Structure and Function of REP34 Implicates Carboxypeptidase Activity in Francisella tularensis Host Cell Invasion*

Received for publication, July 23, 2014, and in revised form, August 30, 2014. Published, JBC Papers in Press, September 17, 2014, DOI 10.1074/jbc.M114.599381

Geoffrey K. Feld, Sahar El-Etr, Michele H. Corzett, Mark S. Hunter, Kamila Belhocine, Denise M. Monack, Matthias Frank, Brent W. Segelke, and Amy Rasley

From the Biosciences and Biotechnology and Physics Divisions, Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California 94550 and the Stanford University School of Medicine, Stanford, California 94305

Francisella tularensis is the etiological agent of tularemia, or rabbit fever. Although F. tularensis is a recognized biothreat agent with broad and expanding geographical range, its mechanism of infection and environmental persistence remain poorly understood. Previously, we identified seven F. tularensis proteins that induce a rapid encystment phenotype (REP) in the free-living amoeba, Acanthamoeba castellanii. Encystment is essential to the pathogen’s long term intracellular survival in the amoeba. Here, we characterize the cellular and molecular function of REP34, a REP protein with a mass of 34 kDa. A REP34 knock-out strain of F. tularensis has a reduced ability to both induce encystment in A. castellanii and invade human macrophages. We determined the crystal structure of REP34 to 2.05-Å resolution and demonstrate robust carboxypeptidase B-like activity for the enzyme. REP34 is a zinc-containing monomeric protein with close structural homology to the metallo-carboxypeptidase family of peptidases. REP34 possesses a novel topology and substrate binding pocket that deviates from the canonical funnelin structure of carboxypeptidases, putatively resulting in a catalytic role for a conserved tyrosine and distinct S1’ recognition site. Taken together, these results identify REP34 as an active carboxypeptidase, implicate the enzyme as a potential key F. tularensis effector protein, and may help elucidate a mechanistic understanding of F. tularensis infection of phagocytic cells.

Francisella tularensis is a Gram-negative bacteria and the cause of the zoonotic disease tularemia, a debilitating, acute disease with high mortality rates if untreated (1, 2). Also known as rabbit fever, due to its historical propensity to infect hunters and trappers, tularemia has recently appeared in diverse geographic regions, where it was either previously unseen (3, 4) or reemerged after more than 20 years dormancy (5). Subspecies of F. tularensis are facultative intracellular pathogens that exhibit remarkably broad host ranges, capable of infecting more than 200 diverse hosts and vectors (6–8). The organism’s geographic and infectious diversity, combined with increased awareness of its potential as a bioterrorism weapon (9) and the absence of an approved vaccine for prophylaxis against tularemia (1) contribute to classification of F. tularensis as a Tier 1 priority pathogen by the Centers for Disease Control.

Berdal et al. (10) first reported the association of F. tularensis with the free-living amoeba, Acanthamoeba castellanii, and proposed that these organisms may act as a potential environmental reservoir for the pathogen in Norway. Using the F. tularensis subspecies holarctica live vaccine strain (LVS), Abd et al. (11) confirmed the ability of F. tularensis to infiltrate A. castellanii cells, further implicating these protozoa in the pathogen’s

The Journal of Biological Chemistry 289: 30668-30679, October 31, 2014

The atomic coordinates and structure factors (code 4OK0) have been deposited in the Protein Data Bank (http://wwpdb.org/).

1 Present address: Theranos, Inc., Palo Alto, CA 94304.
2 To whom correspondence should be addressed: Biosciences and Biotechnology Division, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94550. Tel: 925-423-1284; E-mail: rasley2@llnl.gov.

The abbreviations used are: LVS, live vaccine strain; REP, rapid encystment phenotype; rREP34, recombinant REP34; CPA, CPB, and CPT, carboxypeptidase A, B, and T, respectively; NOV, F. tularensis subspecies novicida U112; PDB, Protein Data Bank; SAD, single wavelength anomalous diffraction; GEMS A, 2-guanidinoethylmercaptosuccinic acid.
environmental persistence. It is well documented that *A. castellanii* can serve as a reservoir for a number of pathogenic microorganisms (12–17) and enhance the virulence of these organisms in murine models of infection (18, 19). Selective pressure for survival from protozoa predation has been suggested to represent a driving force in the evolution of pathogenic bacteria that persist in the environment (17). Given that macrophages in higher organisms possess similar physiology and phagocytic ability as amoebae, protozoa may serve as “training grounds” for intracellular pathogens by priming the bacteria for survival in phagocytic cells (20). Because *F. tularensis* preferentially infects and replicates within host macrophages, probing the Francisella-amoeba interaction may provide insight into pathogenesis as well as environmental persistence.

We recently demonstrated that fully virulent *F. tularensis* isolates infect *A. castellanii* by inducing the amoebae to rapidly encyst (21). The study identified seven putatively secreted *F. tularensis* proteins that may be responsible for this rapid encystment phenotype (REP) in amoeba, a phenomenon required for long term survival of *F. tularensis* in these hosts. Interestingly, two of the genes encoding these REP proteins are deleted in LVS, consistent with the inability of LVS to induce amoeba encystment (21) and with its attenuated infectivity toward macrophages (1). Here, we focus on the gene product of FTN_0149 from *F. tularensis* subspecies novicida U112 (NOV), a REP protein with a molecular mass of 34 kDa (REP34).

The amino acid sequence of REP34 indicates weak sequence homology to the M14 family of metallo carboxypeptidases (22) and even less homology toward metallo carboxypeptidases for which structures are available in the PDB. M14 metallo carboxypeptidases are a diverse and important class of peptidases that catalyze the removal of the C-terminal residue from polyboxypeptidases are a diverse and important class of peptidases which structures are available in the PDB. M14 metallo carboxypeptidase with CPB-like activity from *Porphyromonas gingivalis*, resulted in an avirulent bacterial phenotype (REP) in amoeba, a phenomenon required for long term survival of the pathogen within amoeba, and the implications of other M14 enzymes in virulence, we hypothesized that REP34 does indeed possess an important, yet unknown function. We surmised that a three-dimensional model of REP34 at atomic resolution would aid in devising experiments to test the function and ultimately may provide insights into the role of REP34 during interactions with phagocytic cells.

