A random mutagenesis screen enriched for missense mutations in bacterial effector proteins

Urbanus, Malene L; Zheng, Thomas M; Khusnutdinova, Anna N; Banh, Doreen; O'Connor Mount, Harley; Gupta, Alind; Stogios, Peter J; Savchenko, Alexei; Isberg, Ralph R; Yakunin, Alexander F; Ensminger, Alexander W

G3: Genes, Genomes, Genetics

DOI:
10.1093/g3journal/jkae158

Published: 04/09/2024

Peer reviewed version

A random mutagenesis screen enriched for missense mutations in bacterial effector proteins.

Malene L. Urbanus[1], Thomas M. Zheng[1], Anna N. Khusnutdinova[2,6], Doreen Banh[1], Harley O'Connor Mount[3], Alind Gupta[3], Peter J. Stogios[2], Alexei Savchenko[2,4], Ralph R. Isberg[5], Alexander F. Yakunin[2,6], and Alexander W. Ensminger[1,3*]

Affiliations

[1]Department of Biochemistry, University of Toronto, Toronto, ON, M5G 1M1, Canada.

[2]Department of Chemical Engineering and Applied Chemistry, University of Toronto, ON, M5S 1A4, Canada.

[3]Department of Molecular Genetics, University of Toronto, Toronto, ON, M5G 1M1, Canada.

[4]Department of Microbiology, Immunology & Infectious Diseases, Health Research Innovation Centre, University of Calgary, AB, T2N 4N1, Canada.

[5]Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, MA, 02115, USA

[6]Centre for Environmental Biotechnology, School of Natural Sciences, Bangor University, Bangor, LL57 2UW, UK.

*Corresponding author:

Email: alex.ensminger@utoronto.ca

Running head: Missense mutant enrichment in random mutagenesis

Keywords: *Saccharomyces cerevisiae*; random mutagenesis screen; missense mutation; loss-of-function mutant; bacterial effector; *Legionella pneumophila;* AlphaFold

Abstract

To remodel their hosts and escape immune defenses, many pathogens rely on large arsenals of proteins (effectors) that are delivered to the host cell using dedicated translocation machinery. Effectors hold significant insight into the biology of both the pathogens that encode them and the host pathways that they manipulate. One of the most powerful systems biology tools for studying effectors is the model organism, *Saccharomyces cerevisiae*. For many pathogens, the heterologous expression of effectors in yeast is growth inhibitory at a frequency much higher than housekeeping genes, an observation ascribed to targeting conserved eukaryotic proteins. Abrogation of yeast growth inhibition has been used to identify bacterial suppressors of effector activity, host targets, and functional residues and domains within effector proteins. We present here a yeast-based method for enriching for informative, in-frame, missense mutations in a pool of random effector mutants. We benchmark this approach against three effectors from *Legionella pneumophila*, an intracellular bacterial pathogen that injects a staggering >330 effectors into the host cell. For each protein, we show how *in silico* protein modeling (AlphaFold2) and missense-directed mutagenesis can be combined to reveal important structural features within effectors. We identify known active site residues within the metalloprotease RavK, the putative active site in SdbB, and previously unidentified functional motifs within the C-terminal domain of SdbA. We show that this domain has structural similarity with glycosyltransferases and exhibits *in vitro* activity consistent with this predicted function.

Summary:

Random mutagenesis is a common approach in structure-function studies, but not all mutations are equally informative. Nonsense mutations and frame-shift mutations frequently lead to non-

specific disruptions. In contrast, missense mutations can reveal individual residues, active-site pockets, and other amino acid motifs that are essential for protein function. We present here a methodology that leverages the power of yeast genetics to enrich for informative missense mutations and apply this to the study of effector proteins from the human pathogen *Legionella pneumophila.*

Introduction

For many bacterial pathogens, host manipulation derives from the collective activity of large numbers of translocated proteins (effectors) that are injected into the host cell using dedicated secretion machinery. A striking example is the gram-negative bacterium *Legionella pneumophila* which naturally replicates in freshwater protozoa and is the causative agent of Legionnaires' disease in humans (Fields et al. 2002). The *L. pneumophila* genome encodes the largest effector arsenal described to date (>330 effectors per isolate, or roughly 10% of the proteome) (Burstein et al. 2009; Huang et al. 2011; Zhu et al. 2011), which is injected into the host cell using the Dot/Icm type IVB secretion system (Segal et al. 1998; Vogel et al. 1998). *L. pneumophila* effectors modulate very conserved host processes, such as vesicle trafficking, post-translational modification, protein translation, autophagy, vacuolar function, and the cytoskeleton to avoid lysosomal fusion and to establish a replicative, neutral pH vacuole (Isberg et al. 2009; Escoll et al. 2013; Sherwood and Roy 2016; Qiu and Luo 2017; Mondino et al. 2020; Luo et al. 2021; Shames 2023; Yang et al. 2023). While over 50 effectors have been studied, most of the effectors remain uncharacterized (Finsel and Hilbi 2015; Mondino et al. 2020).

Determining the function of the >330 effectors and the role they play in establishing the *Legionella*-containing vacuole is complicated by extensive genetic redundancy within the

effector arsenal (O'Connor et al. 2011) and the lack of predicted conserved domains or functions for many substrates (Gomez-Valero et al. 2011; Gomez-Valero et al. 2014; Burstein et al. 2016). Only half of the predicted effectors contain conserved domains and many of these are of uncharacterized function (Burstein et al. 2016). Although the amino acid sequence of many effectors may not yield obvious clues to their function, some effectors have structural homology to characterized proteins or domains, along with conserved active site motifs or other signature motifs (Toulabi et al. 2013; Morar et al. 2015; Wong et al. 2015; Urbanus et al. 2016; Pinotsis and Waksman 2017; Kozlov et al. 2018; Lin et al. 2018; Valleau et al. 2018; Black et al. 2019; Sulpizio et al. 2019; Hsieh et al. 2021; Voth et al. 2021). Looking beyond *L. pneumophila*, over 18,000 effector genes (including orthologs, paralogs and unique genes) have been predicted across the entire *Legionella* genus (Burstein et al. 2016; Gomez-Valero et al. 2019). A wealth of novel effector activities and host biology remains to be discovered.

We set out to develop a method to efficiently identify important motifs or amino acid residues in uncharacterized *L. pneumophila* effectors by random mutagenesis and selection for loss-of-function mutations to facilitate the prediction of mechanism and function. As has been observed for the effectors of other pathogens (Lesser and Miller 2001; Valdivia 2004; Siggers and Lesser 2008), the heterologous expression of *L. pneumophila* effectors often leads to inhibition of yeast growth. While the level of inhibition varies between effectors, approximately 10% of the effectors severely inhibit yeast growth when overexpressed (Campodonico et al. 2005; Shohdy et al. 2005; de Felipe et al. 2008; Heidtman et al. 2009; Shen et al. 2009; Guo et al. 2014; Urbanus et al. 2016) such that loss-of-function by random mutagenesis can be selected for as an alleviation of the yeast growth defect. However, a random mutant pool contains many mutations that can potentially cause a loss-of-function phenotype, such as frameshift, nonsense

and missense mutations in the effector or regulatory elements such as the promoter region. While most frameshift and nonsense mutations are so disruptive as to be largely uninformative, missense loss-of-function mutations can be extremely informative by identifying specific residues and motifs essential for protein function. To enrich for full-length missense clones, we used a C-terminal in-frame fusion of the yeast *HIS3* gene to effector genes to complement a yeast strain carrying the *his3Δ1* allele (Brachmann et al. 1998) and selected for the ability to grow on medium lacking histidine, which requires the presence of a full-length fusion protein. A similar strategy (C-terminal *HIS3* fusions) was previously shown to enrich for in-frame human open reading frames amongst a randomly primed pool of cDNAs cloned into a yeast expression vector (Holz et al. 2001). Here, we benchmark this in-frame mutagenesis approach against three *L. pneumophila* effectors previously shown to inhibit yeast growth (Heidtman et al. 2009): SdbA and SdbB, whose functions remain uncharacterized and RavK, a previously described metalloprotease (Liu et al. 2017). We show our approach identifies active site residues within RavK (Liu et al. 2017), the putative active site in SdbB and previously unidentified functional motifs in the C-terminal domain of SdbA. These motifs are part of the donor and acceptor binding regions of glycosyltransferases, which we show share homology with the C-terminal domain of SdbA. Finally, we show that a C-terminal fragment of SdbA exhibits *in vitro* activity consistent with this predicted function.

Materials and Methods

In frame *effector-HIS3* fusion by yeast recombinational cloning

The *Saccharomyces cerevisiae* BY4742 (*MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0*) (Brachmann et al. 1998) strains overexpressing *lpg0275*, *lpg0969* and *lpg2482* (*sdbA*, *ravK*, and *sdbB* respectively) from the high-copy vector pYES2 NT/A (Life Technologies, GAL 1 promoter, N-terminal 6X HIS/Xpress tag and URA3 selectable marker) (Heidtman et al. 2009) were used to create the *effector-HIS3* fusion mutants by yeast recombinational cloning. The *S. cerevisiae HIS3* gene was PCR amplified from pAG423GAL-ccdB (Alberti et al. 2007) using an effector specific forward primer containing the last 50-60 nucleotides of the effector (minus the stop codon) followed by the first 20-30 nucleotides of the *HIS3* sequence and the pYES-HIS3 reverse primer (Table S1). The resulting PCR products were transformed together with XbaI/PmeI digested pYES2 NT/A vector encoding *sdbA*, r*avK* or *sdbB* to BY4742 using the high-efficiency LiOAC/PEG method (Gietz and Schiestl 2007) and plated onto SD-uracil with 2% glucose (SD-Ura/gluc). The resulting transformants were screened by PCR and sequence verified. To confirm that the *HIS3* fusion does not interfere with effector function, the ability of the effector-His3 fusion protein to cause a yeast growth defect was tested by comparing the growth BY4742 with empty vector control, the wild-type effector and the effector-His3 fusion in a yeast spot dilution assay as described previously (Urbanus et al. 2016).

Selection of loss-of-function mutations

The *effector-HIS3* fusion vectors were mutagenized in XL-1 Red (Agilent) as per manufacturer's instructions. XL-1 Red transformants were washed off the transformation plate, grown overnight in 50 ml LB with ampicillin and the resulting mutant plasmid pool was purified using PureYield

Plasmid Midipreps (Promega). The mutant plasmid pool was transformed to BY4742 using the high-efficiency LiOAC/PEG method (Gietz and Schiestl 2007). Four transformation reactions were performed per screen, each using 1 μg of plasmid pool per reaction. One reaction was split in three parts and plated onto different media types to quantify the transformable (intact selection marker and origin of replication) plasmids (SD-Ura/gluc), loss-of-function mutants (SD-uracil +2% galactose (SD-Ura/gal)), and missense loss-of-function mutants (SD-uracil/histidine + 2% galactose (SD-Ura/His/gal)) and incubated for 2-4 days at 30°C. The three remaining transformation reactions were plated onto 150 mm SD-Ura/His/gal plates and allowed to grow until colonies appeared (3-4 days). Plasmids were rescued from missense loss-of-function mutants and transformed to the *E. coli* Top10 strain before sequencing using primers in the vector and effector, if required (Table S1). Details of the sequenced mutants are shown in Table S2.

Analysis of mutant fitness

Liquid growth assays were used to assess the effect of missense loss-of-function mutations on yeast fitness as described (Urbanus et al. 2016) with the following modifications. Overnight cultures of freshly transformed BY4742 with empty vector control, pYES2 NT/A *effector-HIS3* wild-type and mutants were diluted 100-fold into 100 μl of SD-Ura/gal and grown with Breathe-Easy adhesive seals (EK scientific) in a CellGrower robot (S&P robotics) at 30ºC with intermittent shaking. Yeast growth was monitored for 30 h by measuring the OD_{620} every 15 min. Growth fitness was calculated as the ratio of the area under the curve (AUC) of an effector-expressing strain over an empty vector control after 30 h using the R package GrowthCurver

(Sprouffske and Wagner 2016). The average AUC ratio and standard deviation was calculated from three technical replicates.

Expression of missense loss-of-function mutants

Expression levels of the missense loss-of-function mutants were assessed by western blot. BY4742 strains with empty vector controls, wild-type *effector*, *effector-HIS3* fusion and *effector-HIS3* mutant clones were grown overnight in SD-Ura/gluc. To induce expression 10 OD_{600} units of were washed with SD-Ura/gal, resuspended in 5 ml SD-Ura/gal and grown for 6 h at 30°C. Three OD_{600} units were harvested, treated as described (Zhang et al. 2011), resuspended in 100 µl 2X sample buffer and incubated for 5 min at 95°C. Samples were analyzed using SDS-PAGE and western blot using the following antibodies: anti-Xpress (1:5000, catalog nr R910-25, Invitrogen), anti-actin (1:2,500, catalog nr A2066, Sigma-Aldrich), and secondary antibodies anti-Mouse HRP (1:5000) or anti-Rabbit HRP (1:5000) (Cell Signaling Technology, catalog nr 7074 and 7076).

