bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Shotgun Immunoproteomic Approach for the Discovery of Linear B Cell Epitopes in Biothreat Agents *Francisella tularensis* and *Burkholderia pseudomallei*

Patrik D'haeseleer[1*], Nicole M. Collette[1], Victoria Lao[1], Brent W. Segelke[1], Steven S. Branda[2], Magdalena Franco[1]

[1]Lawrence Livermore National Laboratory, Livermore, CA, USA

[2]Sandia National Laboratories, Livermore CA 94550, USA

***Correspondence:**
Patrik D'haeseleer
dhaeseleer2@llnl.gov

Keywords: Francisella, Burkholderia, immunoproteome, B cell epitope, antigen, peptide vaccine

Abstract

Peptide-based subunit vaccines are coming to the forefront of current vaccine approaches, with safety and cost-effective production among their top advantages. Peptide vaccine formulations consist of multiple synthetic linear epitopes that together trigger desired immune responses that can result in robust immune memory. The advantages of peptide epitopes are their simple structure, ease of synthesis, and ability to stimulate immune responses by means that do not require complex 3D conformation. Identification of linear epitopes is currently an inefficient process that requires thorough characterization of previously identified full-length protein antigens, or laborious techniques involving genetic manipulation of organisms. In this study, we apply a newly developed generalizable screening method that enables efficient identification of B cell epitopes in the proteomes of pathogenic bacteria. As a test case, we used this method to identify epitopes in the proteome of *Francisella tularensis* (Ft), a Select Agent with a well-characterized immunoproteome. Our screen identified many peptides that map to known antigens, including verified and predicted outer membrane proteins and extracellular proteins, validating the utility of this approach. We then used the method to identify seroreactive peptides in the less characterized immunoproteome of Select Agent *Burkholderia pseudomallei* (Bp). This screen revealed known Bp antigens as well as proteins that have not been previously identified as antigens. The present workflow is easily adaptable to detecting peptide targets relevant to the immune systems of other mammalian species, including humans (depending upon the availability of convalescent sera from patients), and could aid in accelerating the discovery of B cell epitopes and development of vaccines to counter emerging biological threats.

INTRODUCTION

Development of an effective vaccine against a biothreat agent or emerging pathogen is a costly and cumbersome process that can take years to decades to complete. The identification of antigens that stimulate protective immunity against a pathogen represents a significant bottleneck in the typical vaccine development process. Our study addressed the need to accelerate this process by testing the

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

feasibility of a platform for efficient identification of immunoreactive peptides that could be utilized as candidates for development of peptide-based vaccines.

Peptide-based vaccines represent the next generation of vaccines, with great potential to provide rapid protection against biothreats and emerging pathogens. Peptide vaccine formulations consist of multiple synthetic linear epitopes that together trigger immune responses resulting in robust immune memory. This multi-epitope approach can be broadly protective across divergent strains (*e.g.*, the first universal influenza vaccine to enter phase III clinical trials was a peptide vaccine) and effective for pathogens with complex life cycles (*e.g.*, several malaria peptide vaccines are currently in clinical trials) (1–3). Due to their lack of complex secondary and tertiary structure, peptides can be easily synthesized, multiplexed into vaccine formulations, and efficiently screened for efficacy. Consequently, peptide-based vaccines represent promising candidates for rapid response medical countermeasures against infectious disease.

Current strategies for epitope identification depend upon detection of epitopes within an individual full-length protein, a low-throughput approach that requires prior knowledge of the antigenic protein and its sequence. Technologies to screen for epitopes at the whole proteome level have been developed (e.g., proteomic microarrays, phage and yeast display); however, these technologies require extensive use of synthetic biology and other time-consuming methodologies (e.g., library construction, peptide/protein array preparation, heterologous protein expression) (3–8). Another major disadvantage of display technologies and use of non-native expression systems is that these methods do not reliably replicate the native properties of the antigenic proteins, including their post-translational modifications, which can lead to inaccurate results.

In this study, proteome-wide screening for linear B cell epitopes was achieved using native proteomes isolated from the pathogen of interest and convalescent sera from infected animals. This strategy holds several advantages over the currently available methods for epitope discovery: It does not require prior knowledge of antigenicity or antigen structure, and obviates need for complex and laborious experimental techniques such as preparation of display libraries and heterologous protein expression. Our approach was designed to enable identification of the protein antigen and, importantly, the antigenic regions within the identified antigen, such that these short linear peptides can be immediately synthesized and tested for efficacy in vaccine formulations.

In this study, we focused on two intracellular bacterial pathogens, *Francisella tularensis* (Ft) and *Burkholderia pseudomallei* (Bp), organisms which pose a high risk for misuse as bioweapons and therefore are considered Tier 1 Select Agents by the US Centers for Disease Control and Prevention. The mortality rates of both pathogens are high, and there is currently no licensed vaccine available for either agent (9–11). Humoral immunity plays an important role in developing immune protection to both of these intracellular pathogens, making them good model organisms for the purposes of this study (12–17). In addition, the immunoproteome of Ft has been thoroughly characterized (10,18,19), such that the previously published data could be compared to the datasets generated in our study. We leveraged a merged dataset of 164 previously identified antigens, corresponding to ~10% of Ft proteome. The Bp immunoproteome is not as well characterized compared to that of Ft: our reference dataset contained only 61 previously identified seroreactive proteins, corresponding to ~1% of the Bp proteome (20,21). Consequently, the dataset resulting from the Bp screen has revealed many proteins that have not been previously categorized as antigens.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

MATERIALS AND METHODS

Bacterial strains and culture conditions

Francisella tularensis SCHU S4ΔclpB ("Ft-ΔclpB") was a generous gift from Dr. Wayne Conlan (National Research Council Canada). Stock cultures were prepared by growing Ft-ΔclpB on Chocolate II Agar plates supplemented with hemoglobin and isovitalex (BD 221169) for 48 hours at 37°C. Bacteria were harvested by scraping confluent lawns into Mueller Hinton (MH) broth containing 20% (w/v) sucrose, and stored at -80°C at a concentration 10^8 - 10^9 CFU/mL. *Burkholderia pseudomallei* mutant ΔpurM ("Bp82") was obtained from BEI resources (NR-51280). Frozen stocks were prepared by growing the bacteria to log phase in Luria-Bertani (LB) broth, adding glycerol to achieve 20% (w/v) with the bacteria at a final concentration of 10^8 - 10^9 CFU/mL, and storing aliquots at -80°C. For immunizations, the Ft-ΔclpB and Bp82 bacterial stocks were thawed and diluted in sterile phosphate-buffered saline (PBS) to the specified concentrations used for dosing. For protein extraction purposes, Ft-ΔclpB and Bp82 were propagated to log phase in MH and LB broth, respectively.

Protein extraction and peptide preparation

Ft-ΔclpB and Bp82 were grown to log phase in 300 mL of MH broth or LB broth, respectively, at 37°C with shaking (250 rpm). The bacteria were harvested by centrifugation at 3200 x g for 10 min at 4°C, washed once with 10 mL of PBS, and the pellet flash frozen using dry ice. The bacteria in the pellet were lysed by subjecting them to two freeze-thaw cycles (alternating between room temperature and dry ice). For protein extraction, the lysate was mixed with Bper Complete Bacterial Protein Extraction Reagent (Thermo Fisher Scientific, cat# 89822), and the mixture incubated at room temperature for 15 min with rotational shaking. The mixture was then subjected to two rounds of sonication (1 sec pulses, timed output 10 sec, at 50% power) using a Heat Systems Ultrasonics sonicator (model W-385), and centrifuged at 16,000 x g for 10 min. Proteins were precipitated with acetone and washed twice with ethanol. Air-dried protein pellets were solubilized using 8M urea and Protease Max surfactant (Promega, V2071), then digested with trypsin (Promega, V5111) using the in-solution digestion protocol provided by the manufacturer (Promega, TB373). Completion of the trypsinization reaction was confirmed by gel electrophoresis. The trypsin-digested proteins were filtered using 10K MWCO concentrators (Pierce) at 10,000 x g for 20 min at 20°C, and the filtrates (purified peptides) stored at -20°C.

Mice and immunizations

Mouse immunization studies were carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals and the National Institutes of Health. Appropriate efforts were made to minimize suffering of animals. All animals were housed in ABSL2 conditions in an AAALAC-accredited facility, and the protocol (Protocol 270, renumbered 284, approved 10/09/2017) was approved by the LLNL Institutional Animal Care and Use Committee (IACUC). For immunization, 6 week-old female specific-pathogen-free BALB/c-Elite and C57BL/6J-Elite mice (Charles River) were injected subcutaneously with 10^6 CFU Ft-ΔclpB (BALB/c and C57BL/6J), or intradermally with 10^7 CFU Bp82 (BALB/c), and boosted at 2 weeks. No adjuvants were used. Matched PBS-dosed controls were included for each injection route. Course of infection was monitored by performing daily health scoring and weight measurements. Mice that developed infection wounds (Ft only) were topically treated with Dakin's solution to encourage wound healing, and allowed to remain on test so long as they did not meet humane endpoint criteria (any mice with ~20% body weight loss or overt signs of morbidity were humanely euthanized). Sera from euthanized

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

mice were excluded from analysis due to lack of immunity to the pathogen. Convalescent sera were harvested from resilient mice at 4 weeks post-infection, *via* cardiac puncture terminal bleeding under inhaled isoflurane anesthesia followed by blood fractionation [centrifugation at 3800 x g for 15 min in microtainer serum separator tubes (BD)]. Sera were stored at -80^{o}C.