**EXPERIMENTAL PROCEDURES**

Identification of REP Proteins—*F. tularensis* proteins were identified from bacteria-amoeba coculture subfractions, as described previously (21). Briefly, subfractions found to induce the highest levels of REP were trypsin-digested using standard methods and were analyzed commercially by ProtTech, Inc. (Norristown, PA) using liquid chromatography tandem mass spectrometry (LC-MS-MS). Single polypeptide coverage for each protein was reported for each protein, resulting in 98% confidence. Proteins were identified from LC-MS-MS data using RefSeq from GenBank. In this study, we characterize the cellular and molecular function of REP34 to better understand the biological role of REP in *F. tularensis* infection and environmental persistence. The comprehensive MEROPS peptidase database (27) classifies REP34 and related proteins as a non-peptidase homologue of the M14 family. Given the observation that putatively secreted *F. tularensis* proteins induce the REP phenomenon, which is required for long term survival of the pathogen within amoeba, and the implications of other M14 enzymes in virulence, we hypothesized that REP34 does indeed possess an important, yet unknown function. We surmised that a three-dimensional model of REP34 at atomic resolution would aid in devising experiments to test the function and ultimately may provide insights into the role of REP34 during interactions with phagocytic cells.

**Francisella Growth and Deletion Mutagenesis—Francisella strains were grown overnight in Mueller-Hinton broth (Difco) and plated onto tryptic soy agar as described previously (28). The NOV Δ0149 knock-out mutant was produced as described previously (28). Briefly, a Topo-TA vector (Invitrogen) was used as a surrogate to produce the gene fragment, which was subsequently cloned into a pACYC184 vector in line with a pFNLTTP kanamycin resistance cassette (29). The Δ0149 mutant was selected on modified Mueller-Hinton agar supplemented with kanamycin (15 μg/ml), and gene deletion was confirmed by sequence analysis.

*A. castellanii* Infection Assays—A. castellanii was acclimated to 37 °C for 30 min prior to infection. *A. castellanii* were then infected with NOV overnight cultures at a multiplicity of infection of 10 in triplicate and incubated at 37 °C and 5% CO₂ for 30–60 min. The number of cysts and trophozoites were counted in three random fields of view per well using a light microscope; typically, ~100–200 amoebae were counted in each view.

**Macrophage Infection Assays**—THP-1 cells (ATCC TIB-202) were grown in DMEM (Invitrogen) supplemented with 10% FBS at 37 °C in 24-well tissue culture plates (Costar), similarly to amoeba infection experiments described elsewhere (21). Briefly, THP-1 cells were seeded at a concentration of 2 × 10⁵ cells/well and infected with NOV overnight cultures at a multiplicity of infection of 10. After co-incubation for 30 min, the cells were washed and incubated with fresh DMEM plus 100
Structure and Function of the Carboxypeptidase REP34

μg/ml gentamicin for either 2 h (“entry”) or 24 h (“proliferation”) at 37 °C and 5% CO₂. The cells were then washed once to remove gentamicin and lysed by incubation with 1% saponin (Sigma-Aldrich) for 5 min. Dilutions were plated on supplemented Mueller-Hinton agar to determine viable intracellular cfu counts for the initial inoculum (cfu₂₀), after 2 h (cfu₂₂), and after 24 h (cfu₂₂₄). Entry is expressed as the fraction of cfu₂₂₄ compared with cfu₂₀.

\[
\text{Entry} = \frac{\text{cfu₂₂₄}}{\text{cfu₂₀}} \quad (\text{Eq. 1})
\]

Proliferation is expressed as the fraction of cfu₂₂₄ compared with cfu₂₀.

\[
\text{Proliferation} = \frac{\text{cfu₂₂₄}}{\text{cfu₂₀}} \quad (\text{Eq. 2})
\]

REP34 Cloning and Expression—The NOV gene, FTN_0149, encoding REP34 was cloned into a modified pETBlue expression vector, pETBlueER (Novagen/Merck KGaA, Darmstadt, Germany) by directed cloning (Ndel and BamHI restriction sites). pETBlueER was constructed by replacing the Ncol-Xhol fragment from pETBlue-2 with a 22-bp fragment (5’-CATATGAGAGAGAGGGATCC-3’), creating an Ndel-BamHI MCS. The Xhol site is maintained immediately 3’ of the new 22-bp fragment, but the Ncol site is disrupted by blunting with mung bean polymerase (New England Biolabs, Ipswich, MA).

Both native and selenomethionine REP34 were overexpressed in BL21(DE3) Escherichia coli. Briefly, an overnight starter culture from a single colony was overexpressed in LB, 0.1 mg/liter ampicillin, 37 °C, 220 rpm in shaker flasks and subjected to 1.0 mM isopyrrol β-D-1-thiogalactopyranoside induction at A₆₀₀ = 0.6 for 2 h, pelleted at 4 °C, and stored at −20 °C. Selenomethionine REP34 was produced using the Overnight Express Autoinduction System 2 (Novagen) supplemented with 25 mg/ml l-selenomethionine. Cell pellets were resuspended in buffer A (0.05 mM Na⁺/K⁺ phosphate, pH 8.0, 0.3 mM NaCl, 0.01 mM imidazole) and homogenized on ice using an Emulsiflex-C5 (Avestin Inc., Ottawa, Canada). Lysates clarified by centrifugation at 8000 x g for 20 min were purified by nickel-nitrilotriacetic acid affinity chromatography on a 5-ml HisTrap FF column (GE Healthcare) by first washing with buffer A plus 0.02 mM imidazole, followed by elution in buffer A supplemented with 0.25 mM imidazole. Fractions in the elution peak were further purified by Superdex-200 size exclusion chromatography (GE Healthcare) in 0.01 M HEPES, pH 7.5, 0.05 M NaCl. The dominant peak, judged >90% pure by SDS-PAGE, was concentrated to 9 mg/ml, flash-frozen in liquid N₂, and stored at −80 °C. REP34 + Zn²⁺ samples were supplemented with 1 mM ZnCl₂ prior to performing crystal trials. REP34 site-directed mutants were produced using a modified site-directed mutagenesis protocol (30) (Stratagene, La Jolla, CA) and cloned into pET24b⁺ (Novagen). Expression and purification proceeded as described for WT REP34, except that kanamycin was at 0.05 mg/liter was substituted for ampicillin.