HHpred analysis and sequence alignments

The amino acid sequence of RavK and SdbA (amino acid residues 528-1116) were submitted to the HHpred server (https://toolkit.tuebingen.mpg.de/#/) (Zimmermann et al. 2018) analyzed using MSA generation HHblits Uniclust20_2017_07 and Uniprot20_2016_02, respectively and otherwise default parameters. The resulting alignments were visualized using Boxshade (https://embnet.vital-it.ch/software/BOX_form.html). Amino acid sequence alignments of SdbB with its orthologs or with SidB were generated using T-coffee (Di Tommaso et al. 2011) and visualized using Jalview (Waterhouse et al. 2009) or Boxshade.

Nucleotide sugar donor specificity of the SdbA C-terminal domain

The gene fragment corresponding to SdbA residues 510-1050 was PCR amplified from *L. pneumophila* str Philadelphia-1 genomic DNA and inserted into the pMCSG53 plasmid (Eschenfeldt et al. 2013) by ligation independent cloning, providing an N-terminal 6xHIS-TEV tag. The point mutant E963A was prepared by site-directed mutagenesis using QuikChange™ site-directed mutagenesis kit (Stratagene) according to the manufacturer's protocol. Plasmids were sequenced and transformed into the *E. coli* BL21 DE3 Gold strain for purification. Recombinant proteins were purified to near homogeneity (>95%) using Ni-chelate affinity chromatography on Ni-NTA Superflow resin (Qiagen) using standard protocols. Cultures were grown in TB and expression was induced at an OD_{595} of 0.8 with 0.4 mM IPTG overnight at 16 °C. Cells were harvested by centrifugation at 9,300× *g*, resuspended in 50 mM HEPES pH 7.5, 400 mM NaCl, 5% glycerol, 5 mM imidazole, and lysed by sonication. Lysates were clarified by centrifugation at 21,000× *g* at 4°C and loaded onto gravity flow Ni-NTA agarose columns (Qiagen), followed by washing with 50 mM HEPES pH 7.5, 400 mM NaCl, 5% glycerol, 30 mM imidazole. Proteins were eluted using 50 mM HEPES pH 7.5, 400 mM NaCl, 5% glycerol, 250 mM imidazole and flash-frozen in liquid nitrogen for storage at −80°C. The purity of the protein samples was assessed by SDS-PAGE and visualized by Coomassie Brilliant Blue R.

The nucleotide sugar donor specificity of $SdbA_{510\text{-}1050}$ was assayed using the UDP-Glo Glycosyltransferase Assay (Promega) according to manufacturer's protocol. Briefly, 0.09 µM of purified wild-type and E963A mutant $SdbA_{510\text{-}1050}$ protein was incubated with 100 µM UDP-glucose, UDP-GlcNAc, UDP-glucuron, UDP-galactose or UDP-GalNAc for 1h at 30°C in 50 mM HEPES, pH 7.5, 100 mM KCl, 2 mM $MgCl_2$, 1 mM $MnCl_2$. The hydrolysis of the UDP-

substrate was detected as the release of UDP by the UDP-Glo assay (Promega) after 20 min of incubation with UDP-Glo detection reagent. Luminescence was measured using a SpectraMax M2 plate reader. Three technical repeats were performed per reaction.

The V_{max}, K_m and k_{cat} for wild-type $SdbA_{510-1050}$ with UDP-GlcNAc was determined by incubating 0.16 µM $SdbA_{510-1050}$ with UDP-GlcNAc concentration range of 0.0039 – 2 mM for 1h at 30°C in 50 mM HEPES, pH 7.5, 100 mM KCl, 2 mM $MgCl_2$, 1 mM $MnCl_2$. Three technical repeats were performed per reaction. Kinetic parameters were determined by non-linear curve fitting from the Michaelis Menten plot using GraphPad Prism (version 5.00 for Windows, GraphPad Software).

Results

A random mutagenesis screen to identify regions important for bacterial effector function

To efficiently screen for randomly generated mutations in *L. pneumophila* effectors that cause a loss-of-function phenotype and represent full-length protein rather than frameshift or nonsense mutations, we applied a yeast method first designed to select for human cDNA inserts that contain intact open-reading frames (Holz et al. 2001). This method leverages tools developed for plasmid selection and protein expression in the model organism *Saccharomyces cerevisiae* (budding yeast). A strain with several auxotrophic alleles (*his3Δ1, leu2Δ0, met15Δ0, ura3Δ0;* genes involved in histidine, leucine, methionine, and uracil pathways) (Brachmann et al. 1998), allows for maintenance of a yeast plasmid encoding the wild-type allele as a selection marker. The wild-type alleles complement the auxotrophic allele and allow growth on medium lacking histidine, leucine, methionine or uracil (Sikorski and Hieter 1989).

We cloned the *S. cerevisiae HIS3* gene in-frame behind *L. pneumophila* effector genes on a high-copy galactose-inducible yeast expression plasmid with a *URA3* selection marker (Fig. 1). After confirming that the C-terminal His3 fusion does not interfere with the yeast growth phenotype, and therefore likely does not interfere with effector function, we generated a pool of random mutants using the *E. coli* mutator strain XL-1 Red. We then transformed this mutant plasmid pool to yeast and monitored growth on different media types. To assess the number of vectors with an intact backbone, where the *URA3* marker can complement the *ura3Δ0* allele of BY4742 and the vector has an intact origin of replication (ORI 2μ), the transformed pool was grown on medium lacking uracil and with glucose (SD-Ura/gluc) to repress expression of the *effector-HIS3* fusion (Fig. 1, step 3I). To look at the efficiency of the random mutagenesis step, we grew the transformed pool on medium lacking uracil and with galactose to induce expression of the *effector-HIS3* fusion (SD-Ura/gal) selecting all mutations that caused a loss of function in the effector (Fig. 1, step 3II). This can be caused by promoter mutations that disrupt expression or missense, nonsense, or frameshift mutations. Finally, to specifically select for full-length missense loss-of-function mutation, we grew the transformed pool on medium with galactose and lacking uracil and histidine (SD-Ura/His/gal), which requires the production of full-length effector-His3 fusion protein and mutations in the effector gene that disrupt effector activity (Fig. 1, step 3III).

Missense loss-of-function screen identifies important conserved SdbB residues

As a proof of principle, we looked at *sdbB* which causes a severe yeast growth phenotype when expressed (Fig. 2a) (Heidtman et al. 2009) and is part of the *sidB* paralog family, whose members are predicted to be lipases from the α/β hydrolase enzyme family (Luo and Isberg

2004). After verifying that the *sdbB-HIS3* fusion was still capable of causing a yeast growth defect (Fig. 2a), we created a *sdbB-HIS3* random mutagenesis pool and quantified the number of colony-forming units (CFUs) on the different selection media. While 1.89% of the transformable plasmids carried a mutation that allowed for growth on SD-Ura/gal medium indicating some type of loss-of-function mutation (Fig. 2b), only 0.03% of the transformable plasmids carried a mutation allowing growth on SD-Ura/His/gal medium–a condition that requires the expression of a full-length fusion protein. The efficiency of the histidine selection step was verified by sequencing 20 clones from each condition. The loss-of-function clones selected on SD-Ura/gal consisted of 16 frameshift mutations, 2 nonsense mutations, 1 combination of a missense and frameshift mutation and 1 missense mutation in the *sdbB* gene (Fig. 2c and d). In contrast, all 20 loss-of-function clones selected on SD-Ura/His/gal contained only missense mutations (Fig. 2c and f). In both conditions, a number of mutations were recovered several times suggesting that sequencing additional clones would yield few new mutations. To confirm that the identified mutations indeed rescued the *sdbB*-induced growth defect in yeast, we compared the growth of *sdbB* wild-type and mutants to an empty vector control in a liquid growth curve assay. Using the area under the growth curve (AUC) at 30 h, which encompasses differences in every growth phase, we calculated the fitness of the *sdbB* strains as the ratio AUC of *sdbB*/AUC empty vector control. The wild-type and *HIS3*-fused *sdbB* caused a severe yeast growth defect, while the *sdbB* loss of function mutants showed a fitness of 60-90% compared to the empty vector control (Fig. 2e and g).

The positions of the frameshift and nonsense mutations in SdbB (Fig. 2d) indicate that a large part of the protein is required for function, as even a nonsense mutation at S365, 84 amino acid residues from the C-terminus, almost completely rescued activity. The missense loss-of-

function mutations (Fig. 2d and f) target three amino acid residues (SD-Ura/gal: D273 and SD-Ura/His/gal: G116, H351) that are invariant in SdbB orthologs from *L. pneumophila* and other *Legionella* species (Burstein et al. 2016) (Fig. S1), suggesting they are essential for function or structure. The G189E mutation is part of the GXS/CXG motif that is highly conserved across the SdbB orthologs and is predicted by NCBI Conserved Domain search (Marchler-Bauer et al. 2017) to align with the so-called nucleophile elbow of the nucleophile-acid-base triad of the α/β hydrolase active site (Brenner 1988; Ollis et al. 1992; Schrag and Cygler 1997) (Fig. S2).

When the missense mutants are mapped onto the SdbB AlphaFold2 model (Jumper et al. 2021; Varadi et al. 2021) all but C309 localize in the vicinity of the putative catalytic cysteine (C187), including the invariant D273 and H351 residues captured in the screen, suggesting they are the remaining residues of the catalytic triad (Fig. 2h). A key strength of forward genetic approaches is to identify functionally important residues independent of bioinformatic predictions and conservation. Some of the functionally important residues that we identify using our method are conserved and some are not. As an example of the latter, two loss-of-function mutations targeting residues S279 and L281 localize near conserved residues within the SdbB AlphFold2 model (consistent with their apparently essential role in protein function) yet are themselves not conserved between *sdbB* orthologs. We verified expression of all missense loss-of-function mutants, and all are expressed at a higher level than wild-type *sdbB* with C309 having the lowest expression (Fig. S3a).

Thus, the *sdbB* example demonstrates that functionally important amino acid residues can be efficiently identified using the random mutagenesis method in conjunction with the histidine selection for full-length protein. Importantly, this approach significantly reduces the number of

sequenced clones required to identify amino acid residues or regions of interest, by approximately 60-fold in the case of *sdbB.*

Missense loss-of-function screen identifies the active site of the characterized effector RavK

To benchmark the missense loss-of-function screen on a more characterized effector, we looked at *ravK* which also causes a severe yeast growth defect (Heidtman et al. 2009; Liu et al. 2017). RavK is a small, soluble metalloprotease that specifically cleaves host actin, and directed substitutions within the predicted active site motif HExxH abolish both its activity and toxicity to yeast (Liu et al. 2017). After confirming that the *ravK-HIS3* fusion was still able to cause a yeast growth when expressed (Fig. 3b), we subjected this plasmid to random mutagenesis, transformed the pool into yeast, and then selected for loss-of-function mutants on SD-Ura/His/gal medium. Of the 14 loss-of-function clones we sequenced, one clone contained a large, in-frame deletion from amino acid residue 70 to residue 166, which was unexpected but confirms the strength of the histidine selection for maintaining open reading frames. All other loss-of-function clones were caused by single point mutations resulting in missense mutations (Fig. 3a). In the growth assay, wild-type and *HIS3*-fused *ravK* almost completely inhibited yeast growth, while the *ravK* mutants displayed a fitness of 70-90% compared to the empty vector control. All *ravK* mutant are expressed at a similar level (Fig. S3b). The loss-of-function mutations all map to the first half of RavK, suggesting that the N-terminal half of RavK is essential for RavK function. This agrees with the previous study which identified the active site motif (H95ExxH99) in the N-terminal half of the protein and found that the 50 C-terminal residues of RavK can be deleted without any effect on its activity on actin (Liu et al. 2017).