SDS-PAGE and Western analysis

Western analysis was performed to confirm seropositivity of infected mice. Bacterial lysates were prepared using Bper Complete Bacterial Protein Extraction Reagent (Thermo Fisher Scientific, cat# 89822), combined with Laemmli loading buffer (BioRad), and boiled at 95°C for 5 min. Samples were loaded onto 4-15% acrylamide gels (Mini-Protean TGX, BioRad) and separated by electrophoresis at 120 V for 1 hr. The proteins were transferred from the gels to nitrocellulose membranes (BioRad). Membranes were blocked with Tris-buffered saline plus 0.05% Tween 20 (TBS-T) plus 5% nonfat dry milk, at room temperature for 1 hr or at 4^{o}C for 16 hrs. The membranes were hybridized with mouse sera at 1:500 dilution in TBS-T plus 5% milk, at room temperature for 2 hrs; washed three times with TBS-T; and then incubated with goat anti-mouse antibodies conjugated to HRP (Pierce, prod#1858413), at 1:5000 dilution in TBS-T plus 5% milk, at room temperature for 1 hr. After three TBS-T washes, the membranes were developed using SuperSignal^{TM} West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific).

Enzyme-linked immunosorbent assay (ELISA)

ELISA was performed to assess the level of seropositivity of infected mice. Wells were coated with bacterial lysates and incubated at 4°C for 16 hrs. After three washes with PBS plus 0.1% Tween-20 (PBS-T), sera from infected mice diluted to 1:100 with PBS were added to the wells and incubated at room temperature for 1 hr. Following four PBS-T washes, the wells were incubated for 1 hr with Recombinant Protein A/G peroxidase (Pierce, cat#32490) diluted at 1:5000 with PBS. After four PBS-T washes, 1-Step ABTS Substrate Solution (cat# 37615) was added, and after 15 min incubation any colorimetric changes in the wells were detected using a microplate reader (Tecan M200 Pro).

Affinity purification of immunoreactive peptides

Magnetic beads coated with protein G (Invitrogen, cat#10007D) were used to capture antibodies from sera from infected mice, following the manufacturer's "Dynabeads Protein G immunoprecipitation" protocol (MAN0017348). The antibody-coated beads were then incubated with purified peptides at room temperature for 45 min. Following three PBS washes, immunoreactive peptides were eluted from the beads using citrate buffer (pH 3). Input, unbound, and eluate fractions were flash frozen with dry ice and stored at -20°C. As a negative control, antibodies from uninfected (PBS treated) mice were used to detect any background resulting from nonspecific binding of peptides to beads or antibodies.

Mass spectrometry (MS)

The input, unbound, and eluate fractions recovered from antibody-coated beads (see preceding section) were desalted using an Empore SD solid phase extraction plate; lyophilized; reconstituted in 0.1% TFA; and analyzed *via* LC-MS/MS by MS Bioworks (Ann Arbor, Michigan), using a Waters M-Class UPLC system interfaced to a ThermoFisher Fusion Lumos mass spectrometer. Peptides were loaded on a trapping column and eluted over a 75 µm analytical column at 350 nL/min. Both columns were packed with Luna C18 resin (Phenomenex). A 2 hr gradient was employed. The mass spectrometer was operated in a data dependent HCD mode, with MS and MS/MS performed in the

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Orbitrap at 60,000 FWHM resolution and 15,000 FWHM resolution, respectively. The instrument was run with a 3 sec cycle for MS and MS/MS.

MS data processing

Data were analyzed using Mascot (Matrix Science) with the following parameters: Enzyme: Trypsin/P; Database: UniProt *F. tularensis* SCHU S4 or UniProt *B. pseudomallei* strain 1026b (forward and reverse appended with common contaminants and mouse IgG sequences); Fixed modification: Carbamidomethyl (C); Variable modifications: Oxidation (M), Acetyl (N-term), Pyro-Glu (N-term Q), Deamidation (N/Q); Mass values: Monoisotopic; Peptide Mass Tolerance: 10 ppm; Fragment Mass Tolerance: 0.02 Da; Max Missed Cleavages: 2; Mascot DAT files were parsed into Scaffold Proteome Software for validation, filtering and to create a non-redundant list per sample. Data were filtered using 1% protein and peptide FDR and requiring at least one unique peptide per protein.

Bioinformatic analysis

Each experiment typically consisted of three sets of data: “Input” (total bacterial peptides without affinity purification), “Control” (peptides purified from beads coated with antibodies from uninfected mice), and “Experiment” (peptides purified from beads coated with antibodies from infected mice).

LC-MS/MS data were analyzed at the peptide level based on the Total Ion Current (TIC, total area under the MS2 curve), rather than rolling up peptide scores into a protein abundance metric as would be done in standard proteomics. Input datasets were first normalized against each other based on median ratios for the peptides occurring in every Input dataset. The more sparse Control and Experiment datasets were then normalized against their respective Input dataset based on median ratios as well. Since each animal can be expected to raise a different set of antibodies, we counted how often peptides occurred more abundantly in the experiment vs control, rather than focusing on the average log fold change in abundance. Each peptide was assigned an enrichment score, by adding +1, 0, or -1 based on whether the experimental peptide level was greater than, equal to, or lower than the control level in each experiment. Statistical significance was evaluated by randomizing this matrix of +1/0/-1 values.

Average Amino Acid Conservation Scores (AAACS) were calculated using the ConSurf web server (22) with default parameter values, using near full-length protein structure homology models from SWISS-MODEL or crystal structures from PDB where available. The AAACS for the peptide is the average conservation score for the residues in the peptides, with negative scores indicating more highly conserved regions (23).

In addition to AAACS, we also scored peptides based on how many complete sequenced genomes of pathogenic *B. pseudomallei* and *F. tularensis* they occurred in, similar to the conservation analysis in EpitoCore (24). We downloaded proteomes for all 110 *B. pseudomallei* strains with complete genome sequences available through NCBI. For *F. tularensis*, 36 strains with complete genomes were available through NCBI, but several of these corresponded to the less-pathogenic *novicida*, *holartica* and *mediasiatica* subspecies, so we decided to focus exclusively on the 17 available *F. tularensis* subsp. *tularensis* complete genomes. We identified homologs with ≥90% sequence identity to the proteins containing our top scoring peptides in Tables 1 and 2, and then scored each peptide based on how often they had a 100% identical hit in each homolog.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

We used two state-of-the-art computational B cell epitope prediction tools to evaluate all of the peptides in our proteomic data that match the proteins in Tables 1 and 2. Peptides were submitted to the iBCE-EL web server for scoring (25). In addition, proteins were submitted to the Bepipred Linear Epitope Prediction 2.0 tool on the IEDB website (26), and peptides were then scored based on their average predicted residue score. For selected proteins with an available structure model, we also used the Discotope 2.0 web server for prediction of potentially discontinuous B-cell epitopes from protein 3D-structure (27).

RESULTS

Overview of immunoproteome screen

In this study, we tested the feasibility of proteome-wide screening for linear B cell epitopes using peptide extracts from target bacteria and sera from infected animals. The method requires: (1) isolation of peptides from lysates generated from the target bacteria; (2) challenge of the host (in this case, mouse) with the target bacteria, followed by collection of convalescent serum; (3) mixing of the bacterial peptides and convalescent serum, to allow peptide antigens to bind to their cognate antibodies in the serum; and (4) recovery of bound peptides for identification through mass spectrometry (Figure 1). We applied this method to two bacterial Select Agent pathogens: *Francisella tularensis* and *Burkholderia pseudomallei.* Infection with attenuated strains of these pathogens [*F. tularensis* SCHU S4ΔclpB and *B. pseudomallei* ΔpurM (strain Bp82)] has been shown to stimulate development of protective immunity against their corresponding fully-virulent parental strains (*F. tularensis* SCHU S4 and *B. pseudomallei* K96245, respectively) (28,29), suggesting that convalescent sera recovered from hosts infected with these attenuated pathogens must contain protective antibodies.

Briefly, proteins purified from pathogen lysates were digested with trypsin to generate a peptide library. Mice were infected with a sublethal dose of Ft-ΔclpB or Bp82, and immune status assessed through measurement of seroreactivity to pathogen lysate *via* enzyme-linked immunosorbent assay (ELISA) or Western blot analysis (Figure 2). Antibodies purified from the convalescent sera of infected mice were immobilized on magnetic beads and then incubated with pathogen-derived peptides to allow formation of antigen-antibody complexes. Peptides recovered from the immobilized antibodies were identified *via* liquid chromatography coupled with tandem mass spectrometry.

Bioinformatic identification of enriched antigenic peptides

The peptides recovered from infected mice (Experiment peptidome) were compared to those recovered from mock-infected mice (Control peptidome); a total of 8 pairs of peptidomes were collected for Ft, and 9 pairs for Bp. For Ft, we found that 44 of the recovered peptides had an enrichment score of 6 or greater, whereas only 20.1 +/- 6.1 peptides would be expected at random ($p=5x10^{-5}$). For Bp, 46 peptides had an enrichment score of 6 or greater, whereas only 17.8 +/- 4.3 peptides would be expected at random ($p=3x10^{-12}$). The enriched peptides included those derived from a number of protective antigens identified in previous studies, as well as predicted outer membrane and extracellular proteins (Tables 1 and 2). There were many examples of multiple enriched peptides originating from the same protein (highlighted in bold in the tables), a further indication that enrichment was not random but rather due to immune response to a discrete set of bacterial proteins.