REP34 Crystallization and X-ray Structure Determination—Native and selenomethionine REP34 were screened for crystallization against −300 random conditions designed by CRYSSTOP (31) in 96-well sitting drop vapor diffusion experiments pipetted with the HYDRA+1 (Robbins Scientific, Sunnyvale, CA). Optimization was first achieved using more stringent CRYSSTOP parameters in 96 wells. Further optimization proceeded in manually prepared hanging-drop vapor diffusion trials (32) consisting of 0.5 ml of well solution containing 20% (w/v) polyethylene glycol 4000, 0.1 mM sodium acetate, pH 4.7–5.3, and drops of 1 μl each of protein and well solution at room temperature. Long, icicle-like, heavily twinned crystals appeared overnight with dimensions as large as ~300 × 50 × 30 μm. Crystals were embedded in Paratone-N (Hampton Research, Hayward, CA) and plunged into liquid N₂. X-ray data were collected at the Stanford Synchrotron Radiation Laboratory beamline 12-2 (Pilatus 5M detector) and Advanced Photon Source beamlines GM/CA 23D-D (MARmosaic 300 detector) and Northeastern Collaborative Access Team beamline 24-ID (Pilatus 5M detector) using the microbeam (20 × 20 × 20 μm) (33) at 100 K. Microfocus was essential to collect single crystal diffraction data, which was 90% complete to 2.7 Å resolution. Data were processed using XDS (34), and initial phases were obtained using a single wavelength anomalous diffraction (SAD) data set collected at the selenium edge (12.6 keV, 0.979 Å). Four of the five selenium atoms (with the exception of the N-terminal Met) were identified per molecule using SHELXC/D/E (35), and an initial model was built using BUCCANEER (36). Four REP34 molecules were identified in the asymmetric unit.

Zn²⁺ co-crystal SAD data (9.66 keV, 1.28 Å) collected from a native REP34 crystal extended the resolution to 2.05 Å (2.4 Å with Friedel pairs unmerged). These data, 95% complete, were processed in HKL2000/3000 (37). One zinc ion was identified per molecule using SAD/molecular replacement phasing in PHASER with the initial model as the search criteria (38). Model refinement was carried out in PHENIX using B-factor refinement and atomic displacement parameters individually with 4-fold non-crystallographic symmetry torsion restraints (39). Manual model building was performed with the Coot graphical user interface (40) using simulated annealing composite omit maps calculated in PHENIX and Shake&wARP maps (41) to reduce model bias. The final model consists of four REP34 chains of total root mean square deviation 300 random conditions designed by

Structural Analysis—Structural homologues for REP34 were identified using the 3D-BLAST search algorithm (42). Structural alignments were performed with the MatchMaker function in CHIMERA (43) using the default parameters. Solvent-accessible pockets were calculated using the CASTp online server (44) by submitting single chain coordinates for REP34, CPT (PDB code 1OBR (45)), and CPB (PDB code 1Z5R (46)) with solvent molecules removed. Structures were prepared for Poisson-Boltzman electrostatic potential calculations using PDB entry 2PQR (47), and the calculations were performed using APBS (48).

Carboxypeptidase Activity Spectrophotometry—Hydrolysis of hippuryl-l-arginine (Sigma-Aldrich) was monitored by...
change in absorbance at 254 nm, as described previously (49), with slight modification for 96-well plates, using a BioTek Synergy HT microplate reader (Winooski, VT). Briefly, 3 μl of protein (recombinant REP34 (rREP34) WT and point mutants) was added to each well (final protein concentration in the well was 1.5 μg/ml), and 97 μl of hippuryl-L-Arg and prepared fresh in reaction buffer (0.025 M Tris-HCl, pH 7.65, 0.1 M NaCl, 1.0 mM ZnCl₂) was added and mixed with a multichannel pipette, mixing twice before measurement. For REP34 concentration dependence experiments, 10 μl of protein was added to each well, followed by 190 μl of hippuryl-L-Arg and mixing. Absorbance was measured every 7–10 s for 5 min, and Vₘ for each run was calculated. Michaelis-Menten kinetic parameters were determined using non-linear regression analysis in Microsoft Excel.

**Inhibition of REP34**—The carboxypeptidase B inhibitor 2-guanidinoethylmercaptosuccinic acid (GEMSA) was purchased from Calbiochem (EMD Millipore, Billerica, MA). 1 mg/ml WT REP34 was incubated with a 10-fold molar excess of GEMSA in buffer B plus 1.0 mM ZnCl₂ at room temperature for 1 h. The carboxypeptidase B inhibitor 2-guanidinoethylmercaptosuccinic acid (GEMSA) was purchased from Calbiochem (EMD Millipore, Billerica, MA). 1 mg/ml WT REP34 was incubated with a 10-fold molar excess of GEMSA in buffer B plus 1.0 mM ZnCl₂ at room temperature for 1 h. The molecular masses of REP34 both with and without GEMSA were analyzed using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry on a Voyager-DE system (PerSpective Biosystems).

**RESULTS**

**REP34 Contributes to Amoeba Encystment**—NOV genes corresponding to proteins believed to be responsible for REP (21) were identified using LC-MS-MS and are presented in Table 1. We constructed a knock-out strain of NOV lacking the FTN_0149 gene (Δ0149) and infected *A. castellanii* with both Δ0149 and wild-type NOV (Fig. 1A). The Δ0149-infected amoebae show significantly less (*p < 0.0001*) encystment compared with wild type NOV-infected amoeba. Although 86 ± 11% of amoeba encysted in response to NOV, only 9 ± 4% formed cysts in response to Δ0149. Significant (*p < 0.005*) encystment was restored by the addition of 10 μg/ml recombinant REP34 (rREP34) to the knock-out strain because 21 ± 8% of amoeba encysted under these conditions. A comparable fraction of amoeba formed cysts for both the addition of rREP34 alone and a phosphate-buffered saline (PBS) control, 4.3 ± 0.3 versus 5 ± 1%, respectively. Thus, our data suggest that FTN_0149 is important for amoebic encystment during *F. tularensis* infection, but the exogenous FTN_0149 gene product, rREP34, is only modestly able to complement the encystment-inducing activity and, in the absence of other *F. tularensis* components, is unable to induce REP.