Four of the loss-of-function mutations that we isolated targeted the active site motif H95ExxH99, three of which are mutations in the E96 codon. Our screen also identified several residues outside of this motif that are critical for RavK function. To investigate why these might be functionally important, we performed an HHpred analysis which looks for structural homologs of proteins (Zimmermann et al. 2018). HHpred identified many hits with homology to the HExxH metalloprotease motif. Among the top five HHpred hits are three small soluble metalloproteases or minigluzincins, anthrax lethal factor and a zinc dependent peptidase from the M48 family (Dalkas et al. 2010; López-Pelegrín et al. 2013; López-Pelegrín et al. 2014) (Fig. S4). Notably, some of the other loss-of-function mutations occur in areas that have homology with structural elements in the minigluzincins contributing to the active site cleft (López-Pelegrín et al. 2013) (Fig. S4). Indeed, when the missense mutations are mapped on the RavK AlphaFold2 model (Jumper et al. 2021; Varadi et al. 2021) (Fig. 3c) they are localized around the active site HExxH including the top rim of the active site cleft.

The C-terminal domain of SdbA is a putative glycosyltransferase

Like *sdbB*, *sdbA* is a member of the s*idB* paralog family (Luo and Isberg 2004). While the function of SdbA remains undefined, experimental evolution of *Legionella* in mouse macrophages selected for parallel *sdbA* nonsense and frameshift mutations in three out of four independent lineages (Ensminger et al. 2012), suggesting that SdbA activity partially restricts growth in this accidental host. While the N-terminal domain of SdbA has homology with SidB (Luo and Isberg 2004), the additional C-terminal domain does not have significant sequence homology to other known proteins (data not shown). Expression of *sdbA* completely inhibits yeast growth (Heidtman et al. 2009), making the missense loss-of-function screen an informative

tool to identify functional residues that might suggest a specific activity inside the eukaryotic cell.

The missense loss-of-function screen in *sdbA* identified 19 mutations in 24 sequenced clones targeting 17 codons (Fig. 4a). In contrast to the smaller genes *sdbB* and *ravK*, the *sdbA* results included several double mutants. Some of these mutations were also recovered as a single loss-of-function mutant with a similar fitness (Fig. 4b). The *sdbA* mutant clones are expressed at a higher (albeit varying) level than wild-type *sdbA*, which could not be detected by western blot (Fig. S3c). All the single mutations that lead to a loss-of-function phenotype fall in the C-terminal domain and concentrate in two regions: G541-GTGHI-S547 and G957-GLSVM-E963. An HHpred homology search (Zimmermann et al. 2018) predicted with high confidence that the C-terminal domain is a glycosyltransferase of the GT-B fold. When comparing the SdbA C-terminal domain with the sequence of *E. coli* MurG, a well-studied member of the GT-B fold glycosyltransferase family, the two mutated regions align with the G-loop 1 and a consensus region in GT-B fold superfamily involved in binding the donor molecule (Ha et al. 2000; Hu et al. 2003; Crouvoisier et al. 2007) (Fig. 4c). Glycosyltranferases hydrolyze UDP-sugar donor molecules and transfer the sugar to the acceptor molecule, which can be a variety of molecules such as small molecules, lipids or proteins (Lairson et al. 2008). In MurG residues A263, L264, L265, E268, Q287 and Q288 contact the donor molecule UDP-GlcNAc (Hu et al. 2003) (Fig. 4c) while the G-loop 1 is thought to be involved in acceptor molecule binding (Ha et al. 2000). Mutations in these motifs abrogate MurG enzymatic activity, including mutation of the residues H18 and E268 (Hu et al. 2003; Crouvoisier et al. 2007), whose corresponding residues in SdbA (H545 and E963) were found to be mutated in our screen. The single missense mutants abrogating SdbA activity were mapped onto the AlphaFold2 model (Jumper et al. 2021; Varadi

et al. 2021) of a SdbA C-terminal fragment (residues 510-1050) with the residues H545 and E963 highlighted in yellow.

To test whether the C-terminal domain of SdbA is indeed a glycosyltransferase, we purified the C-terminal fragment (residues 510-1050) and the equivalent of the MurG E268A inactive mutant in SdbA (E963A) and tested several UDP-sugars as substrate using the UDP-Glo assay (Fig. 4e). Glycosyltransferases can hydrolyze UDP-sugars in the absence of an acceptor molecule (with water acting as an acceptor in the reaction) (Sheikh et al. 2017; Vicente et al. 2023). Indeed wild-type $SdbA_{510\text{-}1050}$ hydrolyzes the UDP-GlcNAc donor, while the E963A mutant does not (Fig. 4e). This suggests that SdbA is a glycosyltransferase with specificity for UDP-GlcNAc and that a mutation in the E963 codon identified by the missense loss-of-function screen abrogates that activity. Using the same assay, we determined the kinetic parameters of UDP-GlcNAc hydrolysis by $SdbA_{510\text{-}1050}$ (Fig. 4f), which revealed high affinity (low micromolar Km) of the enzyme to UDP-GlcNAc. Taken together, these data support the prediction of SdbA glycosyltransferase activity and demonstrate the power of missense mutations and *in silico* predictions to inform the functional determination of effector activity.

Discussion

The identification of functionally important residues through amino acid substitutions is a common tool to interrogate protein activities, test structural predictions, and define protein-protein interaction interfaces. This is typically done through site-directed mutagenesis (reverse genetics) or random mutagenesis (forward genetics). One advantage of random mutagenesis is that it can identify functionally important residues that might be missed by the site-directed mutagenesis approach (e.g. due to lack of apparent sequence homology or incomplete/inaccurate bioinformatic predictions). Nonsense mutations and frameshifts provide limited insight into protein structure or function yet typically make up the majority of loss-of-function mutations recovered after random mutagenesis of a protein sequence. Here, we demonstrate that the combination of a random mutagenesis loss-of-function screen with a selection for full-length protein is highly effective in specifically selecting for loss-of-function missense clones. In fact, all but one of the clones recovered in our assay contained missense mutations, while the remaining one contained an in-frame deletion which included the active site of RavK. The percentage of missense mutant clones for *sdbB* was 0.3% of transformable plasmids, but this number will be different for each gene and experiment. It will depend on the efficiency of the mutagenesis step, the gene length and the number of functionally important codons that can be mutated by a single mutation. We recommend performing one yeast transformation with a new mutant plasmid pool and plate one third on the different selection media as described in the methods (SD-Ura/gluc, SD-Ura/gal and SD-Ura/His/gal). This will indicate whether the random mutagenesis worked and gauge how many yeast transformations are needed to isolate sufficient loss-of-function clones.

We have benchmarked this approach against three effectors in *L. pneumophila*, using existing structural information and advances in protein modeling (AlphaFold) to reveal how a forward genetic approach can both validate existing knowledge and models and reveal novel functional residues that would be difficult to predict using only homology or modeling alone. Notably, our approach correctly identified residues in the previously described active site of RavK, the predicted active site nucleophile motif of SdbB and the remaining putative residues of the catalytic triad of SdbB. Forward genetics can also lead to unexpected results, in our case, this came in the form of insight into the functional divergence of two apparent paralogs. While our approach enriched for several loss-of-function mutations within the α/β hydrolase domain of *sdbB*, in *sdbA*, the loss-of-function mutations that we recovered localized to its C-terminal domain. Based on this mutational profile, we were able to predict that the C-terminal domain of SdbA is a glycosyltransferase, a hypothesis supported by *in vitro* activity towards UDP-GlcNAc. For proteins with even less structural information, such mutational profiling can focus follow up studies designed to link specific domains to activity.

A close examination of our data shows that not all important functional residues were identified in the screens, especially those within the active site residues of RavK and SdbB. This suggests that the mutational space of these proteins has not been saturated, though from a practical perspective we recovered several identical mutations in each, suggesting diminishing returns of sequencing additional clones. To expand sampling of the mutational space, an alternative method of random mutagenesis such as error prone PCR could be used to increase the number of mutations. In a direct comparison of these methods, error prone PCR introduced more mutations than the XL-1 Red mutator strain, though the incidence of multiple mutations per clone would increase (Rasila et al. 2009).

To effectively apply the random mutagenesis missense enrichment selection or extend this approach to bacteria and mammalian cells, several considerations should be taken into account. First, the protein of interest must have a selectable loss-of-function phenotype such as alleviation of growth defect. Growth fitness is a universal phenotype that is easy to measure in bacteria, yeasts and mammalian cell lines. Second, the C-terminal fusion of a selection marker must not interfere with protein function. If the function of protein of interest is inhibited by the C-terminal fusion, it could potentially be overcome by introducing linker regions of varying length and flexibility (Chen et al. 2013) or by using a cleavable linker such as the ubiquitin K0 mutant that is processed by cytosolic deubiquitinases in eukaryotic cells (Bachran et al. 2013). Similarly, the C-terminal selection marker must be able to function as a fusion protein or be liberated by an *in vivo* cleavable linker. To extend the random mutagenesis missense enrichment selection to bacteria, the chloramphenicol acetyltransferase (CAT) gene (which confers resistance to chloramphenicol) is a viable candidate as a C-terminal fusion partner. CAT has been successfully used in protein fusions where it conferred chloramphenicol resistance during colony selection as a C-terminal fusion partner, with increased selection efficiency when mutant fusion proteins were soluble (Maxwell et al. 1999). In mammalian cell lines, a positive selection marker such as blasticidin S deamidase could be used as a C-terminal fusion partner. Blasticidin S deamidase is functional as a C-terminally fused protein (Suarez and McElwain 2009) and confers resistance against blasticidin, which rapidly inhibits mammalian cell growth at a low dose (Kimura et al. 1994). An alternative, if no positive selection marker is available, is GFP, which has been used extensively as a fused localization marker for various cellular compartments and organisms (Margolin 2000; Roessel and Brand 2002; Huh et al. 2003). After a

standard number of generation doublings, GFP-positive cells, indicative of the presence of full-length protein, can be isolated by fluorescence-activated cell sorting.

Our initial results suggest that missense-directed mutagenesis will be a useful tool to help identify potential functions for other bacterial effector proteins, many of which have low sequence homology to characterized proteins (Gomez-Valero et al. 2011; Burstein et al. 2016). Rather than being replaced by *in silico* protein modeling, we show how the two methodologies complement one another and can be used to identify structural features or regions essential for activity against the eukaryotic cell. In some cases, the combined information of functional residues, protein models or sequence conservation may not indicate an apparent activity or function of the effector. Even in these cases, the loss-of-function mutants may prove useful in other assays with growth-based readouts.

In *L. pneumophila* alone 10% of the translocated effectors cause severe yeast growth defects (Campodonico et al. 2005; Shohdy et al. 2005; de Felipe et al. 2008; Heidtman et al. 2009; Shen et al. 2009; Guo et al. 2014; Urbanus et al. 2016) and are possible candidates for the random mutagenesis missense enrichment screen. Additional growth phenotypes are likely to be revealed under other conditions of growth, such as nocodazole, caffeine, high osmolarity, pH, low and high temperature, or yeast deletion strains which are known to potentiate some bacterial effectors (Sisko et al. 2006; Slagowski et al. 2008; Xu et al. 2010; Bosis et al. 2011; Nevo et al. 2014). Identifying functional residues within uncharacterized effectors is a logical first step towards validating *in silico* protein models, predicting effector activity, and designing protein-protein interaction studies.

Data availability

Strains and plasmids are available upon request. Supplemental material available at G3 online. Figure S1 shows an amino acid alignment of SdbB orthologs from seven Legionella species along with the location of missense mutations identified in this screen. Figure S2 shows an amino acid alignment of SidB and SdbB amino acid sequence alignment. Figure S3 contains western blots of effector-HIS3 wildtype and loss-of-function mutants. Figure S4 shows an HHpred alignment of RavK with metalloproteases. Figure S5 shows the purity of the SdbA fragments used in the UDP-Glo Glycosyltransferase assay. Table S1 shows the primers used for yeast recombinational cloning and sequencing. Table S2 lists all the mutations identified in this study.

Acknowledgments

We thank Kamran Rizzolo and Lisa Shao for help with cloning and Shayna Deecker, Beth Nicholson, Jordan Lin, Morgan Petersen and John MacPherson, for their suggestions and careful reading of the manuscript.

Funding

This work was supported by a Project Grant (AWE, AS) from the Canadian Institutes of Health Research (PJT-162256), an NIH NIAID contract #75N93022C00035 (AS), as well as by the NSERC Strategic Network grant IBN (AFK, ANK).

Author contributions

AWE conceived and designed the random mutagenesis with missense enrichment screen and performed the *sdbA* screen, initially under the supervision of RRI. AG refined the screen and created the *effector-HIS3* fusion constructs. DB analyzed the *sdbA* data. DB, HOM and TMZ performed the *sdbB-HIS3* screen and analyzed sequencing data. TMZ performed the benchmarking experiments for *sdbB*. MLU provided experimental supervision, performed the *ravK* screen and analysis, growth fitness analysis with assistance of TMZ, expression analysis and created figures. ANK purified SdbA, performed the enzymatic assays and created figures. PJS, AS, SRI, AFY and AWE provided project supervision and advice. MLU and AWE prepared the paper with input from others.