Note that we used C57BL/6J mice for two of the eight Ft experimental samples, because of previously reported differences in protection and antibody response after immunization of C57BL/6J

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

and BALB/c mice with Ft-ΔclpB by Twine *et al* (30). Analyzing the BALB/c Ft samples separately yielded a very similar set of results as in Table 1, but with lower p-value for the enrichment due to the smaller number of samples (results not shown). Therefore, we decided to combine the data and focus on antibody responses in common between both strains of mice. Although Twine *et al* reported an antibody response against chaperonin protein GroL only in BALB/c mice, our data shows that there are several GroL epitopes that are enriched in samples from both mouse strains (see Table 1 and Figure 4).

Immunoproteomics analysis of the antibody response to *F. tularensis* using human or mouse sera has identified 164 antibody targets out of a total of 1667 proteins (~10% of the entire Ft proteome) (10,18,19). Out of the 1923 peptides that have hits in at least two Ft datasets, 876 peptides match known antigenic proteins. Given those numbers, we would expect only 20 such peptides to show up at random in our list of 44 in table 1, but instead we observe that 38/44 peptides in the list correspond to known antigens - and almost two-fold enrichment ($p=2.79x10^{-9}$). The immune response to *B. pseudomallei* has not been studied in as much depth as for *Francisella*. So even though Bp with 6203 protein coding genes has a genome that is more than three times as large as that of Ft, we found only 61 known antigens identified in previous studies (20,21) (~1% of the entire proteome). Our list of 47 top Bp peptides in Table 1 includes one known antigen, which does not qualify as a statistically significant enrichment primarily because of the much smaller total number of known antigens for Bp.

Prioritizing highly conserved epitopes is a critical consideration for vaccine development, as highly conserved epitopes can induce broadly protective immunity, and reduce the risk that emergence of pathogen variants will render the vaccine ineffective (31). ~90% of the top scoring peptides were found to be present in 90% or more of the fully sequenced pathogenic *F. tularensis* and *B. pseudomallei* strains (see Supplementary tables S1 and S2). In addition, we can target peptides that show even deeper evolutionary conservation based on their Average Amino Acid Conservation Score (AAACS), reflecting parts of the protein that may be important for its function (22) (see Supplementary tables S1 and S2).

Note that while some of the proteins in Tables 1 and 2 have homologs in human and mouse (e.g. mitochondrial DnaK), the peptides recovered here are unique to the bacterial versions. Peptides that are only one or two amino acids different from human or mouse versions are likely less suitable as vaccine candidates and are marked with a subscript 1 or 2 respectively in the tables. For vaccine design, we may also want to prioritize peptides which do not tend to occur in healthy human microbiome.

Figure 3 shows the 46 Ft DnaK peptides that were detected in at least two Experiment samples, including the 8 that are in our list of 44 enriched Ft peptides (Table 1).

Lu et al. (32) used hydrogen/deuterium exchange–mass spectrometry (DXMS) to experimentally identify one discontinuous and four linear B-cell epitopes for a selection of mouse monoclonal antibodies against GroL. Figure 4 shows the 32 Ft GroL peptides that were detected in at least two Experiment samples in our study, including the 4 that are in our list of 44 enriched Ft peptides (Table 1). Note that one of these 4 peptides (DNTTIIDGAGEK) overlaps with a linear epitope (NTTIIDGAGEKEAIAKRINVIK) and a discontinuous epitope (SEDLSMKLEETNM—NTTIIDGAGEKEAIA), while a second enriched peptide (EGVITVEEGK) is directly adjacent to another of the linear epitopes (FEDEL). According to the Immune Epitope Database (IEDB) (33), these are the only experimentally validated B-cell epitopes for Ft. IEDB also lists four *B.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

pseudomallei antigens that have been assayed for B-cell epitopes, none of which overlap with the proteins in Table 1.

DISCUSSION

We have developed a widely applicable shotgun immunoproteomic method that enables efficient identification of B cell epitopes in the proteomes of pathogens. The results of this study have revealed a significant enrichment of peptides derived from previously identified antigens and vaccine candidates, validating the method's efficacy. This method was designed to identify linear epitopes efficiently without the need of genetic manipulation or other experimental techniques that can be costly and labor intensive. Attenuated strains made the optimization of this proof-of-concept study more efficient; however, the availability of an attenuated strain for the target organism does not represent a limitation, as our strategy could be applied to fully virulent strains of pathogens as well. Although the present study was performed using a mouse model, the workflow could be easily adapted to detecting targets relevant to the human immune system, using convalescent sera from patients.

Utilizing peptide antigens for vaccine development has several advantages over typical vaccine development efforts. First, peptide vaccines represent a safer alternative to traditional vaccines, because the vaccine formulation is defined and contents are fully synthetic. Second, peptide vaccines have the potential to decrease the cost and production timeline, due to ease of synthesis and recent advances in improved peptide stability (3,34,35). In addition, once antigenic peptides are identified, screening for efficacy could represent a lesser challenge due to the possibility of multiplexing peptides during *in vivo* trials, rather than a one-at-a-time approach.

Among Ft proteins, the present screen identified multiple peptides for two well-characterized antigens, 60kDa chaperonin GroL (Q5NEE1) and chaperone protein DnaK (Q5NFG7). Both chaperonins have been previously implicated in virulence of Francisella (36–38), and are known to induce antibody production in mice and humans (18,39,40). These chaperonin proteins are important for facilitating folding of nascent proteins as well as post-translational modifications. They are also known as heat-shock proteins, as they protect cellular proteins from environmental stresses such as high temperature and low pH (40,41). Although their cellular localization is predicted to be cytoplasmic, they reportedly also associate with membrane proteins and are released into host cells during infection (40,42–44) perhaps contributing to their ability to stimulate various immune functions, including innate immunity, humoral immunity and cell-mediated immunity (36,40,45–48). Heat-shock proteins are good candidates for subunit vaccine design due to their ability to stimulate various immune responses without the need of adjuvant; in fact, both GroL and DnaK have been exploited for vaccine development efforts targeting *Francisella* and other pathogens (32,40,49,50).

Highly virulent Type A *Francisella* strains such as SCHU S4 can bind host plasminogen to the bacterial cell surface where it can be converted to plasmin, a serine protease that degrades opsonizing antibodies, inhibiting antibody-mediated uptake by macrophages (51,52). Among the 25 Ft proteins listed in Table 1, we find at least 3 that are known to be involved in plasminogen binding in *Francisella* or other pathogens, including conserved hypothetical lipoprotein LpnA (Q5NGE4) (52), fructose-1,6-bisphosphate aldolase (Q5NF78) (53), and elongation factor Tu (Q5NID9) (54). These proteins could make for particularly attractive vaccine targets, because if we can interfere with their function before the pathogen has activated its plasmin-mediated antibody evasion, that would make it more susceptible to other antibodies as well.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Among the antigenic peptides identified in the Bp proteome are those belonging to Type VI secretion system component Hcp-,1 and previously identified antigen 10kDa chaperonin GroES (55). Hcp-1 was previously found to be a major virulence determinant in *Burkholderia* and recognized by sera from infected human patients and animals (56–58). Due to this, Hcp-1 has been interrogated as a potential candidate for *Burkholderia* vaccine development (56–58). Additionally, a peptide from an ankyrin repeat-containing protein (A0A0H3HJC) came up as one of the highest scoring peptides in our study. Ankyrin repeats are typically eukaryotic protein domains involved in protein-protein interactions (59), but have been co-opted by many bacterial pathogens as type IV secreted effector proteins to mimic or manipulate various host functions (60).

Recovery of peptides derived from several supposedly cytosolic enzymes may seem puzzling. However several "housekeeping" enzymes are known to be displayed on the surface of pathogens where they play a role in virulence (61). For example, our top scoring peptides from *B. pseudomallei* include two derived from enolase (A0A0H3HLA6). While enolase is primarily thought of as a key glycolytic enzyme, it is also expressed on the surface of a wide variety of bacterial and fungal pathogens, where it interacts with host plasminogen and is associated with invasion and virulence (62). Antibodies against enolase have been detected in a large variety of infectious and autoimmune diseases (63). It is as yet unknown whether enolase plays the same role in Burkholderia, but the protein is predicted to be present both in the cytoplasm and on the cell surface, and its production was found to be upregulated upon exposure to human lung epithelial cells (64). Other housekeeping proteins in our top scoring results whose homologs in other pathogens are known to play a role in adhesion, invasion, or virulence include elongation factor Tu (Q5NID9), malic enzyme/malate dehydrogenase (A0A0H3HP28, Q5NHC8), and fructose-1,6-bisphosphate aldolase (Q5NF78) (61).

Overall, this immunoproteomic workflow has identified numerous peptides mapping to previously identified antigens and subunit vaccine targets, predicted membrane-associated proteins, as well as uncharacterized proteins. The Ft datasets revealed a significant enrichment of peptides belonging to previously identified antigenic proteins in Experiment samples relative to their respective Control samples, providing validation to this approach. Interestingly, several of these known antigens also yielded multiple top scoring peptides in our analysis. Despite the large amount of prior immunoproteomic analysis on Ft, covering ~10% of the genome, experimentally validated B-cell epitopes are available for only a single protein, and our analysis captures two out of its five known epitopes. Due to the much smaller number of previously identified antigens for Burkholderia, we were not able to tell whether the enrichment in the Bp datasets was significant. More comprehensive immunogenic profiles could be achieved with the use of alternative enzymes with different specificities, since there is a risk of ablating epitopes that contain cut sites recognized by specific enzymes such as trypsin. Alternatively, performing incomplete digestion with one enzyme, or a cocktail of enzymes with different specificities, could improve the yield and diversity of identified epitopes.