**FIGURE 1.** REP34 is implicated in *F. tularensis* host cell invasion. A, ability of NOV strains to induce amoeba to encyst, where Fraction Encysted = cysts/total. Error bars, S.D. of three independent technical replicates (three field-of-view counts, n = 3–9). B, ability of *F. tularensis* strains to enter (Fraction Entry; left scale and black bars) and subsequently survive (Fraction Proliferate; right scale and gray bars) in human monocyte hosts. Histograms and error bars, mean and S.D. of cfu in three independent counts (n = 3). Significance was determined by Student’s t test (***, *p < 0.0001*, **, *p < 0.005*, *, *p < 0.05*). C, crystal structure of REP34 rendered as ribbons in the standard view for funnels (24) and colored by secondary structure: helices (orange), strands (green), and loops (purple). The Zn²⁺ is represented as a cyan sphere. N and C termini are indicated. D, REP34 depicted as a translucent molecular surface, colored as in B and rotated ~90° around the x axis. All figures were prepared in CHIMERA (43).

**REP34 Contributes to *F. tularensis* Entry into Human Monocytes**—Given the hypothesis that protozoa may serve as “training grounds” for intracellular pathogens and the observation that Δ0149 is significantly impaired in its ability to induce REP in *A. castellanii* (Fig. 1A), we hypothesized that the Δ0149 strain might also exhibit a reduced ability to infect human macrophages. To test this hypothesis, we infected human monocytes with NOV, Δ0149, and LVS. Two time points were taken to investigate the ability of *F. tularensis* to enter (2 h) and to replicate (24 h) in human monocytes. Host cell entry was significantly reduced (*p < 0.0001*) in Δ0149 because this mutant strain entered monocytes at a rate of 9.7 ± 0.9% versus 53 ± 14% for NOV. The Δ0149 knockout had less of an effect.
Structure and Function of the Carboxypeptidase REP34

Table 2

Data collection and refinement statistics

|                | REP34 (−Zn<sup>2+</sup>)<sup>a</sup> | REP34 (+Zn<sup>2+</sup>)<sup>a</sup> |
|----------------|----------------------------------------|----------------------------------------|
| PDB code       | 40KO                                   |                                        |
| Data collection|                                        |                                        |
| Wavelength (keV)| 12.6                                   | 9.66                                   |
| Space group     | P1                                     | P1                                     |
| Cell dimensions |                                        |                                        |
| a, b, c (Å)     | 62.79, 63.53, 82.58                    | 62.71, 63.31, 81.69                    |
| Resolution (Å)  | 77.64–2.73 (2.80–2.73)<sup>b</sup>    | 22.00–2.05 (2.09–2.05)<sup>b</sup>    |
| R<sub>merge</sub>, R<sub>free</sub> | 0.085 (0.307)                          | 0.078 (0.437)                          |
| I/σ(I)          | 7.3 (2.3)                              | 9.0 (2.1)                              |
| Completeness (%)| 89.9 (82.2)                            | 94.9 (83.1)                            |
| Redundancy      | 3.1 (3.1)                              | 2.6 (2.6)                              |
| Phasing         |                                        |                                        |
| Figure of merit (FOM) | 0.468               |                                        |
| Figure of merit after Buccaneer | 0.408               |                                        |
| Buccaneer      | 0.727                                  |                                        |
| Refinement      |                                        |                                        |
| No. of reflections | 28,542 (1889)            | 68,913 (4503)                          |
| R<sub>work</sub>, R<sub>free</sub> | 21.3/23.3 (26.1/32.1) |                                        |
| No. of atoms    | Protein 7736                           | 9180                                   |
|                 | Ligand 0                               | 20                                     |
|                 | Water 0                                | 240                                    |
| B Factors       | Protein                                 | 42.4                                   |
|                 | Ligand                                 | 47.4                                   |
|                 | Water                                  | 40.2                                   |
| Root mean square deviations | Bond lengths (Å) | 0.004                                 |
|                 | Bond angles (degrees)                  | 0.849                                  |
| Molphilp statistics | All-atom clashscore                 | 2.05                                   |
|                 | Ramachandran favored                   | 97.9                                   |
|                 | Ramachandran allowed                   | 2.1                                    |
|                 | Ramachandran outliers                  | 0.0                                    |
|                 | Overall score                          | 1.10                                   |

*<sup>a</sup> Data collected from a single selenomethionine crystal.
*<sup>b</sup> Data collected from a single native crystal.
*<sup>c</sup> Values for the highest resolution shell are shown in parenthesis.

(p < 0.05) on *F. tularensis* proliferation in monocytes because Δ0149 increased at a rate of 620 ± 60% over 22 h post-entry versus 1100 ± 300% for NOV. For comparison, LVS, which lacks two REP-encoding genes and does not induce amoeba encystment (21), can enter macrophages similarly to NOV, 48 ± 18%, but proliferation within host cells was severely reduced, 41 ± 32%. Thus, the data suggest that FTN_0149 is involved in the entry of human monocytes but has a marginal effect on either bacterial replication or survival.

**REP34-Zn<sup>2+</sup> Co-crystallization**—Purified rREP34 formed snowflake-like crystals with multiple lattices, which proved recalcitrant to single crystal diffraction studies. Single-crystal data could only be collected with a microbeam x-ray source on larger crystal clusters grown from larger volume drops. An initial REP34 structure was determined by Se-SAD phasing of selenomethionine-derived REP34 to 2.7 Å resolution with a triparental carboxypeptidase T and porcine pancreatic CPB (PDB code 1Z5R (46)) as a representative carboxypeptidase B. Despite sequence identities between REP34 and CPT or CPB of less than 15%, both CPT and CPB align well structurally with REP34 (backbone root mean square deviations of 2.5 Å) (Fig. 2A). Canonical Zn<sup>2+</sup>-carboxypeptidases possess three principal structural elements that contribute to substrate recognition and binding. First, a long loop connecting α3 and α4 (L<sub>3</sub>4) forms the outermost lip of the active site, providing the solvent-exposed rim of the funnel. Second, the active site residues surrounding the Zn<sup>2+</sup> coordination site catalyze the hydrolysis chemistry of the enzyme. Third, the P1′ substrate recognition site is lined with residues that provide shape and chemical complementarity to the C-terminal side chain of a substrate, giving rise to the substrate specificity of the enzyme.