Figure legends

Figure 1: A method to enrich full-length missense mutants in a random mutagenesis screen. A schematic representation of the random mutagenesis screen enriched for missense mutations. 1: *L. pneumophila* effector genes (magenta) causing a severe growth phenotype when expressed in yeast were fused in-frame to the *S. cerevisiae HIS3* gene (teal) on a high-copy (ORI 2µ, green) yeast expression vector with galactose-inducible expression (P_{GAL1}, grey) and an uracil selection marker (*URA3*, brown). 2: The plasmids were randomly mutated in the *E. coli* mutator strain XL-1 Red. The resulting mutant pool contained a variety of mutations: sense (black line), nonsense (stop), missense (red line) and frameshifts (yellow line), of which the latter three can cause loss-of-function phenotypes. 3: The plasmid pool was transformed to the *S. cerevisiae* strain BY4742 and grown under conditions that selected for an intact vector backbone (I, SD -Ura with glucose), all loss-of-function mutations (II, SD-Ura with galactose induction) and for expression of full-length effector-His3 fusion proteins caused by missense loss-of-function mutations (III, SD-Ura/His with galactose induction). The 100 mm plate images each represent 8% of the transformation pool used in the *sdbB* screen.

Figure 2: The full-length ORF enriched random mutagenesis screen identifies missense mutations in the putative active site of SdbB. a) Expressing *sdbB* and the *sdbB* in-frame fusion with *HIS3* caused a severe yeast growth defect as shown in a yeast spot dilution assay. A dilution series of yeast strains carrying an empty vector or plasmids with *sdbB* and *sdbB-HIS3* were spotted on SD-Ura/gluc (uninduced condition) and SD-Ura/gal (inducing conditions) and grown 2 days at 30°C before imaging. **b)** The percentage of loss-of-function mutations in the pool of transformable plasmids selected on SD-Ura/gal (grey) and SD-Ura/His/gal (black) medium. The

average and standard deviation of three independent replicates are shown. **c)** Percentage of the occurrence of frameshift, nonsense and missense mutations in 20 sequenced clones selected on SD-Ura/gal (grey) and SD-Ura/His/gal (black). **d)** A schematic representation of loss-of-function clones selected on SD-Ura/gal. Mutations recovered from the same clone are shown in black and are connected by a horizontal line, while single mutations are shown in red. Mutation type is indicated as a closed triangle (frameshift), open hexagon (nonsense) or closed circle (missense), and the number of symbols reflects the occurrence of the mutation in the dataset. **e)** The fitness of wild-type *sdbB*, *sdbB-HIS3* fusion and loss-of-function mutant strains (selected on SD-Ura/gal) compared to empty vector controls confirmed the loss-of-function phenotype for the *sdbB* random mutagenesis clones. The fitness was determined using growth curve assays (see methods) and calculated as the ratio (wt or mt *sdbB*/empty vector control) of the area under the growth curve at 30 h. The average and standard deviation of three technical replicates are shown. **f)** A schematic representation of SdbB loss-of-function clones selected on SD-Ura/His/gal. Missense loss-of-function mutations are shown in red with a closed circle, the number of symbols reflects the occurrence of the mutation in the dataset. Amino acids shown for presentation purposes are shown in grey. **g)** The fitness of wild-type *sdbB*, *sdbB-HIS3* fusion and loss-of-function mutants (selected on SD-Ura-His/gal) compared to empty vector controls in liquid growth assays confirms the loss-of-function phenotype for the *sdbB* random mutagenesis strains. The average and standard deviation of three technical replicates are shown. **h)** The missense mutations identified by the random mutagenesis screen (**d, f**) are shown in orange on an Alphafold2 model of SdbB (AF-Q5ZSN5-F1-model_v4.pdb), and residues from the putative active site G185xCxG189 not captured by the screen are shown in magenta. Putative catalytic triad C187-D273-H351 residues are shown with sticks and the box shows the enlargement of the

putative catalytic site. The AlphaFold2 model was visualized using the PyMOL Molecular Graphics System, Version 2.2 Schrödinger, LLC.

Figure 3: *ravK* random mutagenesis captures residues lining the RavK active site cleft. a) A schematic representation of RavK and the amino acids changed by mutations causing a *ravK* loss-of-function phenotype when expressed in yeast. Mutated residues are shown in red, amino acids shown for presentation purposes are in grey and the number of symbols reflects the occurrence of the mutation in the dataset. **b)** The fitness of wild-type *ravK*, *ravK-HIS3* and loss-of-function mutants confirms the loss-of-function phenotype for the *ravK* random mutagenesis strains. The fitness was calculated as the ratio (wt or mt *ravK*/empty vector control) of the area under the growth curve at 30 h. The average and standard deviation of three technical replicates are shown. **c)** The Alphafold2 model of RavK (AF-Q5ZWW5-F1-model_v4.pdb) with missense mutations shown in orange. The histidine residues of the active site motive H95-Exx-H99 (Liu et al. 2017) not captured by the screen are shown in red. The residues in the active site cleft are shown as sticks. The AlphaFold2 model was visualized using the PyMOL Molecular Graphics System, Version 2.2 Schrödinger, LLC.

Figure 4: The C-terminal domain of SdbA is a putative glycosyltransferase domain. **a)** A schematic representation of SdbA and the residues changed by the mutations causing a s*dbA* loss-of-function phenotype when expressed in yeast. Mutations recovered from the same clone are shown in black, connected by a line and single mutations alleviating a *sdbA* induced growth defect are shown in red. Amino acids shown for presentation purposes are shown in grey. Black and red closed circles indicate the number of times the mutation was identified. **b)** The

fitness of wild-type *sdbA*, *sdbA-HIS3* and loss-of-function mutants compared to empty vector controls. The fitness was calculated as the ratio (wt or mt *sdbA*/empty vector control) of the area under the growth curve at 30 h. *sdbA* and *sdbA-HIS3* display a severe growth defect, while the loss-of-function mutations rescue growth up to 50% of the empty vector control. The average and standard deviation of three technical replicates are shown. **c)** HHpred alignment of SdbA with 1F0K, *E. coli* MurG (Ha et al. 2000). The G-loop 1 and consensus sequence are shown with identical residues (black) and similar residues (grey) highlighted. SdbA residues that when mutated abrogate activity, are indicated by a red closed circle. Residues in the MurG consensus sequence that contact UDP-GlcNac are indicated by a black star, or a red star if mutating the residue abrogates MurG activity. **d)** The Alphafold2 model of SdbA (AF-Q5ZYT6-F1-model_v4.pdb) with missense mutations shown in orange. Residues H545 and E963 corresponding to residues H18 and E268 in MurG, are shown in yellow. The AlphaFold2 model was visualized using the PyMOL Molecular Graphics System, Version 2.2 Schrödinger, LLC. **e)** The SdbA glycosyltransferase domain uses GlcNAc as a donor substrate. Donor substrate specificity was tested by incubating 0.09 µM of purified fragment (residues 510-1050) of wild-type SdbA or an inactive mutant (E963A) with 100 µM UDP-glucose, UDP-GlcNAc, UDP-glucuron, UDP-galactose or UDP-GalNAc for 1 h at 30°C. The hydrolysis of the UDP-substrate was detected as the release of UDP by the UDP-Glo assay (Promega) after 20 min of incubation with UDP-Glo detection reagent. Luminescence was measured using a SpectraMax M2 platereader and the hydrolysis activity was calculated as mU (nmoles UDP-substrate/min) per mg $SdbA_{510\text{-}1050}$ wt or E963K. Three technical repeats were performed per reaction. **f)** Determination of the kinetic parameters for GlcNAc hydrolysis by $SdbA_{510\text{-}1050}$. Reactions with 0.16 µg purified wild-type C-terminal domain of SdbA (residues 510-1050) and a range of

0.0039 – 2 mM GlcNAc were incubated for 1 h at 30°C. GlcNAc hydrolysis was measured using the UDP-Glo glycosyltransferase assay as described above; three technical replicates were performed per reaction. Kinetic parameters were determined by non-linear curve fitting from the Michaelis-Menten plot.

References

Alberti S, Gitler AD, Lindquist S. 2007. A suite of Gateway® cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. Yeast. 24(10):913–919. doi:10.1002/yea.1502.

Bachran C, Morley T, Abdelazim S, Fattah RJ, Liu S, Leppla SH. 2013. Anthrax Toxin-Mediated Delivery of the *Pseudomonas* Exotoxin A Enzymatic Domain to the Cytosol of Tumor Cells via Cleavable Ubiquitin Fusions. Mbio. 4(3):e00201-13. doi:10.1128/mbio.00201-13.

Black MH, Osinski A, Gradowski M, Servage KA, Pawłowski K, Tomchick DR, Tagliabracci VS. 2019. Bacterial pseudokinase catalyzes protein polyglutamylation to inhibit the SidE-family ubiquitin ligases. Science. 364(6442):787–792. doi:10.1126/science.aaw7446.

Bosis E, Salomon D, Sessa G. 2011. A Simple Yeast-Based Strategy to Identify Host Cellular Processes Targeted by Bacterial Effector Proteins. Plos One. 6(11):e27698. doi:10.1371/journal.pone.0027698.

Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P, Boeke JD. 1998. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast. 14(2):115–132. doi:10.1002/(sici)1097-0061(19980130)14:2<115::aid-yea204>3.0.co;2-2.

Brenner S. 1988. The molecular evolution of genes and proteins: a tale of two serines. Nature. 334(6182):528–530. doi:10.1038/334528a0.

Burstein D, Amaro F, Zusman T, Lifshitz Z, Cohen O, Gilbert JA, Pupko T, Shuman HA, Segal G. 2016. Genomic analysis of 38 *Legionella* species identifies large and diverse effector repertoires. Nat Genet. 48(2):167–175. doi:10.1038/ng.3481.

Burstein D, Zusman T, Degtyar E, Viner R, Segal G, Pupko T. 2009. Genome-Scale Identification of *Legionella pneumophila* Effectors Using a Machine Learning Approach. Plos Pathog. 5(7):e1000508. doi:10.1371/journal.ppat.1000508.

Campodonico EM, Chesnel L, Roy CR. 2005. A yeast genetic system for the identification and characterization of substrate proteins transferred into host cells by the *Legionella pneumophila* Dot/Icm system. Mol Microbiol. 56(4):918–933. doi:10.1111/j.1365-2958.2005.04595.x.

Chen X, Zaro JL, Shen W-C. 2013. Fusion protein linkers: Property, design and functionality. Adv Drug Deliver Rev. 65(10):1357–1369. doi:10.1016/j.addr.2012.09.039.

Crouvoisier M, Auger G, Blanot D, Mengin-Lecreulx D. 2007. Role of the amino acid invariants in the active site of MurG as evaluated by site-directed mutagenesis. Biochimie. 89(12):1498–1508. doi:10.1016/j.biochi.2007.06.011.

Dalkas GA, Chasapis CT, Gkazonis PV, Bentrop D, Spyroulias GA. 2010. Conformational Dynamics of the Anthrax Lethal Factor Catalytic Center. Biochemistry. 49(51):10767–10769. doi:10.1021/bi1017792.

de Felipe KS, Glover RT, Charpentier X, Anderson OR, Reyes M, Pericone CD, Shuman HA. 2008. *Legionella* Eukaryotic-Like Type IV Substrates Interfere with Organelle Trafficking. PLoS Pathog. 4(8):e1000117. doi:10.1371/journal.ppat.1000117.

Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang J-M, Taly J-F, Notredame C. 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res. 39(suppl_2):W13–W17. doi:10.1093/nar/gkr245.

Ensminger AW, Yassin Y, Miron A, Isberg RR. 2012. Experimental Evolution of *Legionella pneumophila* in Mouse Macrophages Leads to Strains with Altered Determinants of Environmental Survival. Plos Pathog. 8(5):e1002731. doi:10.1371/journal.ppat.1002731.

Eschenfeldt WH, Makowska-Grzyska M, Stols L, Donnelly MI, Jedrzejczak R, Joachimiak A. 2013. New LIC vectors for production of proteins from genes containing rare codons. J Struct Funct Genom. 14(4):135–144. doi:10.1007/s10969-013-9163-9.

Escoll P, Rolando M, Gomez-Valero L, Buchrieser C. 2013. Molecular Mechanisms in *Legionella* Pathogenesis. Curr Top Microbiol. 376:1–34. doi:10.1007/82_2013_351.

Fields BS, Benson RF, Besser RE. 2002. *Legionella* and Legionnaires' Disease: 25 Years of Investigation. Clin Microbiol Rev. 15(3):506–526. doi:10.1128/cmr.15.3.506-526.2002.