Interestingly, we find no significant correlation between the peptides experimentally identified using the method described here, and computationally predicted linear B-cell epitope scores generated by state-of-the-art tools such as Bepipred 2.0 (26) and iBCE-EL (25) (see Supplementary tables S1 and S2), nor any significant correlation between the Bepipred 2.0 and iBCE-EL scores themselves. Accurate computational prediction of B-cell epitopes still poses a major challenge (65), highlighting the value of an unbiased experimental method to screen for antibody targets, as presented here. It is also possible that the tryptic peptides evaluated by this method do not score well as B-cell epitopes by tools such as iBCE-EL, which explicitly take into account sequence features at the beginning and end of the epitope. In cases where the tryptic peptide is too short to be used directly as a vaccine

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

candidate (some are as short as 6 residues), we may be able to use these computational tools to guide us in how to extend the peptide beyond its flanking trypsin cleavage sites.

Further confirmation that the identified sequences are B cell epitopes could be achieved through additional *in vitro* and *in vivo* experimentation (e.g., testing the reactivity of immune sera with synthesized candidate epitopes *via* ELISA or immunization studies). High throughput screening of peptides for efficacy is feasible due to recent advancements in solid phase peptide synthesis (SPPS), which enables efficient and cost-effective production of peptide candidates (3). For immunization studies, pools of multiple peptides could be incorporated into vaccine delivery systems containing adjuvants and T-helper epitopes known to stimulate the induction of adaptive immune response against peptide antigens, as reviewed in Skwarczynski et al (3).

Our immunoproteomic method represents a new tool for precise mapping of linear B cell epitopes. Generation of such immunogenic profiles for pathogens could provide an ample pool of candidates for further experimental validation and efficient vaccine development. Accelerating the discovery of B cell epitopes in the proteomes of pathogens will help fuel the development of peptide-based vaccines that have the potential to provide rapid solutions to biothreat agents and emerging pathogens.

Data Availability Statement

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (66) partner repository with the dataset identifier PXD026300 and 10.6019/PXD026300.

Conflict of Interest

MF, NMC and PD are inventors on a provisional patent application for the method for rapid detection of immunogenic epitopes, filed by Lawrence Livermore National Security, LLC.

Author Contributions

PD, NMC and MF contributed to conception and design of the study. NMC performed the in vivo experiments. VL provided laboratory support. VL and MF performed in vitro experimentation. PD performed the bioinformatics analysis. BWS and SSB provided critical input. All authors contributed to manuscript revision, read, and approved the submitted version

Funding

This work was supported by Lawrence Livermore National Laboratory Directed Research and Development Program (LLNL LDRD) Labwide grant (18-LW-039) to MF, and by the LDRD program at Sandia National Laboratories, a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. Work at LLNL was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL IM release number LLNL-JRNL-822446.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Acknowledgements

We thank Dr. Wayne Conlan (National Research Council Canada) for providing *Francisella tularensis* SCHU S4ΔclpB strain. Bp 82 reagent was obtained through BEI Resources, NIAID, NIH: *Burkholderia pseudomallei*, Strain Bp82 (Δ*purM*), NR-51280. Our thanks go to Michael Ford and MS Bioworks team for help with sample preparation troubleshooting and specialized mass spectrometry analyses. We also thank past and present members of our laboratories - Drs. Sahar El-Etr, José Peña, Amy Rasley, and Emilio Garcia - for useful discussions and critical input.

REFERENCES

1. Li W, Joshi MD, Singhania S, Ramsey KH, Murthy AK. Peptide Vaccine: Progress and Challenges. Vaccines Basel. 2014 Jul 2;2(3):515–36.

2. Malonis RJ, Lai JR, Vergnolle O. Peptide-Based Vaccines: Current Progress and Future Challenges. Chem Rev. 2020 Mar 25;120(6):3210–29.

3. Skwarczynski M, Toth I. Peptide-based synthetic vaccines. Chem Sci. 2016 Feb 1;7(2):842–54.

4. Dudek NL, Perlmutter P, Aguilar MI, Croft NP, Purcell AW. Epitope Discovery and Their Use in Peptide Based Vaccines. Curr Pharm Des. 2010;16(28):3149–57.

5. Wang AP, Li N, Zhou JM, Chen YM, Jiang M, Qi YH, et al. Mapping the B cell epitopes within the major capsid protein L1 of human papillomavirus type 16. Int J Biol Macromol. 2018 Oct 15;118:1354–61.

6. Zhao DM, Han KK, Huang XM, Zhang LJ, Wang HL, Liu N, et al. Screening and identification of B-cell epitopes within envelope protein of tembusu virus. Virol J. 2018 Sep 17;15.

7. Bi Y, Jin Z, Wang Y, Mou S, Wang W, Wei Q, et al. Identification of Two Distinct Linear B Cell Epitopes of the Matrix Protein of the Newcastle Disease Virus Vaccine Strain LaSota. Viral Immunol. 2019 Jun;32(5):221–9.

8. Jaenisch T, Heiss K, Fischer N, Geiger C, Bischoff FR, Moldenhauer G, et al. High-density Peptide Arrays Help to Identify Linear Immunogenic B-cell Epitopes in Individuals Naturally Exposed to Malaria Infection. Mol Cell Proteomics. 2019 Apr;18(4):642–56.

9. Dienst FT. Tularemia: a perusal of three hundred thirty-nine cases. J State Med Soc. 1963 Apr;115:114–27.

10. Fulton KM, Zhao X, Petit MD, Kilmury SLN, Wolfraim LA, House RV, et al. Immunoproteomic analysis of the human antibody response to natural tularemia infection with Type A or Type B strains or LVS vaccination. Int J Med Microbiol IJMM. 2011 Nov;301(7):591–601.

11. Gibney KB, Cheng AC. Reducing the melioidosis burden: public health, chronic disease prevention, or improved case management? Lancet Infect Dis. 2019 Aug;19(8):800–2.

12. Mara-Koosham G, Hutt JA, Lyons CR, Wu TH. Antibodies Contribute to Effective Vaccination against Respiratory Infection by Type A Francisella tularensis Strains. Infect Immun. 2011 Apr 1;79(4):1770–8.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

13. Ray HJ, Cong Y, Murthy AK, Selby DM, Klose KE, Barker JR, et al. Oral Live Vaccine Strain-Induced Protective Immunity against Pulmonary Francisella tularensis Challenge Is Mediated by CD4+ T Cells and Antibodies, Including Immunoglobulin A. Clin Vaccine Immunol CVI. 2009 Apr;16(4):444–52.

14. Rhinehart-Jones TR, Fortier AH, Elkins KL. Transfer of immunity against lethal murine Francisella infection by specific antibody depends on host gamma interferon and T cells. Infect Immun. 1994 Aug;62(8):3129–37.

15. Hogan RJ, Lafontaine ER. Antibodies Are Major Drivers of Protection against Lethal Aerosol Infection with Highly Pathogenic Burkholderia spp. mSphere [Internet]. 2019 Jan 2 [cited 2021 Mar 15];4(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6315082/

16. Jones SM, Ellis JF, Russell P, Griffin KF, Oyston PCF. Passive protection against Burkholderia pseudomallei infection in mice by monoclonal antibodies against capsular polysaccharide, lipopolysaccharide or proteins. J Med Microbiol. 2002;51(12):1055–62.

17. Khakhum N, Bharaj P, Myers JN, Tapia D, Kilgore PB, Ross BN, et al. Burkholderia pseudomallei ΔtonB Δhcp1 Live Attenuated Vaccine Strain Elicits Full Protective Immunity against Aerosolized Melioidosis Infection. mSphere [Internet]. 2019 Jan 2 [cited 2021 Mar 15];4(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6315081/

18. Kilmury SLN, Twine SM. The francisella tularensis proteome and its recognition by antibodies. Front Microbiol. 2010;1:143.

19. Nakajima R, Escudero R, Molina DM, Rodríguez-Vargas M, Randall A, Jasinskas A, et al. Towards Development of Improved Serodiagnostics for Tularemia by Use of Francisella tularensis Proteome Microarrays. J Clin Microbiol. 2016 Jul;54(7):1755–65.

20. Felgner PL, Kayala MA, Vigil A, Burk C, Nakajima-Sasaki R, Pablo J, et al. A Burkholderia pseudomallei protein microarray reveals serodiagnostic and cross-reactive antigens. Proc Natl Acad Sci U S A. 2009 Aug 11;106(32):13499–504.

21. Yi J, Simpanya MF, Settles EW, Shannon AB, Hernandez K, Pristo L, et al. Caprine humoral response to Burkholderia pseudomallei antigens during acute melioidosis from aerosol exposure. PLoS Negl Trop Dis. 2019 Feb;13(2):e0006851.

22. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 2016 Jul 8;44(W1):W344-350.

23. Ren J, Ellis J, Li J. Influenza A HA's conserved epitopes and broadly neutralizing antibodies: a prediction method. J Bioinform Comput Biol. 2014 Oct;12(5):1450023.

24. Fiuza TS, Lima JPMS, de Souza GA. EpitoCore: Mining Conserved Epitope Vaccine Candidates in the Core Proteome of Multiple Bacteria Strains. Front Immunol. 2020;11:816.