The REP34 structure exhibits both similar and unique features relative to its closest structural homologues. When observed from the standard view for funnelins (24), the active site Zn<sup>2+</sup> is positioned at the C terminus of the central four-strand parallel sheet, as expected (Fig. 1C). In contrast, REP34 also contains additional structural elements, namely an N-terminal helix, α0<sup>4</sup> and an insertion helix, αΔ′ (Fig. 2, A and B) on the side opposite the central parallel sheet. While the canonical funnelin loop L<sub>α3α4</sub> that defines the right edge of the substrate-binding cleft is quite long, containing ~60 residues, the homologous loop in REP34 is much shorter, containing only ~20 residues (Fig. 2C). Furthermore, REP34 α6 is positioned away from the active site cleft, whereas the corresponding helix and preceding loop in metallocarboxypeptidases forms a wall of the active site and donates a conserved tyrosine, which stabilizes the P1′ amino acid of the substrate. Taken together, this combination of the two additional helices, substantially shorter L<sub>α3α4</sub> and rearranged α6 produces a distinct cleft leading to the Zn<sup>2+</sup>+, when compared with the “funnel” shape of recognized metallocarboxypeptidases (Fig. 1D). Thus, REP34 possesses the central α/β hydrolase fold indicative of metallocarboxypeptidases but has unique structural features that probably redefine how the enzyme interacts with its target peptide.

As a result of these structural differences surrounding the active site cleft, REP34 possesses a larger substrate cavity compared to X-ray fluorescence signatures consistent with the presence of zinc. A new data set collected at the zinc edge (9.66 keV) was used to determine the zinc-containing REP34 structure by molecular replacement using the zinc-free structure as the search model. Continuous electron density covering almost all four copies of REP34 in the asymmetric unit was evident, and residues 4–303 were placed, except for a disordered loop, 233–242, which is disordered in all four copies in the asymmetric unit (Fig. 1, C and D).

**REP34 Possesses an α/β Hydrolase-like Fold**—REP34 is structurally homologous to Zn<sup>2+</sup>-carboxypeptidases with funnelin-like architecture, possessing the α/β<sub>ε</sub>/β<sub>ε</sub> topology of α/β hydroases. To simplify analysis and discussion, here we use CPT from *Thermoactinomyces vulgaris* (PDB code 1OBR (45)) as a representative homologous bacterial carboxypeptidase T and metallocarboxypeptidases but has unique structural features that probably redefine how the enzyme interacts with its target peptide.
pared with CPT and CPB (Fig. 3). This cavity occupies a volume of 1793 Å³ for REP34, which is more than twice as large as the 732-Å³ cavity in CPB and more than 3 times the 432-Å³ volume for CPT, as calculated by CASTp (44). Factors that appear to contribute to the larger cavity include a wider rim at the protein surface, a wider internal cavity where the putative substrate is expected to reside during hydrolysis, and an additional narrow cavity extending through the entire molecule. The cavity linking the substrate-binding pocket to the rest of the pocket on the opposite side of REP34 is too narrow to fit a polypeptide. Overall, REP34 contains a large substrate-binding cavity that appears accessible for polypeptide binding and subsequent hydrolysis.

**REP34 Active Site—** The AZ n²/H₁₁₀₀¹ coordination site represents the centerpiece of the REP34 active site, consisting of side chains His⁸⁶, Glu⁸⁹, and His¹⁷¹ and a coordinated water molecule in a distorted tetrahedral geometry (Fig. 4A). Anomalous difference maps from SAD data collected at the zinc edge pinpoint the metal ion’s position. The map clearly reveals the location of the Zn²⁺, and prominent 2Fo – F₁ density corresponding to the coordinating side chains is also evident. Therefore, REP34 contains a Zn²⁺ coordinated by an HXXEH motif typical of metallo-carboxypeptidases.

In addition to the residues coordinating the Zn²⁺, funnelins have five other conserved residues involved in catalysis (Fig. 2C). Using the standard CPA/B sequence numbering, these include Glu²⁷⁰, which polarizes the nucleophilic water and donates a hydrogen to the amino group of the P₁/-leaving group; Asn¹⁴⁴ and Arg¹⁴⁵, which stabilize the P₁/C terminus during catalysis; Tyr-2⁴⁸, which stabilizes the P₁ amide and the P₁/C terminus; and Arg-¹²⁷, which polarizes the carbonyl of the P₁ amino acid during the tetrahedral intermediate. REP34 residues Glu-²⁷², Asn-¹³⁸, Arg-¹³⁹, and Arg-¹²⁹ overlap quite well with CPA/B residues Glu-²⁷⁰, Asn-¹⁴⁴, Arg-¹⁴⁵, and Arg-¹²⁷, respectively (Fig. 4B). The loop containing CPA/B Tyr-²⁴⁸ is largely absent from REP34 (Fig. 2C); however, Tyr-⁸⁸ may be a compensatory substitution in REP34, affording at least some of the intermediate stabilization that Tyr-²⁴⁸ provides in CPA/B. Sequence and structural secondary assignments for REP34, CPT, and CPB. Active site and Zn²⁺-coordinating residues are highlighted (red and blue, respectively), and REP34 α₀ and α₄⁺ are identified with stars (yellow).
Structure and Function of the Carboxypeptidase REP34

FIGURE 3. Substrate binding pocket comparison for REP34 and related carboxypeptidases. Shown are surface representations of REP34 (purple, left), CPT (PDB 10BR [45], blue, middle), and CPB (PDB 1ZG7 [46], cyan, right), where substrate binding pocket surfaces (gold) are computed using the CASTp server [44]. The top views are of each molecule looking into the pocket, and the bottom views are 90° rotations around the x axis with a clipping plane placed at the center of the pocket. For CPT and CPB, the top view is the standard view, and for REP34, the view is −90° rotation around the x axis relative to the standard view. CPB contains a bound inhibitor (sticks, pink) to aid in visualization of the binding pocket-substrate interaction.