Finsel I, Hilbi H. 2015. Formation of a pathogen vacuole according to *Legionella pneumophila*: how to kill one bird with many stones. Cell Microbiol. 17(7):935–950. doi:10.1111/cmi.12450.

Gietz RD, Schiestl RH. 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc. 2(1):31–34. doi:10.1038/nprot.2007.13.

Gomez-Valero L, Rusniok C, Carson D, Mondino S, Pérez-Cobas AE, Rolando M, Pasricha S, Reuter S, Demirtas J, Crumbach J, et al. 2019. More than 18,000 effectors in the *Legionella* genus genome provide multiple, independent combinations for replication in human cells. Proc Natl Acad Sci. 116(6):2265–2273. doi:10.1073/pnas.1808016116.

Gomez-Valero L, Rusniok C, Cazalet C, Buchrieser C. 2011. Comparative and Functional Genomics of *Legionella* Identified Eukaryotic Like Proteins as Key Players in Host–Pathogen Interactions. Front Microbiol. 2:208. doi:10.3389/fmicb.2011.00208.

Gomez-Valero L, Rusniok C, Rolando M, Neou M, Dervins-Ravault D, Demirtas J, Rouy Z, Moore RJ, Chen H, Petty NK, et al. 2014. Comparative analyses of *Legionella* species identifies genetic features of strains causing Legionnaires' disease. Genome Biol. 15(11):505. doi:10.1186/s13059-014-0505-0.

Guo Z, Stephenson R, Qiu J, Zheng S, Luo Z-Q. 2014. A *Legionella* effector modulates host cytoskeletal structure by inhibiting actin polymerization. Microbes Infect. 16(3):225–236. doi:10.1016/j.micinf.2013.11.007.

Ha S, Walker D, Shi Y, Walker S. 2000. The 1.9 Å crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. Protein Sci. 9(6):1045–1052. doi:10.1110/ps.9.6.1045.

Heidtman M, Chen EJ, Moy M, Isberg RR. 2009. Large-scale identification of *Legionella pneumophila* Dot/Icm substrates that modulate host cell vesicle trafficking pathways. Cell Microbiol. 11(2):230–248. doi:10.1111/j.1462-5822.2008.01249.x.

Holz C, Lueking A, Bovekamp L, Gutjahr C, Bolotina N, Lehrach H, Cahill DJ. 2001. A Human cDNA Expression Library in Yeast Enriched for Open Reading Frames. Genome Res. 11(10):1730–1735. doi:10.1101/gr.181501.

Hsieh T-S, Lopez VA, Black MH, Osinski A, Pawłowski K, Tomchick DR, Liou J, Tagliabracci VS. 2021. Dynamic remodeling of host membranes by self-organizing bacterial effectors. Science. 372(6545):935–941. doi:10.1126/science.aay8118.

Hu Y, Chen L, Ha S, Gross B, Falcone B, Walker D, Mokhtarzadeh M, Walker S. 2003. Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. Proc National Acad Sci. 100(3):845–849. doi:10.1073/pnas.0235749100.

Huang L, Boyd D, Amyot WM, Hempstead AD, Luo Z, O'Connor TJ, Chen C, Machner M, Montminy T, Isberg RR. 2011. The E Block motif is associated with *Legionella pneumophila* translocated substrates. Cell Microbiol. 13(2):227–245. doi:10.1111/j.1462-5822.2010.01531.x.

Huh W-K, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK. 2003. Global analysis of protein localization in budding yeast. Nature. 425(6959):686–691. doi:10.1038/nature02026.

Isberg RR, O'Connor TJ, Heidtman M. 2009. The *Legionella pneumophila* replication vacuole: making a cosy niche inside host cells. Nat Rev Microbiol. 7(1):13–24. doi:10.1038/nrmicro1967.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. Nature. 596(7873):583–589. doi:10.1038/s41586-021-03819-2.

Kimura M, Takatsuki A, Yamaguchi I. 1994. Blasticidin S deaminase gene from *Aspergillus terreus* (BSD): a new drug resistance gene for transfection of mammalian cells. Biochim Biophys Acta (BBA) - Gene Struct Expr. 1219(3):653–659. doi:10.1016/0167-4781(94)90224-0.

Kozlov G, Wong K, Gehring K. 2018. Crystal structure of the *Legionella* effector Lem22. Proteins Struct Funct Bioinform. 86(2):263–267. doi:10.1002/prot.25427.

Lairson LL, Henrissat B, Davies GJ, Withers SG. 2008. Glycosyltransferases: Structures, Functions, and Mechanisms. Annu Rev Biochem. 77(1):521–555. doi:10.1146/annurev.biochem.76.061005.092322.

Lesser CF, Miller SI. 2001. Expression of microbial virulence proteins in *Saccharomyces cerevisiae* models mammalian infection. EMBO J. 20(8):1840–1849. doi:10.1093/emboj/20.8.1840.

Lin Y-H, Lucas M, Evans TR, Abascal-Palacios G, Doms AG, Beauchene NA, Rojas AL, Hierro A, Machner MP. 2018. RavN is a member of a previously unrecognized group *of Legionella pneumophila* E3 ubiquitin ligases. Plos Pathog. 14(2):e1006897. doi:10.1371/journal.ppat.1006897.

Liu Y, Zhu W, Tan Y, Nakayasu ES, Staiger CJ, Luo Z-Q. 2017. A *Legionella* Effector Disrupts Host Cytoskeletal Structure by Cleaving Actin. PLoS Pathog. 13(1):e1006186. doi:10.1371/journal.ppat.1006186.

López-Pelegrín M, Cerdà-Costa N, Cintas-Pedrola A, Herranz-Trillo F, Bernadó P, Peinado JR, Arolas JL, Gomis-Rüth FX. 2014. Multiple Stable Conformations Account for Reversible Concentration-Dependent Oligomerization and Autoinhibition of a Metamorphic Metallopeptidase. Angewandte Chemie Int Ed. 53(40):10624–10630. doi:10.1002/anie.201405727.

López-Pelegrín M, Cerdà-Costa N, Martínez-Jiménez F, Cintas-Pedrola A, Canals A, Peinado JR, Marti-Renom MA, López-Otín C, Arolas JL, Gomis-Rüth FX. 2013. A Novel Family of Soluble Minimal Scaffolds Provides Structural Insight into the Catalytic Domains of Integral Membrane Metallopeptidases. J Biol Chem. 288(29):21279–21294. doi:10.1074/jbc.m113.476580.

Luo J, Wang L, Song L, Luo Z-Q. 2021. Exploitation of the Host Ubiquitin System: Means by *Legionella pneumophila*. Front Microbiol. 12:790442. doi:10.3389/fmicb.2021.790442.

Luo Z-Q, Isberg RR. 2004. Multiple substrates of the *Legionella pneumophila* Dot/Icm system identified by interbacterial protein transfer. P Natl Acad Sci Usa. 101(3):841–846. doi:10.1073/pnas.0304916101.

Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 45(D1):D200–D203. doi:10.1093/nar/gkw1129.

Margolin W. 2000. Green Fluorescent Protein as a Reporter for Macromolecular Localization in Bacterial Cells. Methods. 20(1):62–72. doi:10.1006/meth.1999.0906.

Maxwell KL, Mittermaier AK, Forman-Kay JD, Davidson AR. 1999. A simple *in vivo* assay for increased protein solubility. Protein Sci. 8(9):1908–1911. doi:10.1110/ps.8.9.1908.

Mondino S, Schmidt S, Rolando M, Escoll P, Gomez-Valero L, Buchrieser C. 2020. Legionnaires' Disease: State of the Art Knowledge of Pathogenesis Mechanisms of *Legionella*. Annu Rev Pathol: Mech Dis. 15(1):1–28. doi:10.1146/annurev-pathmechdis-012419-032742.

Morar M, Evdokimova E, Chang C, Ensminger AW, Savchenko A. 2015. Crystal structure of the *Legionella pneumophila* lem10 effector reveals a new member of the HD protein superfamily. Proteins Struct Funct Bioinform. 83(12):2319–2325. doi:10.1002/prot.24933.

Nevo O, Zusman T, Rasis M, Lifshitz Z, Segal G. 2014. Identification of *Legionella pneumophila* Effectors Regulated by the LetAS-RsmYZ-CsrA Regulatory Cascade, Many of Which Modulate Vesicular Trafficking. J Bacteriol. 196(3):681–692. doi:10.1128/jb.01175-13.

O'Connor TJ, Adepoju Y, Boyd D, Isberg RR. 2011. Minimization of the *Legionella pneumophila* genome reveals chromosomal regions involved in host range expansion. Proc National Acad Sci. 108(36):14733–14740. doi:10.1073/pnas.1111678108.

Ollis DL, Cheah E, Cygler M, Dijkstra B, Frolow F, Franken SM, Harel M, Remington SJ, Silman I, Schrag J, et al. 1992. The α / β hydrolase fold. Protein Eng Des Sel. 5(3):197–211. doi:10.1093/protein/5.3.197.

Pinotsis N, Waksman G. 2017. Structure of the WipA protein reveals a novel tyrosine protein phosphatase effector from *Legionella pneumophila*. J Biol Chem. 292(22):9240–9251. doi:10.1074/jbc.m117.781948.

Qiu J, Luo Z-Q. 2017. *Legionella* and *Coxiella* effectors: strength in diversity and activity. Nat Rev Microbiol. 15(10):591–605. doi:10.1038/nrmicro.2017.67.

Rasila TS, Pajunen MI, Savilahti H. 2009. Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, *Escherichia coli* mutator strain, and hydroxylamine treatment. Anal Biochem. 388(1):71–80. doi:10.1016/j.ab.2009.02.008.

Roessel P van, Brand AH. 2002. Imaging into the future: visualizing gene expression and protein interactions with fluorescent proteins. Nat Cell Biol. 4(1):E15–E20. doi:10.1038/ncb0102-e15.

Schrag JD, Cygler M. 1997. [4] Lipases and αβ hydrolase fold. Methods Enzym. 284:85–107. doi:10.1016/s0076-6879(97)84006-2.

Segal G, Purcell M, Shuman HA. 1998. Host cell killing and bacterial conjugation require overlapping sets of genes within a 22-kb region of the *Legionella pneumophila* genome. Proc National Acad Sci. 95(4):1669–1674. doi:10.1073/pnas.95.4.1669.

Shames SR. 2023. Eat or Be Eaten: Strategies Used by *Legionella* to Acquire Host-Derived Nutrients and Evade Lysosomal Degradation. Infect Immun. 91(4):e00441-22. doi:10.1128/iai.00441-22.

Sheikh MO, Halmo SM, Patel S, Middleton D, Takeuchi H, Schafer CM, West CM, Haltiwanger RS, Avci FY, Moremen KW, et al. 2017. Rapid screening of sugar-nucleotide donor specificities of putative glycosyltransferases. Glycobiology. 27(3):206–212. doi:10.1093/glycob/cww114.

Shen X, Banga S, Liu Y, Xu L, Gao P, Shamovsky I, Nudler E, Luo Z. 2009. Targeting eEF1A by a *Legionella pneumophila* effector leads to inhibition of protein synthesis and induction of host stress response. Cell Microbiol. 11(6):911–926. doi:10.1111/j.1462-5822.2009.01301.x.

Sherwood RK, Roy CR. 2016. Autophagy Evasion and Endoplasmic Reticulum Subversion: The Yin and Yang of *Legionella* Intracellular Infection. Annu Rev Microbiol. 70(1):413–433. doi:10.1146/annurev-micro-102215-095557.

Shohdy N, Efe JA, Emr SD, Shuman HA. 2005. Pathogen effector protein screening in yeast identifies *Legionella* factors that interfere with membrane trafficking. P Natl Acad Sci Usa. 102(13):4866–4871. doi:10.1073/pnas.0501315102.

Siggers KA, Lesser CF. 2008. The Yeast *Saccharomyces cerevisiae*: A Versatile Model System for the Identification and Characterization of Bacterial Virulence Proteins. Cell Host Microbe. 4(1):8–15. doi:10.1016/j.chom.2008.06.004.

Sikorski RS, Hieter P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics. 122(1):19–27. doi:10.1093/genetics/122.1.19.

Sisko JL, Spaeth K, Kumar Y, Valdivia RH. 2006. Multifunctional analysis of Chlamydia-specific genes in a yeast expression system. Mol Microbiol. 60(1):51–66. doi:10.1111/j.1365-2958.2006.05074.x.