25. Manavalan B, Govindaraj RG, Shin TH, Kim MO, Lee G. iBCE-EL: A New Ensemble Learning Framework for Improved Linear B-Cell Epitope Prediction. Front Immunol. 2018;9:1695.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

494 26. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell
495 epitope prediction using conformational epitopes. Nucleic Acids Res. 2017 Jul 3;45(W1):W24–9.

496 27. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts
497 of method development and improved benchmarking. PLoS Comput Biol. 2012;8(12):e1002829.

498 28. Conlan JW, Shen H, Golovliov I, Zingmark C, Oyston PC, Chen W, et al. Differential ability of
499 novel attenuated targeted deletion mutants of Francisella tularensis subspecies tularensis strain
500 SCHU S4 to protect mice against aerosol challenge with virulent bacteria: effects of host
501 background and route of immunization. Vaccine. 2010 Feb 17;28(7):1824–31.

502 29. Propst KL, Mima T, Choi KH, Dow SW, Schweizer HP. A Burkholderia pseudomallei
503 deltapurM mutant is avirulent in immunocompetent and immunodeficient animals: candidate
504 strain for exclusion from select-agent lists. Infect Immun. 2010 Jul;78(7):3136–43.

505 30. Twine S, Shen H, Harris G, Chen W, Sjostedt A, Ryden P, et al. BALB/c mice, but not C57BL/6
506 mice immunized with a ΔclpB mutant of Francisella tularensis subspecies tularensis are protected
507 against respiratory challenge with wild-type bacteria: association of protection with post-
508 vaccination and post-challenge immune responses. Vaccine. 2012 May 21;30(24):3634–45.

509 31. Eickhoff CS, Terry FE, Peng L, Meza KA, Sakala IG, Van Aartsen D, et al. Highly Conserved
510 Influenza T cell epitopes Induce Broadly Protective Immunity. Vaccine. 2019 Aug
511 23;37(36):5371–81.

512 32. Lu Z, Rynkiewicz MJ, Madico G, Li S, Yang CY, Perkins HM, et al. B-cell epitopes in GroEL of
513 Francisella tularensis. PLoS One. 2014;9(6):e99847.

514 33. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The Immune Epitope
515 Database (IEDB): 2018 update. Nucleic Acids Res. 2019 Jan 8;47(D1):D339–43.

516 34. Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in
517 vaccine design. Nat Rev Drug Discov. 2007 May;6(5):404–14.

518 35. Sun T, Han H, Hudalla GA, Wen Y, Pompano RR, Collier JH. Thermal stability of self-
519 assembled peptide vaccine materials. Acta Biomater. 2016 Jan 15;30:62–71.

520 36. Noah CE, Malik M, Bublitz DC, Camenares D, Sellati TJ, Benach JL, et al. GroEL and
521 lipopolysaccharide from Francisella tularensis live vaccine strain synergistically activate human
522 macrophages. Infect Immun. 2010 Apr;78(4):1797–806.

523 37. Pechous RD, McCarthy TR, Zahrt TC. Working toward the future: insights into Francisella
524 tularensis pathogenesis and vaccine development. Microbiol Mol Biol Rev. 2009 Dec;73(4):684–
525 711.

526 38. Weiss DS, Brotcke A, Henry T, Margolis JJ, Chan K, Monack DM. In vivo negative selection
527 screen identifies genes required for Francisella virulence. Proc Natl Acad Sci U A. 2007 Apr
528 3;104(14):6037–42.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

39. Havlasova J, Hernychova L, Brychta M, Hubalek M, Lenco J, Larsson P, et al. Proteomic analysis of anti-Francisella tularensis LVS antibody response in murine model of tularemia. Proteomics. 2005 May;5(8):2090–103.

40. Huntley JF, Conley PG, Hagman KE, Norgard MV. Characterization of Francisella tularensis outer membrane proteins. J Bacteriol. 2007 Jan;189(2):561–74.

41. Ericsson M, Tarnvik A, Kuoppa K, Sandstrom G, Sjostedt A. Increased synthesis of DnaK, GroEL, and GroES homologs by Francisella tularensis LVS in response to heat and hydrogen peroxide. Infect Immun. 1994 Jan;62(1):178–83.

42. Henderson B, Allan E, Coates AR. Stress wars: the direct role of host and bacterial molecular chaperones in bacterial infection. Infect Immun. 2006 Jul;74(7):3693–706.

43. Hickey TB, Thorson LM, Speert DP, Daffe M, Stokes RW. Mycobacterium tuberculosis Cpn60.2 and DnaK are located on the bacterial surface, where Cpn60.2 facilitates efficient bacterial association with macrophages. Infect Immun. 2009 Aug;77(8):3389–401.

44. Lee BY, Horwitz MA, Clemens DL. Identification, recombinant expression, immunolocalization in macrophages, and T-cell responsiveness of the major extracellular proteins of Francisella tularensis. Infect Immun. 2006 Jul;74(7):4002–13.

45. Ashtekar AR, Zhang P, Katz J, Deivanayagam CC, Rallabhandi P, Vogel SN, et al. TLR4-mediated activation of dendritic cells by the heat shock protein DnaK from Francisella tularensis. J Leukoc Biol. 2008 Dec;84(6):1434–46.

46. Kol A, Bourcier T, Lichtman AH, Libby P. Chlamydial and human heat shock protein 60s activate human vascular endothelium, smooth muscle cells, and macrophages. J Clin Invest. 1999 Feb;103(4):571–7.

47. Wallin RP, Lundqvist A, More SH, von Bonin A, Kiessling R, Ljunggren HG. Heat-shock proteins as activators of the innate immune system. Trends Immunol. 2002 Mar;23(3):130–5.

48. Valentino MD, Maben ZJ, Hensley LL, Woolard MD, Kawula TH, Frelinger JA, et al. Identification of T-cell epitopes in Francisella tularensis using an ordered protein array of serological targets. Immunology. 2011 Mar;132(3):348–60.

49. Ashtekar AR, Katz J, Xu Q, Michalek SM. A mucosal subunit vaccine protects against lethal respiratory infection with Francisella tularensis LVS. PLoS One. 2012;7(11):e50460.

50. Khan MN, Shukla D, Bansal A, Mustoori S, Ilavazhagan G. Immunogenicity and protective efficacy of GroEL (hsp60) of Streptococcus pneumoniae against lethal infection in mice. FEMS Immunol Med Microbiol. 2009 Jun;56(1):56–62.

51. Crane DD, Warner SL, Bosio CM. A novel role for plasmin mediated degradation of opsonizing antibody in the evasion of host immunity by virulent, but not attenuated, Francisella tularensis. J Immunol Baltim Md 1950. 2009 Oct 1;183(7):4593–600.

52. Clinton SR, Bina JE, Hatch TP, Whitt MA, Miller MA. Binding and activation of host plasminogen on the surface of Francisella tularensis. BMC Microbiol. 2010 Mar 12;10:76.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

53. Shams F, Oldfield NJ, Wooldridge KG, Turner DPJ. Fructose-1,6-bisphosphate aldolase (FBA)-a conserved glycolytic enzyme with virulence functions in bacteria: “ill met by moonlight.” Biochem Soc Trans. 2014 Dec;42(6):1792–5.

54. Kunert A, Losse J, Gruszin C, Hühn M, Kaendler K, Mikkat S, et al. Immune evasion of the human pathogen Pseudomonas aeruginosa: elongation factor Tuf is a factor H and plasminogen binding protein. J Immunol Baltim Md 1950. 2007 Sep 1;179(5):2979–88.

55. Varga JJ, Vigil A, DeShazer D, Waag DM, Felgner P, Goldberg JB. Distinct human antibody response to the biological warfare agent Burkholderia mallei. Virulence. 2012 Oct 1;3(6):510–4.

56. Burtnick MN, Brett PJ, Harding SV, Ngugi SA, Ribot WJ, Chantratita N, et al. The cluster 1 type VI secretion system is a major virulence determinant in Burkholderia pseudomallei. Infect Immun. 2011 Apr;79(4):1512–25.

57. Schell MA, Ulrich RL, Ribot WJ, Brueggemann EE, Hines HB, Chen D, et al. Type VI secretion is a major virulence determinant in Burkholderia mallei. Mol Microbiol. 2007 Jun;64(6):1466–85.

58. Whitlock GC, Deeraksa A, Qazi O, Judy BM, Taylor K, Propst KL, et al. Protective response to subunit vaccination against intranasal Burkholderia mallei and B. pseudomallei challenge. Procedia Vaccinol. 2010;2(1).

59. Li J, Mahajan A, Tsai M-D. Ankyrin repeat: a unique motif mediating protein-protein interactions. Biochemistry. 2006 Dec 26;45(51):15168–78.

60. Pan X, Lührmann A, Satoh A, Laskowski-Arce MA, Roy CR. Ankyrin Repeat Proteins Comprise a Diverse Family of Bacterial Type IV Effectors. Science. 2008 Jun 20;320(5883):1651–4.

61. Pancholi V, Chhatwal GS. Housekeeping enzymes as virulence factors for pathogens. Int J Med Microbiol IJMM. 2003 Dec;293(6):391–401.

62. Lottenberg R, Minning-Wenz D, Boyle MD. Capturing host plasmin(ogen): a common mechanism for invasive pathogens? Trends Microbiol. 1994 Jan;2(1):20–4.

63. Terrier B, Degand N, Guilpain P, Servettaz A, Guillevin L, Mouthon L. Alpha-enolase: a target of antibodies in infectious and autoimmune diseases. Autoimmun Rev. 2007 Jan;6(3):176–82.