FIGURE 4. Active site of REP34. A, close-up view of the Zn²⁺ binding site colored by atom (carbon (purple), nitrogen (blue), oxygen (red), and Zn²⁺ (cyan)), where Zn²⁺ and the coordinated water molecule are depicted as spheres. Electron density for a simulated annealing composite omit map calculated at the end of refinement to 2.05 Å resolution and contoured to 1.0σ is depicted as a mesh (gray). An anomalous difference map revealing the location of the Zn²⁺ ion and calculated at the end of refinement to 2.4 Å resolution is represented by a mesh (orange) contoured to 8.5σ. B, MatchMaker-generated structural alignment of the active sites for REP34 (purple) and CPB (PDB Code 1ZG7 [46], yellow), with atoms colored as in A. Residue numbering corresponds to REP34, except for CPB Tyr-248. C, MatchMaker-generated overlay of REP34 with CPA4-cleaved hexameric peptide (PDB code 2PCU [50], green) and CPT-bound sulfamoyl arginine (PDB code 4GM5, orange). The P1’ residue of the former and the sulfamoyl group of the latter bound substrates are omitted for clarity. The composite omit electron density as in A, contoured to 1.6σ for the ordered acetate bound to REP34, is depicted as a mesh (gray).

carboxypeptidase A4 (CPA4) in complex with a hexapeptide post-bond cleavage (PDB code 2PCU [50]) and a CPT-sulfamoyl arginine co-complex. However, the presence of o4’ in REP34 severely clashes with the polypeptide, implying that a substrate must adopt a different conformation when bound to REP34, as compared with other structurally related funnelin substrate complexes (Fig. 4C). To test this hypothesis, we incubated REP34 with GEMSA, a small molecule known to inhibit CPB enzymes. Subsequent mass spectrometry analysis revealed that both the GEMSA-incubated and free enzyme samples were the same mass, indicating that GEMSA does not covalently modify REP34. Albeit a negative result, this is consistent with the structural differences that impart a unique substrate-binding pocket in REP34 relative to CPB, occluding the binding of the inhibitor GEMSA to REP34.

REP34 was crystallized in the presence of 0.1 M sodium acetate, and strong electron density consistent with a single acetate anion is present in the active site of each REP34 molecule in the asymmetric unit (Fig. 4C). Acetate resembles the C terminus of peptides, and the high concentration in the crystallization condition probably resulted in it being trapped in the active site of REP34. The oxygen atoms of the acetate carboxyl interact with the side chains of REP34 Arg-129, Asn-138, and Arg-139. These residues correspond to CPA/B amino acids that participate in P1’ stabilization in a similar manner. Furthermore, α alignment of CPA/B/T structures containing inhibitors or molecules with either cleaved P1’ amino acids (50) or P1’ mimics (46, 51, 52) (PDB entry 4GM5) places these carboxyl groups in close proximity to the bound acetate ion in the REP34 structure. Therefore, this trapped acetate is consistent with the P1’ C terminus of a cleaved target peptide in the product-bound state. Carboxypeptidase Activity of REP34—From the crystal structure, it appears that REP34 could be an active carboxypeptidase, despite the MEROPS annotation, so we assayed for carboxy-
peptidase activity using rREP34. We measured the change in absorbance at 254 nm caused by the release of hippuric acid from the model substrates hippuryl-L-Phe and hippuryl-L-Arg for CPA and CPB, respectively. Incubation of REP34 with hippuryl-L-Phe did not result in any appreciable release of hippuric acid over a range of protein concentrations (data not shown). Therefore, we deduced that the protein has little, if any, CPA activity. In the presence of hippuryl-L-Arg, we measured considerable hydrolysis, with absorbance changes similar to previous studies of purified porcine pancreatic CPB (49). By varying the concentration of REP34 in the reaction, we find that hippuryl-L-Arg hydrolysis is enzyme concentration-dependent, plateauing around 6 μM/ml protein, for 1.0 mM substrate (Fig. 5A). Varying the substrate concentration while holding [REP34] at 1.5 μg/ml, we observed a kinetic profile that fits a Michaelis-Menten model for the range of substrate concentrations measured (Fig. 5B). We were unable to measure a constant background signal for hippuryl-L-Arg concentrations higher than 5.0 mM; therefore, these measurements were excluded from further analysis. Furthermore, the calculated \( K_m \) for these kinetic measurements was ~18 mm, which is considerably higher than an achievable substrate concentration under these conditions. Although we cannot confirm whether REP34 follows a classical kinetic profile for the model substrate hippuryl-L-Arg, we conclude that REP34 appears to have robust CPB-like activity.

As a structural homologue of CPB (Figs. 2 and 4), we hypothesized that REP34 has a similar catalytic mechanism. To test this hypothesis, we produced alanine point mutations for each of the homologous putative catalytic residues in REP34: Y88A, R129A, N138A, R139A, and E272A. We then measured the ability of these mutants to hydrolyze hippuryl-L-Arg relative to WT (Fig. 5C). For these experiments, the concentration of initial hippuryl-L-Arg was kept constant at 1.0 mM, and hydrolysis was determined as the maximum change in absorbance. The REP34 mutants E272A and Y88A showed the greatest loss in activity, resulting in 18 ± 2 and 23 ± 13% activities relative to WT, respectively. The REP34 substitutions R139A and N138A have activities of 33 ± 7 and 49 ± 16%, respectively. In contrast, the mutant R129A hydrolyzed hippuryl-L-Arg similarly to WT, with a 97 ± 15% activity. Therefore, residues Glu-272 and Tyr-88 contribute significantly to REP34 catalysis, whereas Arg-129 appears to perform a nonessential role in catalysis.

Identification of Putative P1 Recognition Site—The metallo-
peptidases confer substrate specificity by shape and chemical complementarity to the P1 residue in the substrate

![Figure 5](https://example.com/fig5.png)

**FIGURE 5. REP34 carboxypeptidase activity.** The rate of hydrolysis of the carboxypeptidase B model substrate hippuryl-L-Arg was measured by absorbance change at 254 nm for REP34 concentration dependence (A) and hippuryl-L-Arg concentration dependence (B). The line represents nonlinear regression best fit according to the Michaelis-Menten equation. 1 unit = hydrolysis of 1.0 μmol of hippuryl-L-Arg/min at pH 7.65, 25 °C. C, percentage activity of REP34 active site mutants to hydrolyze hippuryl-L-Arg, as compared with WT REP34. Error bars, S.D. for multiple trials (n = 3–5).