Slagowski NL, Kramer RW, Morrison MF, LaBaer J, Lesser CF. 2008. A Functional Genomic Yeast Screen to Identify Pathogenic Bacterial Proteins. Plos Pathog. 4(1):e9. doi:10.1371/journal.ppat.0040009.

Sprouffske K, Wagner A. 2016. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. Bmc Bioinformatics. 17(1):172. doi:10.1186/s12859-016-1016-7.

Suarez CE, McElwain TF. 2009. Stable expression of a GFP-BSD fusion protein in *Babesia bovis* merozoites. Int J Parasitol. 39(3):289–297. doi:10.1016/j.ijpara.2008.08.006.

Sulpizio A, Minelli ME, Wan M, Burrowes PD, Wu X, Sanford EJ, Shin J-H, Williams BC, Goldberg ML, Smolka MB, et al. 2019. Protein polyglutamylation catalyzed by the bacterial calmodulin-dependent pseudokinase SidJ. Elife. 8:e51162. doi:10.7554/elife.51162.

Toulabi L, Wu X, Cheng Y, Mao Y. 2013. Identification and Structural Characterization of a *Legionella* Phosphoinositide Phosphatase. J Biol Chem. 288(34):24518–24527. doi:10.1074/jbc.m113.474239.

Urbanus ML, Quaile AT, Stogios PJ, Morar M, Rao C, Leo RD, Evdokimova E, Lam M, Oatway C, Cuff ME, et al. 2016. Diverse mechanisms of metaeffector activity in an intracellular bacterial pathogen, *Legionella pneumophila*. Mol Syst Biol. 12(12):893. doi:10.15252/msb.20167381.

Valdivia RH. 2004. Modeling the Function of Bacterial Virulence Factors in *Saccharomyces cerevisiae*. Eukaryot Cell. 3(4):827–834. doi:10.1128/ec.3.4.827-834.2004.

Valleau D, Quaile AT, Cui H, Xu X, Evdokimova E, Chang C, Cuff ME, Urbanus ML, Houliston S, Arrowsmith CH, et al. 2018. Discovery of Ubiquitin Deamidases in the Pathogenic Arsenal of *Legionella pneumophila*. Cell Reports. 23(2):568–583. doi:10.1016/j.celrep.2018.03.060.

Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, et al. 2021. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 50(D1):D439–D444. doi:10.1093/nar/gkab1061.

Vicente JB, Guerreiro ACL, Felgueiras B, Chapla D, Tehrani D, Moremen KW, Costa J. 2023. Glycosyltransferase 8 domain-containing protein 1 (GLT8D1) is a UDP-dependent galactosyltransferase. Sci Rep. 13(1):21684. doi:10.1038/s41598-023-48605-4.

Vogel JP, Andrews HL, Wong SK, Isberg RR. 1998. Conjugative Transfer by the Virulence System of *Legionella pneumophil*a. Science. 279(5352):873–876. doi:10.1126/science.279.5352.873.

Voth K, Pasricha S, Chung IYW, Wibawa RR, Zainudin ENHE, Hartland EL, Cygler M. 2021. Structural and Functional Characterization of *Legionella pneumophila* Effector MavL. Biomol. 11(12):1802. doi:10.3390/biom11121802.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 25(9):1189–1191. doi:10.1093/bioinformatics/btp033.

Wong K, Kozlov G, Zhang Y, Gehring K. 2015. Structure of the *Legionella* Effector, lpg1496, Suggests a Role in Nucleotide Metabolism. J Biol Chem. 290(41):24727–24737. doi:10.1074/jbc.m115.671263.

Xu L, Shen X, Bryan A, Banga S, Swanson MS, Luo Z-Q. 2010. Inhibition of Host Vacuolar H+-ATPase Activity by a *Legionella pneumophila* Effector. PLoS Pathog. 6(3):e1000822. doi:10.1371/journal.ppat.1000822.

Yang Y, Mei L, Chen J, Chen X, Wang Z, Liu L, Yang A. 2023. *Legionella pneumophila*-mediated host posttranslational modifications. J Mol Cell Biol. 15(5):mjad032. doi:10.1093/jmcb/mjad032.

Zhang T, Lei J, Yang H, Xu K, Wang R, Zhang Z. 2011. An improved method for whole protein extraction from yeast *Saccharomyces cerevisiae*. Yeast. 28(11):795–798. doi:10.1002/yea.1905.

Zhu W, Banga S, Tan Y, Zheng C, Stephenson R, Gately J, Luo Z-Q. 2011. Comprehensive Identification of Protein Substrates of the Dot/Icm Type IV Transporter of *Legionella pneumophila*. Plos One. 6(3):e17638. doi:10.1371/journal.pone.0017638.

Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol. 430(15):2237–2243. doi:10.1016/j.jmb.2017.12.007.

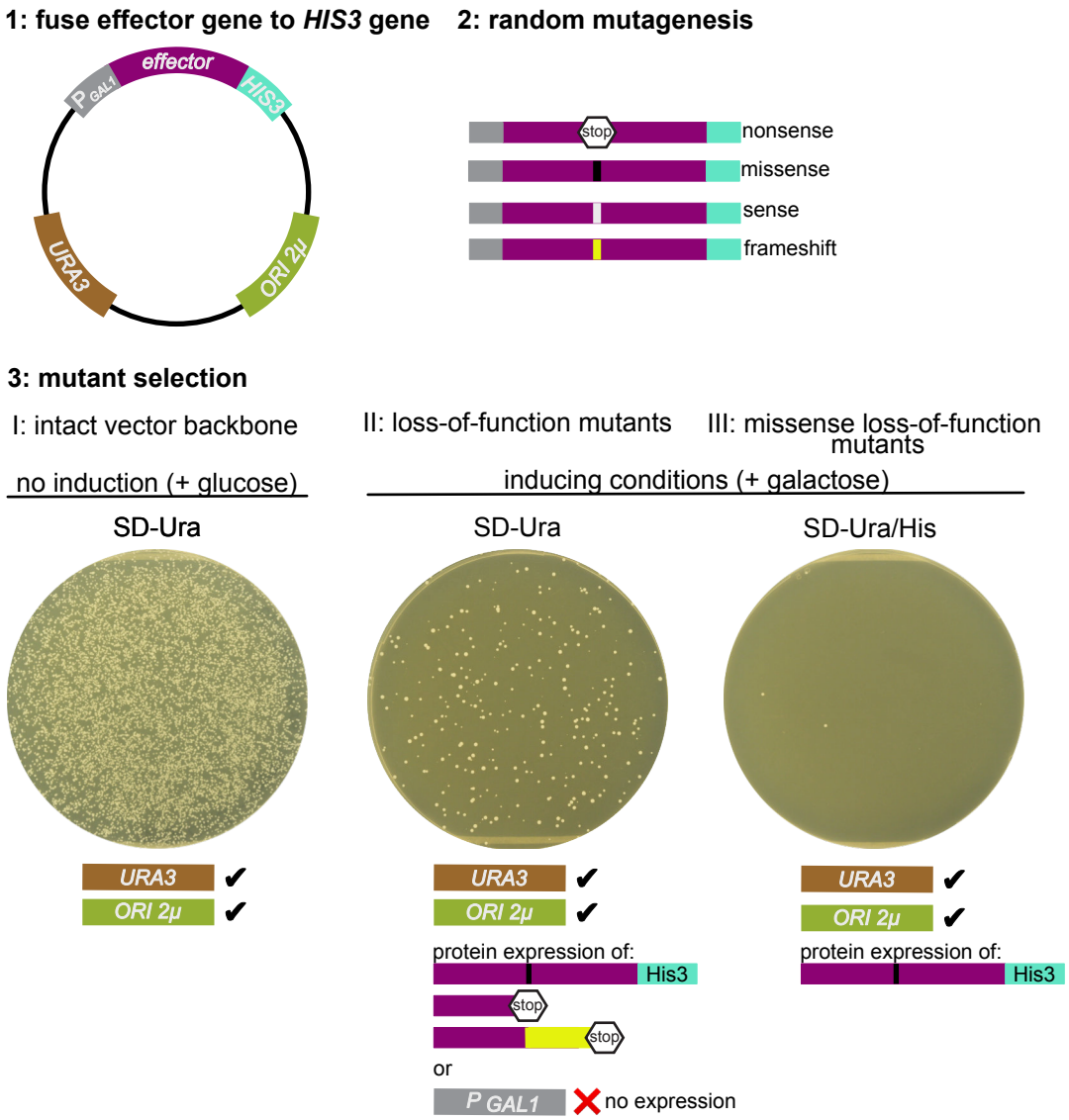

1: fuse effector gene to HIS3 gene
2: random mutagenesis
effector
P GAL1
HIS3
URA3
ORI 2μ
stop
nonsense
missense
sense
frameshift
3: mutant selection
I: intact vector backbone
II: loss-of-function mutants
III: missense loss-of-function mutants
no induction (+ glucose)
inducing conditions (+ galactose)
SD-Ura
SD-Ura
SD-Ura/His
URA3
ORI 2μ
protein expression of:
His3
stop
stop
or
P GAL1
no expression

Figure 1

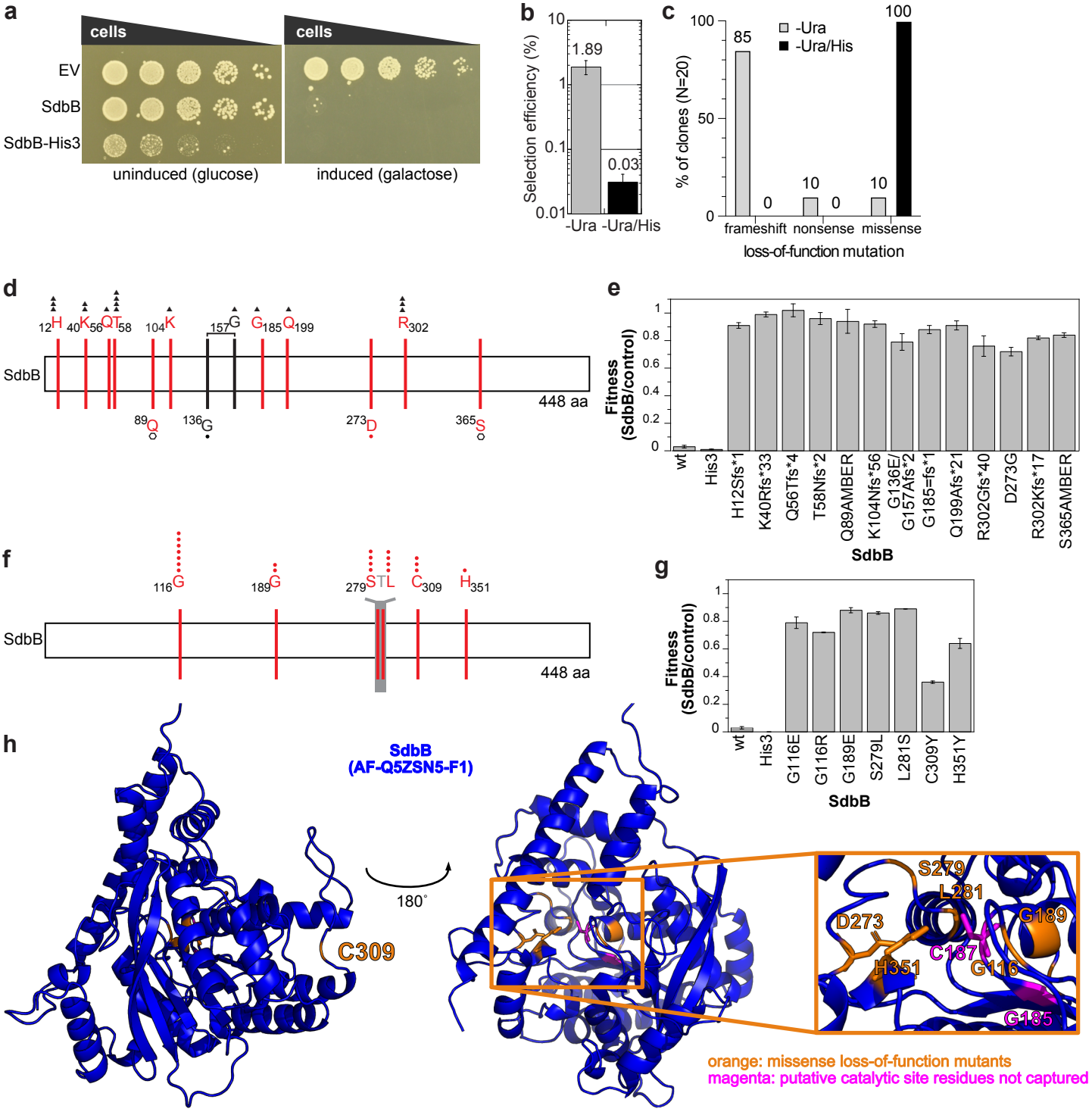

a
cells
cells
EV
SdbB
SdbB-His3
uninduced (glucose)
induced (galactose)
b
Selection efficiency (%)
10
1
0.1
0.01
1.89
0.03
-Ura
-Ura/His
c
% of clones (N=20)
100
50
0
85
0
10
0
10
100
-Ura
-Ura/His
frameshift
nonsense
missense
loss-of-function mutation
d
12H
40K
56QT58
104K
157G
G185
Q199
R302
SdbB
448 aa
89Q
136G
273D
365S
e
Fitness (SdbB/control)
1
0.8
0.6
0.4
0.2
0
wt
His3
H12Sfs*1
K40Rfs*33
Q56Tfs*4
T58Nfs*2
Q89AMBER
K104Nfs*56
G136E/ G157Afs*2
G185=fs*1
Q199Afs*21
R302Gfs*40
D273G
R302Kfs*17
S365AMBER
SdbB
f
116G
189G
279STL
C309
H351
SdbB
448 aa
g
Fitness (SdbB/control)
1
0.8
0.6
0.4
0.2
0
wt
His3
G116E
G116R
G189E
S279L
L281S
C309Y
H351Y
SdbB
h
SdbB
(AF-Q5ZSN5-F1)
C309
180˚
S279
L281
D273
G189
C187
H351
G116
G185
orange: missense loss-of-function mutants
magenta: putative catalytic site residues not captured