64. Al-Maleki AR, Mariappan V, Vellasamy KM, Tay ST, Vadivelu J. Altered Proteome of Burkholderia pseudomallei Colony Variants Induced by Exposure to Human Lung Epithelial Cells. PloS One. 2015;10(5):e0127398.

65. Sun P, Guo S, Sun J, Tan L, Lu C, Ma Z. Advances in In-silico B-cell Epitope Prediction. Curr Top Med Chem. 2019;19(2):105–15.

66. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. 2019 Jan 8;47(Database issue):D442–50.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

FIGURES

Figure 1: Immunoproteome screening workflow. Schematic overview of high throughput approach for identification of seroreactive peptides in the proteomes of pathogens.

Figure 2: **A**. Representative course of mouse infection to obtain immune sera. Mice were infected with a sublethal dose of Bp and their weight monitored. The degree of weight loss correlates to the amount of antibodies detected in the sera. **B**. Representative Western blot of immune sera vs. non-immune sera. Bp protein lysates were analyzed by Western blotting using sera from infected and uninfected mice (Mouse 1–3). Antibodies from sera with the strongest signal are purified in this study and used to screen for immunogenic peptides. **C**. Representative ELISA results obtained from mice infected with Bp and Ft (red) in comparison with uninfected mice (PBS-treated mice, blue). Sera of some mice infected with Ft did not yield positive results because Ft infection led to lethal outcome and mice had to be euthanized during the course of immunization. Graphs represent two technical replicates for sera collected from each mouse.

Figure 3: Scoring for the 46 F. tularensis DnaK peptides detected in at least two Experiment samples. Each horizontal line segment indicates the position of a peptide along the length of the 642aa DnaK protein, and its vertical position within each figure panel indicates its score for the metric indicated. The default score threshold suggested for each tool is shown with a horizontal line. **A**. Peptide enrichment score based on our proteomics results. An enrichment score of 8 indicates that the peptide was detected in greater abundance in all 8 Experiment samples relative to their respective Control samples. The threshold for inclusion in Table 1 was an enrichment score of ≥6**. B.** B-cell epitope prediction score generated using iBCE-EL. Peptides scoring >0.35 were predicted to be likely B-cell epitopes. **C.** B-cell epitope prediction score generated using Bepipred 2.0. The per-amino acid scores are indicated by the line graph. Regions of the protein scoring >0.5 were predicted to likely contain B-cell epitopes. **D.** B-cell epitope prediction score generated using Discotope 2.0. The per-amino acid scores are indicated by the line graph. Regions of the protein scoring >-0.37 were predicted to likely contain B-cell epitopes. **E.** Average Amino Acid Conservation Score (AAACS) based on Consurf analysis. Lower scores indicate greater degrees of evolutionary conservation. **F.** Number of fully sequenced F. tularensis subsp. tularensis genomes (17 analyzed) in which each peptide occurs.

Figure 4: The 32 F. tularensis GroL peptides detected in at least two Experiment samples. Horizontal line segments indicate the position of each peptide along the length of the 544aa GroL protein sequence. **A.** Peptide enrichment score based on our proteomics results, with a score of 8 indicating that the peptide was found in greater abundance in all 8 Experiment samples relative to their respective Control samples. The threshold for inclusion in Table 1 was a score of ≥6 or better. **B.** B-cell epitopes identified by DXMS by Lu et al. (32).

TABLES

Table 1: List of top scoring immunoreactive peptides identified for Francisella tularensis. The columns under “scores” indicate whether the peptide was over or underrepresented in each of the 8 experimental samples compared to its control sample. Green: experiment>control. Red: experiment<control. White: peptide undetected in both experiment and control. Dark colors indicate >2-fold difference in relative abundance. Proteins with multiple top scoring peptides are highlighted in bold.

[a]: known antigen. [i]: inner membrane. [p]: periplasmic. [o]: outer membrane. [e]: extracellular.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

[1]: peptide sequence is only a single amino acid away from a human or mouse peptide. [2]: peptide is only two amino acids away from a human or mouse peptide.

Protein name	Accession	Peptide	Scores
Aminotransferase AspC1	Q5NGG1	LPIDDAEK[2]	
Glutamate dehydrogenase Gdh	Q5NHR7[a]	FHPSVYSGIIK	
Pyruvate dehydrogenase acetyltransferase AceF	Q5NEX3[a]	VSQGSLILK[2]	
60 kDa chaperonin GroL	Q5NEE1[a]	DRVDDALHATR[2]	
Chaperone protein DnaK	Q5NFG7[a]	NTADNLIHSSR	
Chaperone protein DnaK	Q5NFG7[a]	SSSGLSEEDIEK	
60 kDa chaperonin GroL	Q5NEE1[a]	DNTTIIDGAGEK	
60 kDa chaperonin GroL	Q5NEE1[a]	EGVITVEEGK	
Catalase-peroxidase KatG	Q5NGV7[a]	AVAQVYAENGNEQK	
Malate dehydrogenase Mdh	Q5NHC8[a]	FSGVPDNK[1]	
Outer membrane protein 26 Omp26	Q5NES2[o]	EIPADQLGTIK	
Succinate dehydrogenase flavoprotein SdhA	Q5NIJ3[a,i]	ITILATGGAGR	
ATP synthase subunit alpha AtpA	Q5NIK5[a]	GEVATDLTSPIEK	
Elongation factor Ts Tsf	Q5NHX9[a]	ESGKPAEIIEK	
Elongation factor Ts Tsf	Q5NHX9[a]	TVEAETLGAYIHGSK	
Chaperone protein DnaK	Q5NFG7[a]	IAGLEVK[1]	
Cell division protein FtsZ	Q5NI93[a]	KETEVVTGASNAPK	
Trigger factor Tig	Q5NH48	GGVDTFENEIK	
ATP synthase subunit alpha AtpA	Q5NIK5[a]	SVDQALQTGIK	
Catalase-peroxidase KatG	Q5NGV7[a]	NDNLSPQSVDLSPLR	
Isocitrate dehydrogenase [NADP] Idh	Q5NET6[a]	VADIELETK[2]	
Fructose-1,6-bisphosphate aldolase FbaB	Q5NF78[a]	KINIDTDLR	
Glutamate dehydrogenase Gdh	Q5NHR7[a]	GFVHDPEGITTDEK	
Succinate--CoA ligase [ADP-forming] beta SucC	Q5NHF3[a]	PANFLDVGGGATK[1]	
Chaperone protein DnaK	Q5NFG7[a]	KVPYAVIK[2]	
Malonyl CoA-ACP transacylase	Q5NF69[a]	EPTTAVVQNFDAK	
Peroxiredoxin	Q5NHA9[a]	KVPNVTFK[2]	
Chaperone protein DnaK	Q5NFG7[a]	IINEPTAAALAYGVDSK	
Conserved hypothetical lipoprotein LpnA	Q5NGE4[a,o]	ATVYTTYNNNPQGSVR	
Elongation factor Tu Tuf	Q5NID9[a]	TTVTGVEMFR	
Succinate--CoA ligase [ADP-forming] beta SucC	Q5NHF3[a]	EVAESLIGK[1]	
30S ribosomal protein S1 RpsA	Q5NI98[a]	KIELWDR[2]	
Elongation factor Tu Tuf	Q5NID9[a]	HYAHVDCPGHADYVK[1]	
Transcription elongation factor GreA	Q5NFC6[a]	IVGEDEADIDNQK	
60 kDa chaperonin GroL	Q5NEE1[a]	SFGTPTITK[2]	
Aconitate hydratase AcnA	Q5NII1[a]	GIPLVILAGK[1]	
Chaperone protein DnaK	Q5NFG7[a]	AYAEQAQAAVAQGGAK	
Chaperone protein DnaK	Q5NFG7[a]	FHDLVTAR[2]	
Outer membrane protein 26 Omp26	Q5NES2	DGSVGWVK[1]	
3-oxoacyl-ACP reductase FabG	Q5NF68	VALVTGASR[1]	
Chaperone protein DnaK	Q5NFG7[a]	ALEDAGLSK[2]	
Enoyl-ACP reductase [NADH] FabI	Q5NGQ3[i]	TLAASGISNFK	
Aconitate hydratase AcnA	Q5NII1[a]	TAHTTTFEALAR	
Elongation factor Ts Tsf	Q5NHX9[a]	LDVGEGIEK[1]	

Table 2: List of top scoring immunoreactive peptides identified for Burkholderia pseudomallei. The columns under "scores" indicate whether the peptide was over or underrepresented in each of the 9 experimental samples compared to its control sample. Green: experiment>control. Red: experiment<control. White: peptide undetected in both experiment and control. Dark colors indicate

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

>2-fold difference in relative abundance. Proteins with multiple top scoring peptides are highlighted in bold.

[a]: known antigen. [i]: inner membrane. [p]: periplasmic. [o]: outer membrane. [e]: extracellular.