![Figure 6](https://example.com/fig6.png)

**FIGURE 6. S1′ recognition site in REP34.** REP34 and CPB (PDB 1Z5R (45)) are aligned and colored as in Fig. 4. The region of interest is enclosed in the box. A, putative S1′ recognition sites in REP34, with the molecular surface colored gray, B (top), electrostatic surface rendering of REP34, where computed negative, neutral, and positive charges are colored red, white, and blue, respectively. The active site cleft is boxed and magnified (bottom), where the positions of the Zn\(^{2+}\), Glu-256, and Asp-177 are indicated. A clipping plane above the cleft and approximately parallel to the page was applied for clarity. C, comparison of the S1′ sites in REP34 and CPB, with the molecular surface for CPB colored gray. CPB Zn\(^{2+}\) is colored dark gray; all labels correspond to CPB elements except for REP34 Cα4.
vides additional polar surface to both putative S1’ sites (Fig. 6A).

Distance geometry homology suggests that Glu-256 is more consistent with the P1’ complementing residue. The distances from the δ- and γ-carbons for REP34 Glu-256 and Asp-177 to the Zn^{2+} are 12.0 and 6.8, respectively, whereas Asp-175, Asp-178, and Asp-276 are more remote (>13 Å). The corresponding distances for the P1’ basic side chain recognition residues in CPB (Asp-255) and CPT (Asp-260) are 10.5 and 10.1 Å, respectively. Therefore, REP34 Glu-256 appears well positioned to complement a positively charged P1’ residue during substrate binding. Whereas REP34 Asp-177 is considerably closer to the Zn^{2+} than either aspartic acid in CPB or CPT, it also points more directly toward the Zn^{2+}; thus, it may interact less strongly with a P1’ basic residue.

DISCUSSION

Infection assays involving Δ0149 implicate a role for REP34 in both the induction of REP and the entry of F. tularensis in human monocytes (Fig. 1, A and B). As shown in Fig. 1A and elsewhere (21), F. tularensis NOV induces REP in amoeba, and Δ0149 NOV strain has a significantly reduced ability to induce this phenotype. The ability of the FTN_0149 gene to induce encystment may be more complicated than simply expressing and secreting REP34 because rREP34 does not restore REP activity to wild-type NOV levels, and the addition of rREP34 alone did not elicit REP. It is certainly plausible that some of the other REP-inducing genes (Table 1) may also encode for effector proteins, and the Δ0149 mutant may disrupt the REP machinery in ways supplementary to REP34 expression. Thus, we suspect that REP34 may work in concert with other REP proteins to induce REP.

It has been suggested that interactions between intracellular pathogens, including F. tularensis, and environmental amoebae might have resulted in these bacteria evolving phagocytic cell evasion tactics (17), which may in turn enhance their ability to infect macrophages (14). Our data suggest that REP34 plays a significant role in host cell entry but less of a role in either the survival or replication of the pathogen within the host cell (Fig. 1B). Interestingly, the LVS strain, which lacks two of the genes encoding REP3s, is capable of entering monocytes similarly to NOV but lacks the ability to survive and replicate. Further research into the targets of REP34 and other REP proteins may shed light on their roles in mediating interactions of F. tularensis with phagocytic cells.

The crystal structure of REP34 presented here provides preliminary evidence for the carboxypeptidase activity of REP34. The presence of a α/β hydrolase core renders REP34 a close structural homologue of the M14 funnelin family of Zn^{2+}-dependent carboxypeptidases with overall root mean square deviations of ~2.5 Å despite sequence similarities of less than 15%. The MEROPS peptidase database classifies REP34 and similar proteins as inactive members of the M14 family, probably because a structural rearrangement of the active site results in a displacement of the REP34 counterpart for a catalytically important Tyr (Tyr-248 for CPB and Tyr-255 for CPT). We show that REP34 is, indeed, an active carboxypeptidase (Fig. 5). Furthermore, REP34 Tyr-88, which is highly conserved in the subfamily, may represent a compensatory substitution for the canonical catalytic Tyr. This is consistent with the Y88A mutant showing an ~80% decrease in carboxypeptidase activity relative to wild type (Fig. 5C).

Analysis of the REP34 structure provides clues as to the nature of the enzyme’s activity and substrate binding. Compared with other funnelins, REP34 has a shorter Lα3α4 and contains additional helices α0 and α4’, giving rise to a larger predicted binding cleft (Fig. 3) and a unique structural scaffolding for the active site (Figs. 4C and 6C). These differences probably result in novel enzyme-substrate interactions. Biochemical data provide strong evidence for CPB-like activity for REP34 (Fig. 5), and the structure provides a basis to predict a S1’ recognition site on the enzyme, which presents negative charges to complement the predicted basic P1’ residue of the target substrate (Fig. 6, A and B). In addition, a trapped acetate ion is present in the active site (Fig. 4C) and probably approximates the C terminus of a P1’ amino acid of the target protein in the post-hydrolysis state. Taken together, these data strongly support the CPB-like activity of REP34 in catalyzing the hydrolysis of a C-terminal basic residue from an unknown target protein.

Based on proposed catalytic mechanisms for other metallo-carboxypeptidases, we developed a schematic mechanism (Fig. 7) for the hydrolysis of a C-terminal basic residue that is consistent with the REP34 structure (Figs. 2–4 and 6) and biochemical data (Fig. 5) presented here. Contacts between substrate polypeptide atoms and REP34 residues Asn-138, Arg-139, and Glu-272 as well as the catalytic water and Zn^{2+} during the catalytic cycle are assumed to be roughly identical to funnelin active site residues described elsewhere (24). The exceptions are Tyr-88, which adopts a geometry different from that of its presumed functional homologue in CPB, Tyr-248 (Fig. 4B), and Arg-129, which from mutagenesis experiments appears to be superfluous to the catalytic mechanism (Fig. 5C).

Structural superposition of CPB and REP34 places CPB Tyr-248 in close proximity to the acetate position in the REP34 structure, which is assumed to mimic the P1’ C terminus (Fig. 4C). This Tyr is presumed to stabilize the polypeptide substrate by contacting both the negative charge of the P1’ carboxyl and the P1 amide. Although Tyr-88 can still potentially stabilize the P1 amide, Tyr-88 is not close enough to interact with the acetate. Instead, Tyr-88 is within hydrogen-bonding distance of the catalytic water. In the transition state, the catalytic water adds to the P1 peptide carbon, forming a tetragonal bond; Tyr-88 would putatively contact the P1 carboxyl position. This positions the Tyr-88 hydroxyl to stabilize the “anionic hole” created by the partial negative charge on the carboxyl following water addition. CPB Arg-127 is expected to provide this polarizing stability in the canonical mechanism; however, the homologous residue in REP34, Arg-129, does not appear to play as critical a role because REP34 R129A activity is not reduced in vitro (Fig. 5C). Given that the solution side chain pKa values of Tyr and Arg are greater than 10, it is not unreasonable for Tyr-88 to provide a positively charged dipole stabilization of the negatively charged substrate carbonyl, which is chemically similar to the role of Arg-127 in CPA/A. Therefore, we propose that this alternative role for REP34 Tyr-88 in stabilizing the substrate P1 carbonyl may compensate for the R129A substitu-
A co-crystal structure containing REP34 bound to a substrate mimic would be informative in further defining the catalytic mechanism of the enzyme and provide a template for the rational design of countermeasures against REP34 and REP34-like peptidases in other pathogenic organisms.