Figure 2

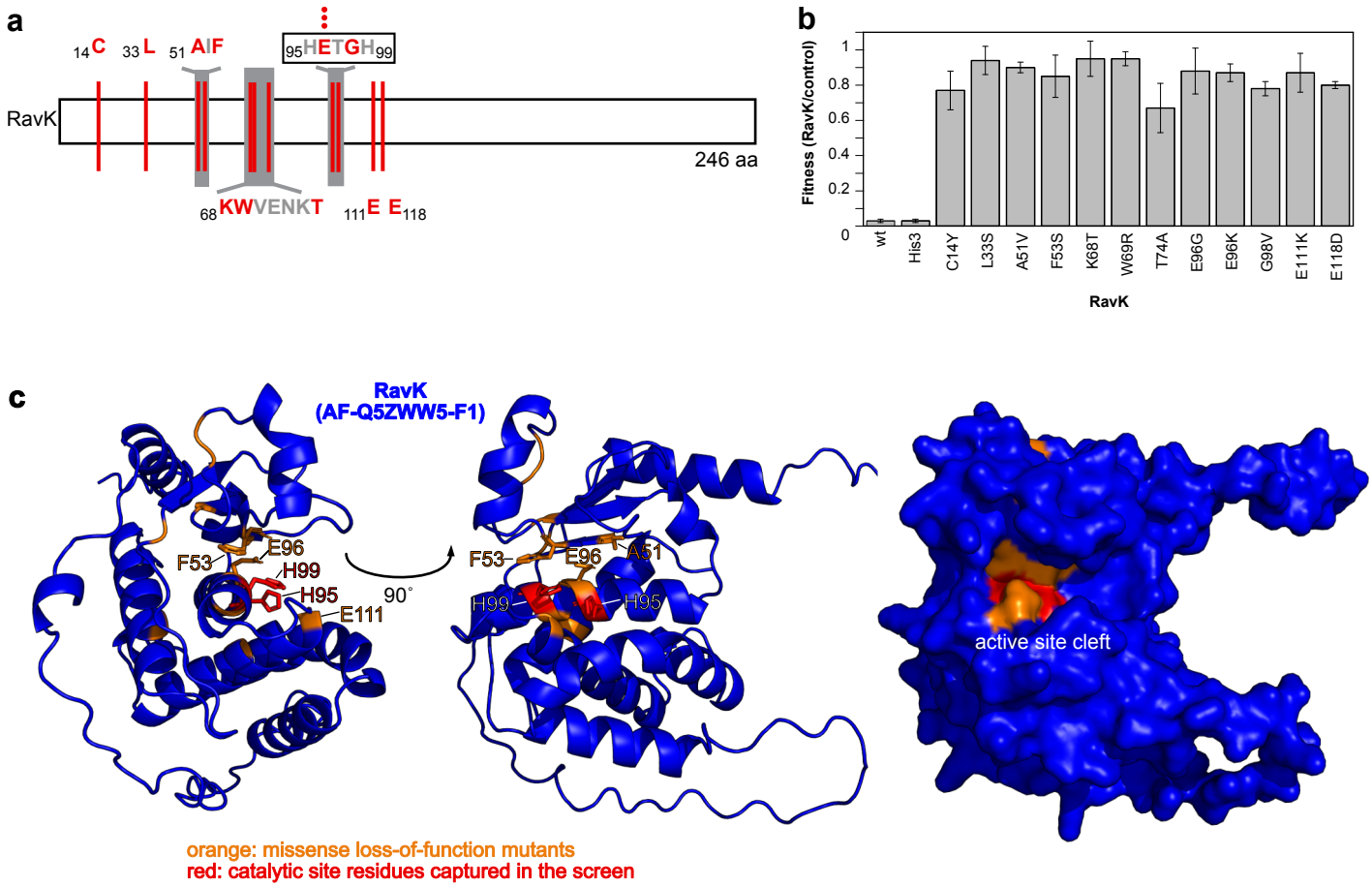
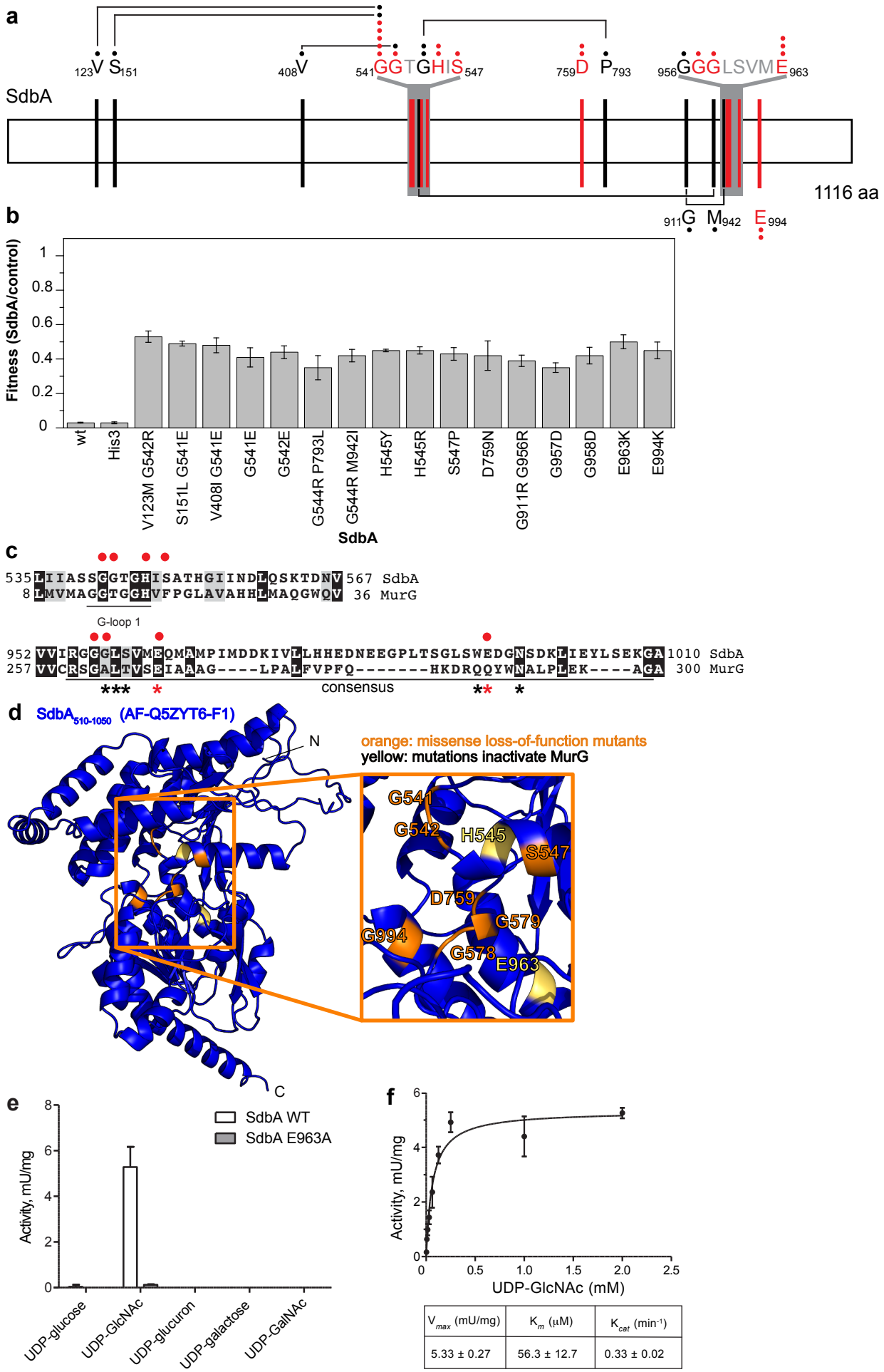

a
14C
33L
51AIF
95HETGH99
RavK
246 aa
68KWVENKT
111E E118
b
Fitness (RavK/control)
1
0.8
0.6
0.4
0.2
0
wt
His3
C14Y
L33S
A51V
F53S
K68T
W69R
T74A
E96G
E96K
G98V
E111K
E118D
RavK
c
RavK
(AF-Q5ZWW5-F1)
F53
E96
H99
H95
E111
90˚
F53
E96
A51
H99
H95
active site cleft
orange: missense loss-of-function mutants
red: catalytic site residues captured in the screen