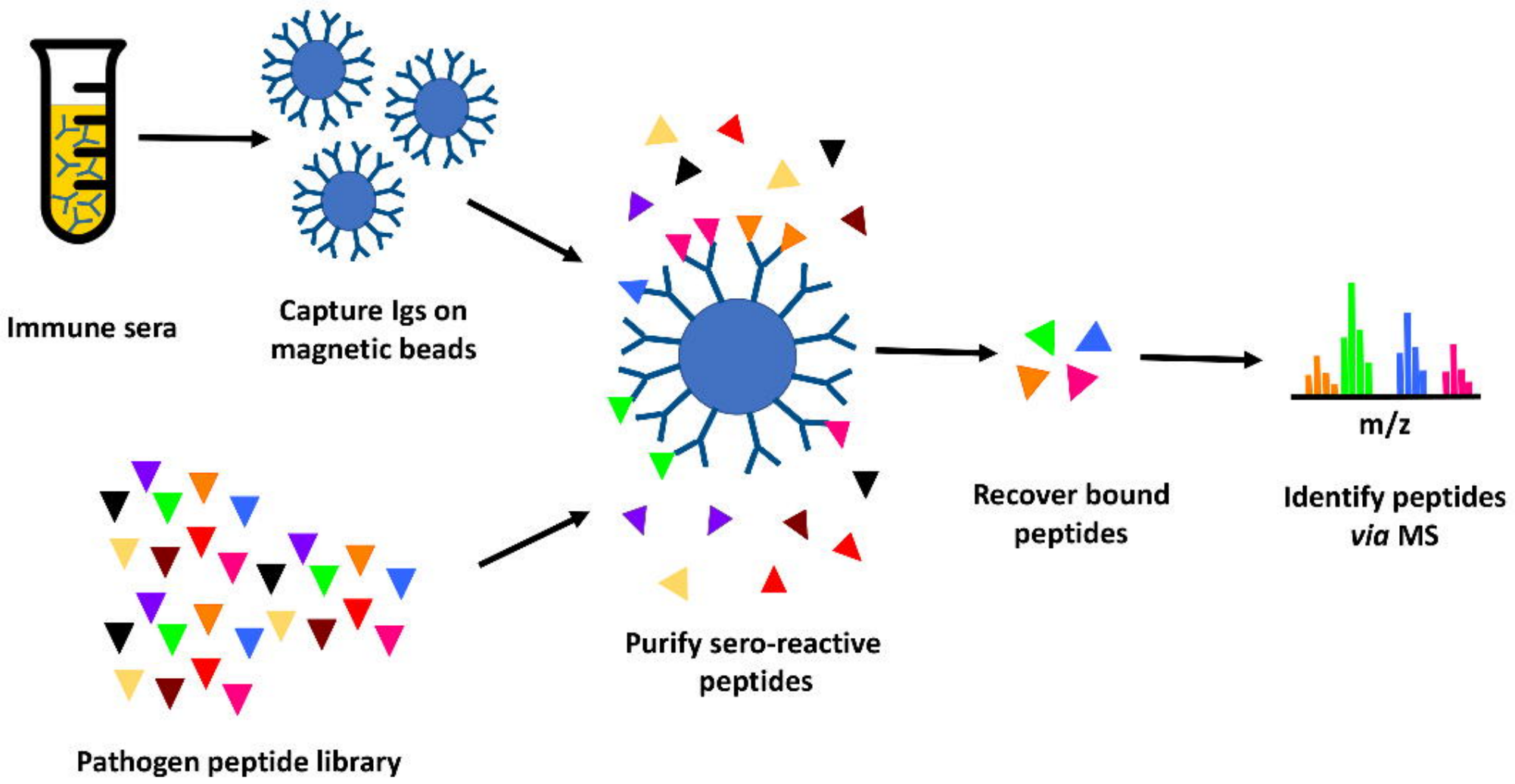
[1]: peptide sequence is only a single amino acid away from a human or mouse peptide. [2]: peptide is only two amino acids away from a human or mouse peptide.

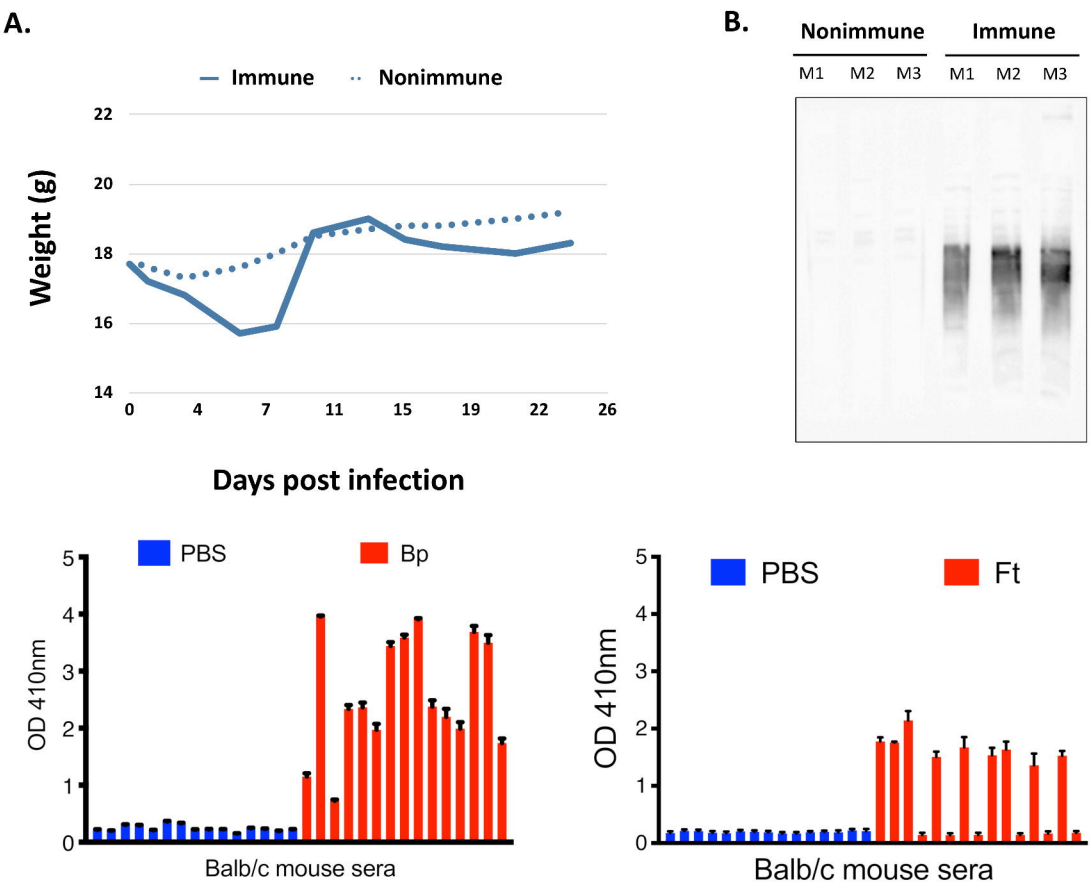
Protein name	Accession	Peptide	Scores
Aspartate--tRNA(Asp/Asn) ligase AspS	A0A0H3HT48	TGAQDGDIIFFAADR	
Adenylosuccinate synthetase PurA	A0A0H3HJJ2	QDQIGITLANVGK	
Dihydrolipoyl dehydrogenase OdhL	A0A0H3HQK7	FPFSINGR[2]	
Ankyrin repeat-containing protein	A0A0H3HJC7	IGDAPAPNAQK	
Phosphoribosylformylglycinamidine synthase PurL	A0A0H3HPH9	GATETFVVLPR	
DNA-directed RNA polymerase subunit beta RpoB	A0A0H3HT47	STGPYSLVTQQPLGGK	
50S ribosomal protein L6 RplF	A0A0H3HQ22	GYRPPEPYK	
DNA-directed RNA polymerase subunit beta RpoC	A0A0H3HP07	ISLYATTVGR	
Enolase Eno	A0A0H3HLA6	GIANSILIK[2]	
Uncharacterized protein	A0A0H3HWA2	IDCLTNAYTAR	
DNA gyrase subunit A GyrA	A0A0H3HKL0	INVVLPVR[2]	
Aspartate-semialdehyde dehydrogenase Asd	A0A0H3HW74	VTGTLSVPVGR	
Malic enzyme	A0A0H3HP28	AALLSNSNFGSAPSASSR	
50S ribosomal protein L10 RplJ	A0A0H3HUR4	AQTVVLAEYR	
50S ribosomal protein L6 RplF	A0A0H3HQ22	AIIANAVHGVTK	
Glutamine synthetase GlnA	A0A0H3HL61	ALNAITNPTTNSYK	
Nucleoside diphosphate kinase Ndk	A0A0H3HJK0[e]	NVIGQIYSR[2]	
Antioxidant protein LsfA	A0A0H3HGZ9	LIITYPASTGR	
UDP-glucose 4-epimerase	A0A0H3HFV2	GYSVLEVVR	
Enolase Eno	A0A0H3HLA6	SAIVDIIGR[2]	
Acetyl-CoA acetyltransferase	A0A0H3HTT4	LPLSVGCTTINK	
KHG/KDPG aldolase Eda	A0A0H3HGE0	FGVSPGLTR[2]	
10 kDa chaperonin GroES	A0A0H3HH83[a]	TASGIVIPDAAAEKPDQGEVLAIGPGKR	
Saccharopine dehydrogenase	A0A0H3HIF5	HGQLVQDVFTR	
Citrate synthase GltA	A0A0H3HYU5	YSIGQPFVYPR	
Aspartate--tRNA(Asp/Asn) ligase AspS	A0A0H3HT48	YVAAHHPFTSPK	
Gamma-aminobutyraldehyde dehydrogenase	A0A0H3HQU5	SVLAAAAGNLK[2]	
Peptide chain release factor 2 PrfB	A0A0H3HL96	SYVLDQSR[2]	
Polyketide non-ribosomal peptide synthase	A0A0H3HWL5[i]	AWFIPLSAR[2]	
Transcription termination/antitermination NusG	A0A0H3HPU8	VTGFVGGAR[2]	
Beta sliding clamp DnaN	A0A0H3HFM1	FTFGQVELVSK	
Malate synthase AceB	A0A0H3HIT5	IATLIVRPR[2]	
PTS system, EIIA component	A0A0H3HRL4	ISGHHLEVTPAIR	
Phosphoenolpyruvate synthase PpsA	A0A0H3HJ13	IFILQARPETVK	
Thiol:disulfide interchange protein DsbA	A0A0H3HTS6[p]	NYNIDGVPTIVVQGK	
RND family efflux transporter MFP subunit BpeA	A0A0H3HQZ3[i]	AQANLATQNALVAR	
Inosine-5'-monophosphate dehydrogenase GuaB	A0A0H3HJ23	LVGIVTNR[1]	
Periplasmic maltose-binding protein MalE	A0A0H3HG39[p]	VNWLYINK	
Putative extracellular ligand binding protein	A0A0H3HWC6[p]	VVATDAQQGPALADYAK	
Acid phosphatase AcpA	A0A0H3HV11[e]	NIVVIYAENR	
NADH-quinone oxidoreductase subunit F NuoF	A0A0H3HPW5	EGTGWLYR[2]	
Type VI secretion system Hcp-1	A0A0H3HE88[e]	IGGNQGGNTQGAWSLTK	
50S ribosomal protein L23 RplW	A0A0H3HT35	AAVELLFK[2]	
50S ribosomal protein L6 RplF	A0A0H3HQ22	LTLVGVGYR	
50S ribosomal protein L17 RplQ	A0A0H3HPQ2	LFDVLGPR[2]	
Aconitate hydratase	A0A0H3HVV9	IVLESVLR[1]	