A BLAST search of nonredundant genetic homologues to REP34 reveals a number of bacterial genera that also may produce similar enzymes. Homologous sequences (~50%) are found in Vibrio species, including the causative agent of cholera, V. cholerae, as well as the primarily marine bacteria Shewanella and Alteromas. Generally, these genes are annotated as either hypothetical proteins or some derivative of zinc-containing metalloenzymes, including hydrolases and desuccinylases; however, all of these sequences contain a conserved Tyr corresponding to REP34 Tyr-88, nestled between Zn$^{2+}$-coordinating His and Glu residues. In the case of V. cholerae, the putative homologue (gene code VCA0936) is classified as a conserved hypothetical protein and is present on chromosome 2. VCA0936 has been implicated in biofilm production by avirulent V. cholerae strains, which may contribute to the tenacity of these organisms in their marine ecosystems (53). It remains to be seen whether this gene product is active in Vibrio or other bacteria as we demonstrate for Francisella and whether it plays any role in virulence and/or environmental persistence.

How does the carboxypeptidase function of REP34 provide a mechanistic basis for the ability of F. tularensis to infect phagocytic host cells? In the absence of specific cellular and molecular mechanistic detail, we can only speculate as to the exact nature of the enzyme’s biological role; however, numerous possibilities exist. For instance, REP34 may have a specific host target protein that, when cleaved, would interrupt lysis signaling pathways or promote phagocytosis of the bacterial invader. Given the structural deviation from conventional funnels, REP34 may have a more specific target than a general basic C terminus, thus limiting the polypeptide sequence space of the host target. Furthermore, the identification of other potential REP-inducing proteins in F. tularensis suggests a complementary role among the various proteins because rREP34 alone is insufficient to complement the REP phenotype to a level approaching wild-type NOV (Fig. 1A). Future efforts to elucidate the host cell targets of the peptidase REP34 and other REP proteins may further shed light on their functions during F. tularensis infection of phagocytic cells.

Acknowledgments—We thank D. Finley and the Francis Laboratory at the University of California, Berkeley, for assistance with MALDI-TOF; D. Cascio and M. Sawaya of the UCLA-Department of Energy Institute for Genomics and Proteomics for assistance with crystalllographic software; M. Soltis and the 12-2 beamline staff at the Stanford Synchrotron Radiation Laboratory; C. Ogata and the GM/CA 23-ID beamline staff and M. Capel, K. Rajashankar, N. Sukumar, J. Schuermann, I. Kourinov, F. Murphy, S. Banerjee, K. Perry and D. Neau of the Northeastern Collaborative Access Team beamline 24-ID at the Advanced Photon Source.

REFERENCES

1. Oyston, P. C. (2008) Francisella tularensis: unravelling the secrets of an intracellular pathogen. J. Med. Microbiol. 57, 921–930
2. Clemens, D. L., and Horwitz, M. A. (2007) Uptake and intracellular fate of Francisella tularensis in human macrophages. Ann. N.Y. Acad. Sci. 1105, 160–186
3. Eliasson, H., and Back, E. (2007) Tularaemia in an emergent area in Sweden: an analysis of 234 cases in five years. Scand. J. Infect. Dis. 69, 880–889
4. Gürçan, S., Eskiciak, M., Varol, G., Uzun, C., Tatman-Ötkun, M., Sakru, N., Karadenizli, A., Karagöl, C., and Ötkun, M. (2006) Tularaemia re-emerging in European part of Turkey after 60 years. Jpn. J. Infect. Dis. 59, 391–393
5. Feldman, K. A., Enscore, R. E., Lathrop, S. L., Matyas, B. T., McGuill, M., Schriefer, M. E., Stiles-Enos, D., Dennis, D. T., Petersen, L. R., and Hayes, E. B. (2001) An outbreak of primary pneumonic tularemia on Martha’s Vineyard. New Engl. J. Med. 345, 1601–1606
6. Ellis, J., Oyston, P. C., Green, M., and Titball, R. W. (2002) Tularemia. Clin. Microbiol. Rev. 15, 631–646
7. Nigrovic, L. E., and Wingerter, S. L. (2008) Tularemia. Infect. Dis. Clin. North Am. 22, 489–504, ix
8. Petersen, J. M., Mead, P. S., and Schriefer, M. E. (2009) Francisella tularensis: an arthropod-borne pathogen. Vet. Res. 40, 7
9. Oyston, P. C., Sjostedt, A., and Titball, R. W. (2004) Tularaemia: bioterrorism defence renews interest in Francisella tularensis. Nat. Rev. Microbiol.
51. Cho, J. H., Kim, D. H., Chung, S. J., Ha, N. C., Oh, B. H., and Yong Choi, K. (2002) Insight into the stereochemistry in the inhibition of carboxypepti-
dase A with N-(hydroxyaminocarbonyl)phenylalanine: binding modes of an enantiomeric pair of the inhibitor to carboxypeptidase A. *Bioorg. Med. Chem.* 10, 2015–2022

52. Reverter, D., Fernández-Catalán, C., Baumgartner, R., Pfänder, R., Huber, R., Bode, W., Vendrell, J., Holak, T. A., and Avilés, F. X. (2000) Structure of a novel leech carboxypeptidase inhibitor determined free in solution and in complex with human carboxypeptidase A2. *Nat. Struct. Biol.* 7, 322–328

53. Mueller, R. S., McDougald, D., Cusumano, D., Sodhi, N., Kjelleberg, S., Azam, F., and Bartlett, D. H. (2007) *Vibrio cholerae* strains possess multiple strategies for abiotic and biotic surface colonization. *J. Bacteriol.* 189, 5348–5360