</tr>
<tr>
<td></td>
<td><em>pafA</em>::ΦMycMarT7</td>
<td>Comparable to wild type</td>
<td>Hyper-susceptible</td>
<td>Hyper-resistant</td>
<td>Reduced replication in acute phase; bacterial load in lungs declines during chronic phase; mild pathology</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>ΔprcBA::hyg³</td>
<td>Impaired for growth on solid media, small growth defect in liquid media</td>
<td>ND</td>
<td>ND</td>
<td>Persistence defect</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>ΔprcBA::hyg³ + PrcBA-T1A</td>
<td>Comparable to wild type</td>
<td>Comparable to wild type</td>
<td>ND</td>
<td>ND</td>
<td>Reduced replication in acute phase; persistence defect; immunopathology mutant</td>
</tr>
<tr>
<td></td>
<td>Δmpa::hyg³</td>
<td>Smaller colony size, increased doubling time (22 h)</td>
<td>ND</td>
<td>ND</td>
<td>Reduced replication in acute phase; bacterial load in lungs declines during chronic phase; mild pathology</td>
<td>70</td>
</tr>
<tr>
<td><em>M. tuberculosis</em> CDC155</td>
<td><em>dop</em>::ΦMycMarT7</td>
<td>Comparable to wild type</td>
<td>Hyper-susceptible</td>
<td>ND</td>
<td>Reduced replication in acute phase; bacterial load in lungs declines during chronic phase; mild pathology</td>
<td>49</td>
</tr>
<tr>
<td><em>M. smegmatis</em> mc²155</td>
<td>Δ<em>dop</em>::kan⁴</td>
<td>Clumping in liquid media</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>prcBA::kan⁴</td>
<td>Comparable to wild type</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>18</td>
</tr>
</tbody>
</table>

*ND, not determined.*
Microarray analysis further identified a novel copper-responsive regulon, expression of which is down-regulated in the mpa and pafA mutants. A common repressor, RicR (regulated in copper repressor), binds to a palindromic sequence (5′-TACCC-N2-G/AGGTA) located between the -10/-35 sites in the promoter of several physically unlinked genes. In the presence of copper, RicR dissociates from promoters, resulting in transcriptional derepression of ricR itself, mymT (a copper metalllothionein), lpqS (a putative lipoprotein), Rv2963 (a putative permease), and an operon (socAB [small ORF induced by copper A and B]) of unknown function. ΦMycoMarT7 transposon mutant of ricR in the CDC1551 background grows slower than wild-type M. tuberculosis and never reaches the same culture density under routine culture conditions. Perhaps not surprisingly, a ricR mutant is hyper-resistant to copper in vitro, most likely because one or more of the gene products of the regulon combat toxicity associated with excess levels of this transition metal. For example, MymT is robustly produced in a ricR null mutant (83) and binds up to six reduced copper ions to protect M. tuberculosis from copper-mediated toxicity (84).

It is noteworthy that several of the genes underlying regulation by RicR (lpqS, mymT, and socAB) are only found in pathogenic mycobacteria, suggesting that their function is critical to adapt to the specific niche inside the animal host. Interestingly, nonproteasomal ATP-dependent proteases have been implicated in the regulation of virulence gene expression in other bacteria (reviewed in reference 28). This study, as well as others that recently reported the identification of copper-responsive regulators in different pathogens, contribute to an emerging notion of copper as an important antimicrobial defense mechanism exhibited by macrophages (reviewed in reference 85).

The mechanism for deregulation of the RicR and Zur regulons in proteasome-mediated degradation-deficient mutants of M. tuberculosis is unknown at this point. However, it is interesting that the transcriptome of these strains appears to respond to low metal conditions, possibly as a result of accumulation of one or more metal-binding proteins in degradation-deficient strains. In response to inflammation, plasma nutrient availability is altered to restrict the growth of pathogenic invaders: as a result of the acute phase response, zinc levels are decreased, while the amount of copper in the plasma increases (reviewed in reference 86). Hence, adaptation to the particular environment inside the host to achieve metal homeostasis is critical during infection.

OUTLOOK AND REMAINING QUESTIONS

The discovery of the PPS in M. tuberculosis began a period of exciting findings in mycobacteria. Post-translational protein modifiers, previously thought limited to eukaryotes, have since been shown to exist in a similar fashion in archaea, where conjugation of two ubiquitin-like proteins termed SAMP1/SAMP2 (small archaeal modifier proteins) presumably targets proteins for proteasomal degradation (87). While much progress can be denoted in elucidating the biochemistry of pypulation, some major questions remain and others have emerged.

For one, the regulation of pypulation is poorly understood. While a myriad of enzymes is devoted to target selection in the eukaryotic ubiquitylation pathway, one enzyme appears to fulfill this role in mycobacteria. Future studies will have to be directed at understanding whether or not additional cellular factors aid mycobacteria. The mechanism of how the proteasome affects transcription is not understood, it is remarkable that it appears connected to metal homeostasis. Reaction of NO with iron-sulfur center of an enzyme appears to fulfill this role in mycobacteria. Future studies will have to be directed at understanding whether or not additional cellular factors aid mycobacteria.

The presence of a proteasome provides a fitness advantage to mycobacteria during infection. However, the link between proteasome function and pathogenesis has not yet been fully elucidated. Is the accumulation of an individual substrate toxic to the cell, resulting in the observed NO hypersensitivity? Or is the large overlap between nitrosoproteome and pypylome indicative of the proteasome’s role to relieve multifactorial insults in response to stress? Although the mechanism of how the mycobacterial proteasome affects transcription is not understood, it is remarkable that it appears connected to metal homeostasis. Reaction of NO with iron-sulfur clusters has been shown to destabilize the transition metal center. Release of free iron can promote the production of hydroxyl radicals through Fenton chemistry (reviewed in references 3, 4). Additionally, NO can displace copper from M. tuberculosis proteins (84). Hence, it is tempting to speculate that a combination of proteolysis-dependent and independent actions of the proteasome influence the bacterial transcriptional response to oxidative and nitrosative stress.
Completing the characterization of the PPS will undoubtedly broaden our understanding of M. tuberculosis biology. The importance of this pathway for virulence makes the pupulation machinery an attractive drug target. One of the exciting new endeavors of the young field of bacterial proteasome biology will be the design of chemotherapeutics against one of the world’s most successful pathogens.

ACKNOWLEDGMENT
Pup-proteasome research in the Darwin lab is supported by NIH grants AI088075 and HL92774 and by the Irma T. Hirschl Trust. K.H.D. holds an Investigators in the Pathogenesis of Infectious Disease Award from the Burroughs Wellcome Fund.

REFERENCES
The Pup-Proteasome System of Mycobacteria