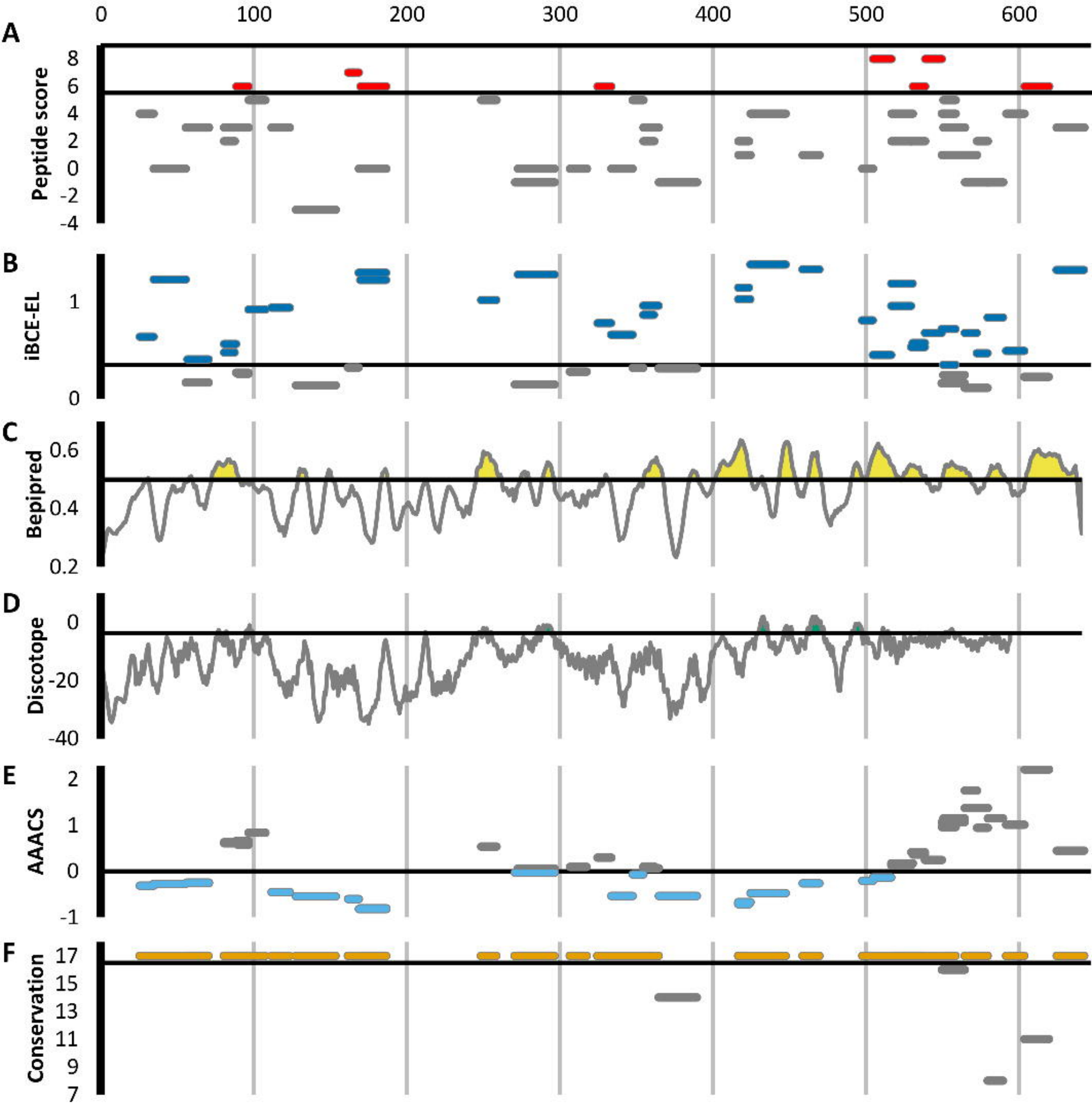
bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447172; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

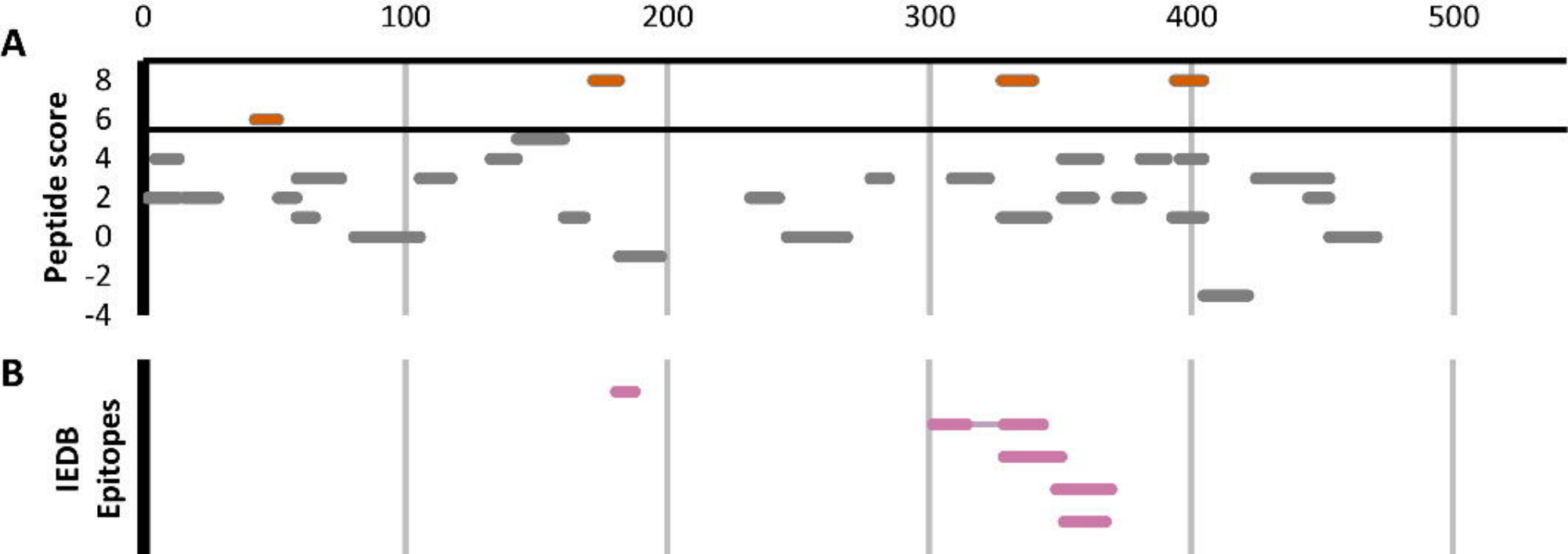
658

659


Immune sera
Capture Igs on magnetic beads
Pathogen peptide library
Purify sero-reactive peptides
Recover bound peptides
m/z
Identify peptides *via* MS


A.
Immune
Nonimmune
Weight (g)
22
20
18
16
14
0
4
7
11
15
19
22
26
Days post infection
B.
Nonimmune
Immune
M1
M2
M3
M1
M2
M3
PBS
Bp
OD 410nm
5
4
3
2
1
0
Balb/c mouse sera
PBS
Ft
OD 410nm
5
4
3
2
1
0
Balb/c mouse sera


0
100
200
300
400
500
600
A
Peptide score
8
6
4
2
0
-2
-4
B
iBCE-EL
1
0
C
Bepipred
0.6
0.4
0.2
D
Discotope
0
-20
-40
E
AAACS
2
1
0
-1
F
Conservation
17
15
13
11
9
7


A
Peptide score
8
6
4
2
0
-2
-4
0
100
200
300
400
500
B
IEDB
Epitopes