Figure 3



V_{max} (mU/mg)	K_m (μM)	K_{cat} (min^{-1})
5.33 ± 0.27	56.3 ± 12.7	0.33 ± 0.02

Figure 4

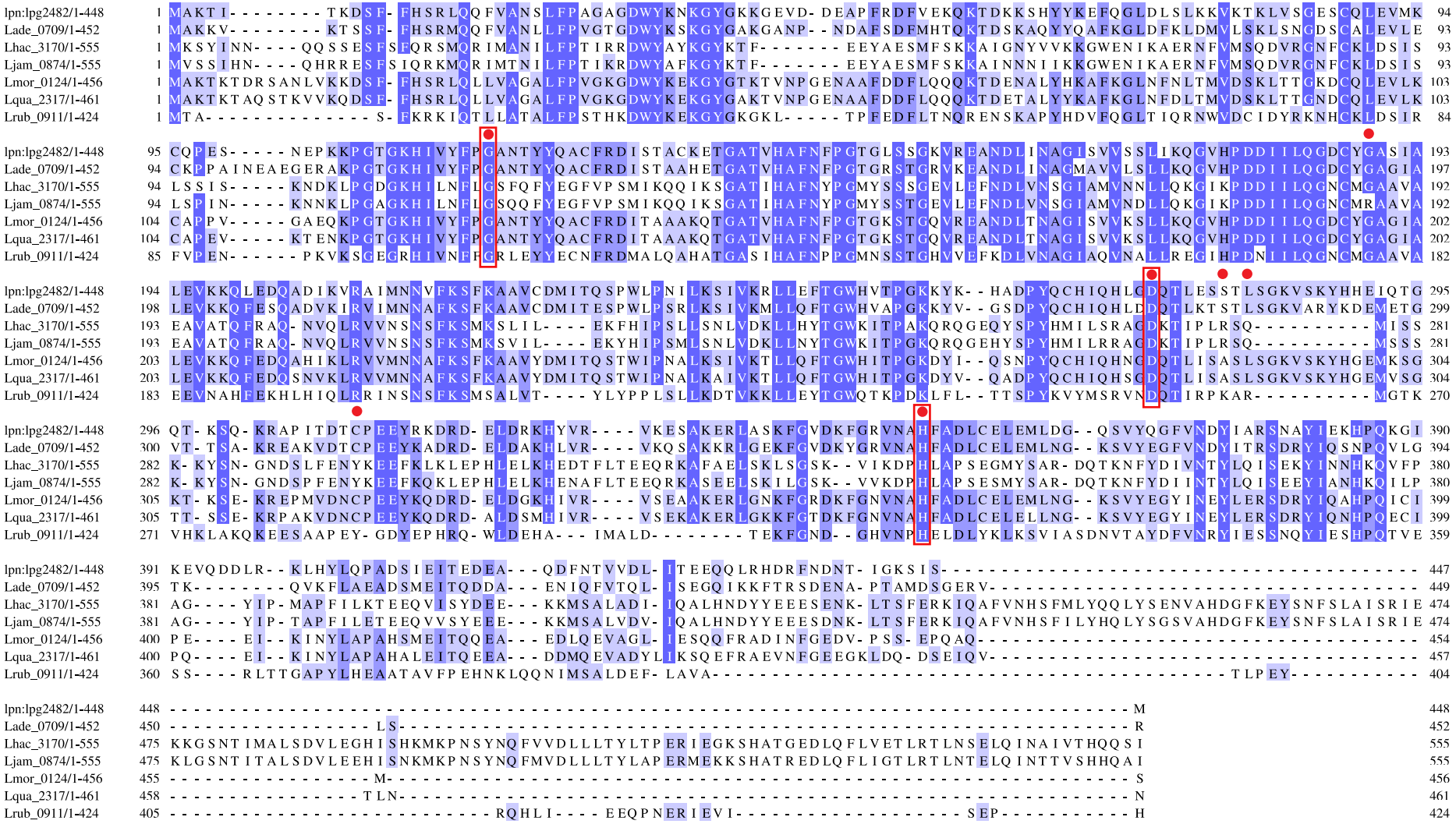


Figure S1: Random mutagenesis targets conserved residues in SdbB orthologs from seven *Legionella* species. The sequences of SdbB orthologs identified by Burstein *et al.* (Nat Genet 48(2): 167-175, 2016) were aligned using T-coffee, visualized with Jalview and coloured by % identity. The missense mutations identified by the random mutagenesis screen are indicated with a red closed circle above the SdbB (Lpg2482) sequence. Three of the seven mutations target invariant residues: G116, D273 and H351 shown in red boxes.

```
SidB    1 MAKIYNAPKPKYSGWEWFKFIAIRTVFPPVLLWDLIKIGANKLLGEWVSGLVLPAQNENF
SdbB    1 MAKTITKDSFFHSRLQ--QFVA-NSLFPAG-------------AGDWYKNKGYGKKGE-V

SidB   61 DDLAISDDTVSNYNED------------D--LICEKHDVITHDGAHLDTFEVRHRSQESI
SdbB   44 DDEAPFRDFVEKQKTDKKSHYYKEFQGLDLSLKKVKTKLVSGESCQLEVMKCQPESNEPK

SidB  107 DPKYQKYIINLVGNGMCYEHIIDDIKEDSKALKANVIGFNLRGVGQSTGKAKSSEDLVAD
SdbB  104 KPGTGKHIVYFPGANTYYQACFRDISTACKETGATVHAFNFPGTGLSSGKVREANDLINA

                                GxS/CxG
SidB  167 GIAQVQRLLDQGVSPQNITLKGHSLGAGVASLVAQHFHQLGQ-PINLFNSRSFSTITNFL
SdbB  164 GISVVSSLIKQGVHPDDIILQGDCYGASIALEVKKQLEDQADIKVRAIMNNVFKSFKAAV

SidB  226 VGHMRLERDEIGRAIGHKDSTVGTILGWLAKPFIKFGVALAKWEINAGSAFKSVPEAYKD
SdbB  224 C----------------DMITQSPWLPNILKSIVKRLLEFTGWHVTPGKKYKHA-DPYQC

SidB  286 YIVVRSRKEIRGERIDDAVIPHYASIHKELASERHKKKAEIDEEIANLDDIIRKADPLAK
SdbB  267 HIQHLGDQTLESSTLSGKVSKYHHEIQT--GQTKSQKRAPITDTC--PEEYRKDRDELDR

SidB  346 PGLANARDALVQAREKIKSDRKMETDVQYANGHNSDWNALHNRSGKSA-QTFFREFVQRT
SdbB  323 KHYVRVKE---SAKERLASKFGVD-KFGRVNAHFADLCELEMLDGQSVYQGFVNDYIARS

SidB  405 EA---DHAVKSIPEIN---------------------------------------------
SdbB  379 NAYIEKHPQKGIKEVQDDLRKLHYLQPADSIEITEDEAQDFNTVVDLITEEQQLRHDRFN

SidB      ----------
SdbB  439 DNTIGKSISM
```

Figure S2: SidB and SdbB amino acid sequence alignment. The amino acid sequences of SidB and SdbB were aligned with T-coffee and visualized with Boxshade, where identical residues are shown in black and similar residues in grey. The active site motif GxS/CxG predicted by NCBI conserved domain search (Marchler-Bauer A et al., Nucleic Acids Res.45(D)200-3, 2017) for SidB is indicated, suggesting that SdbB C187 is part of the active site catalytic triad.

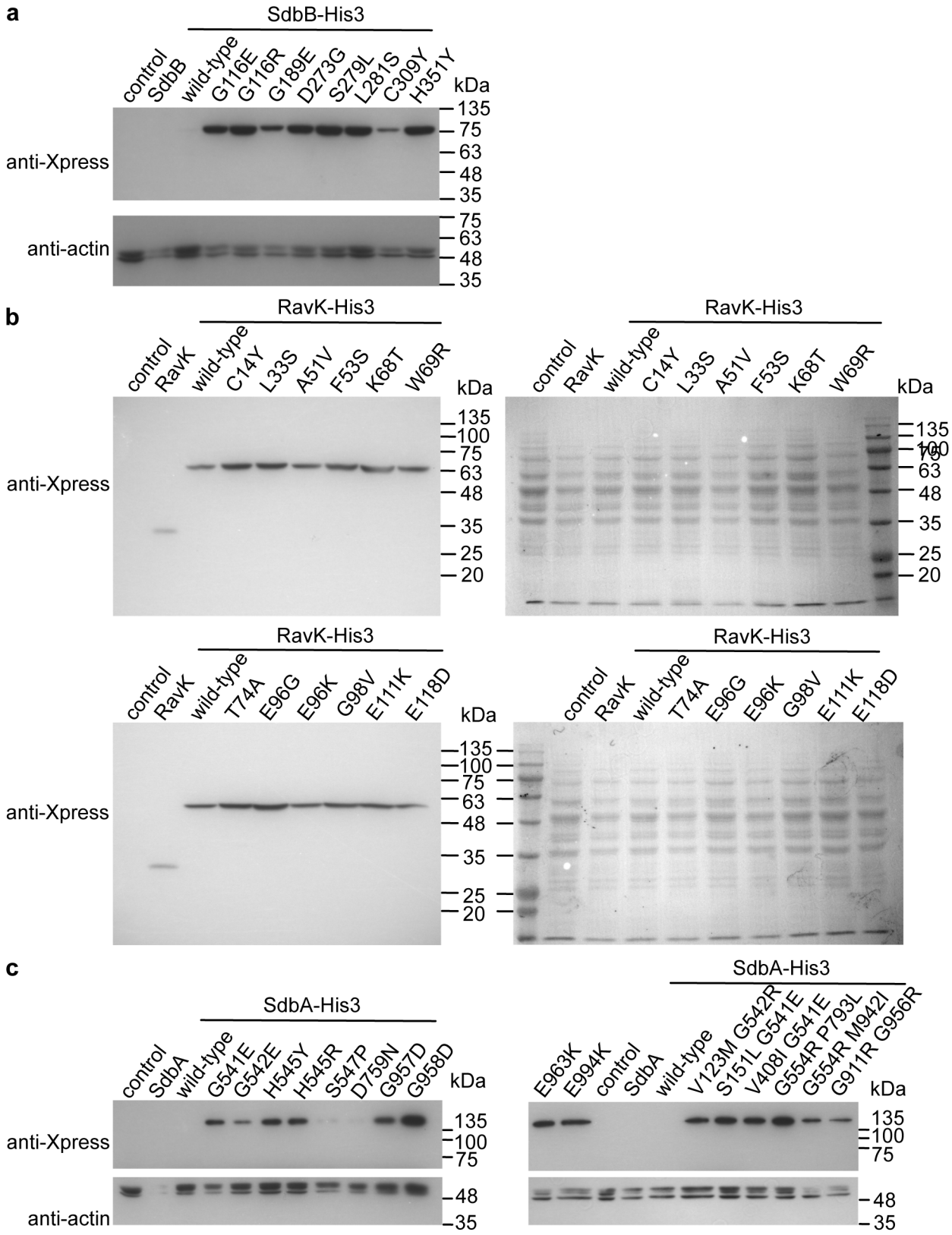


Figure S3. Expression of *effector-HIS3* wild-type and loss-of-function mutants. Yeast strains carrying pYES2NT/A effector-HIS3 mutant clones were grown overnight in SD-Ura/gluc, washed and diluted in SD-Ura/gal and grown for 6 h. Samples were analyzed by SDS-PAGE and western blot using the anti-Xpress antibody for the effectors and anti-actin as a loading control for SdbB and SdbA. **a)** Expression of Xpress-tagged SdbB, SdbB-His3 wild-type and mutants. Wild-type SdbB and SdbB-His3 are not detectable, but the loss-of-function mutants are. SdbB-His3 C309 has a lower expression level, but higher than the wild-type SdbB. **b)** Expression of Xpress-tagged RavK, RavK-His3 wild-type and mutants. Wild-type and mutant RavK can be detected in all samples. As RavK is an actin protease, the Ponceau S stain is shown as a loading control. **c)** Expression of Xpress-tagged SdbA, SdbA-His3 wild-type and mutants. Wild-type SdbA and SdbA-His3 are not detectable, but the loss-of-function mutants are. Clones with missense mutations S547P and D759N have a lower expression level, but higher than wild-type SdbA-His3.

```
                                                                                                β2
ravK       1 MVSLEHIQKLISECRKLGKDGLDNGTNGLIPELEIDVVPPSAFLG-VGNNPAIFVNSKTY
4QHJ_B     1 MKDRKILNEILSNTINE------LNLNDKKANIKIKIKPLKRKIA-SISL----------
4JIU_A     1 ----------------------------------------------LG-YQRPVKVRIRPLK-
4JIX_B     1 ----------------------------------------------LN-INESINIEIKPMK-
2L0R_A     1 ------------------------------------------------------------
3C37_A     1 -VNDPEVQRYVDKVGKRLLSGARAV----EFDYVFKVVKDDSVNAFAIPGGRVYVHTGLL
             rim of active site cleft
                                  β3
ravK      60 KLMRTTHEKWVENKTIVFKSYLL-S-QPAIKIIGAIVHETGHAFN------------V--
4QHJ_B    44 -----------TNKTIYINKNILPY-LSDEEIRFILAHELLHLKY------------GKY
4JIU_A    16 ---MSIARVSFKYGTITLDPAVL-N-LEEEEMFYILIHELAHLKAETSYHSSSFWREV--
4JIX_B    16 ---QKIASFSFKTKTLRLNKYVV-ENFDEELLHYIILHELIHFKI---------------
2L0R_A     1 ------------------------------NDSEGFIHEFGHAVD------------D--
3C37_A    56 KAAD-----------------------NETELAGVLAHEINHAVA---------------

ravK     104 -----AAKIPNT-EANACIFEIEVLMRLFQVKSPLLLGCTELDMQSYFKSRLTDYNKC--
4QHJ_B    80 HINEFEEELLFL-FPNKEAILFNLINKLFQKK----------------------------
4JIU_A    69 -----EKVFPGE-R--AKEIEDRIMTKLQR-------------------------------
4JIX_B       ------------------------------------------------------------
2L0R_A    17 -----YAGYLLDKNQSDLVTNSKKFIDIFKEEGSNLTSYGRTNEAEFFAEAFRLMHSTDH
3C37_A       ------------------------------------------------------------

ravK     156 ------VKDCQCLAEMVEFITHQ
4QHJ_B       -----------------------
4JIU_A       -----------------------
4JIX_B       -----------------------
2L0R_A    72 AERLKVQKNAPKTFQFINDQIKF
3C37_A       -----------------------
```

Figure S4: HHpred alignment of RavK with metalloproteases. Alignment of the HHpred top hits; MJ1213 (4QHJ) (M. López-Pelegrín et al., Angewandte Chemie Int Ed. 53, 10624–10630, 2014), proabylysin (4JIU), projannalysin (4JIX) (M. López-Pelegrín et al., J. Biol. Chem. 288, 21279–21294, 2013), anthrax lethal factor (2L0R) (G. A. Dalkas et al., Biochemistry. 49, 10767–10769, 2010) and a zinc-dependent endopeptidase of the M84 family (3C37) with RavK is shown where grey and black background indicate the level of amino acid residue conservation as 50% or more similarity or 50% or more identity, respectively. Structural elements of proabylysin (4JIU) are indicated on top of the alignment and mutations that abolish RavK activity are indicated by a red closed circle.

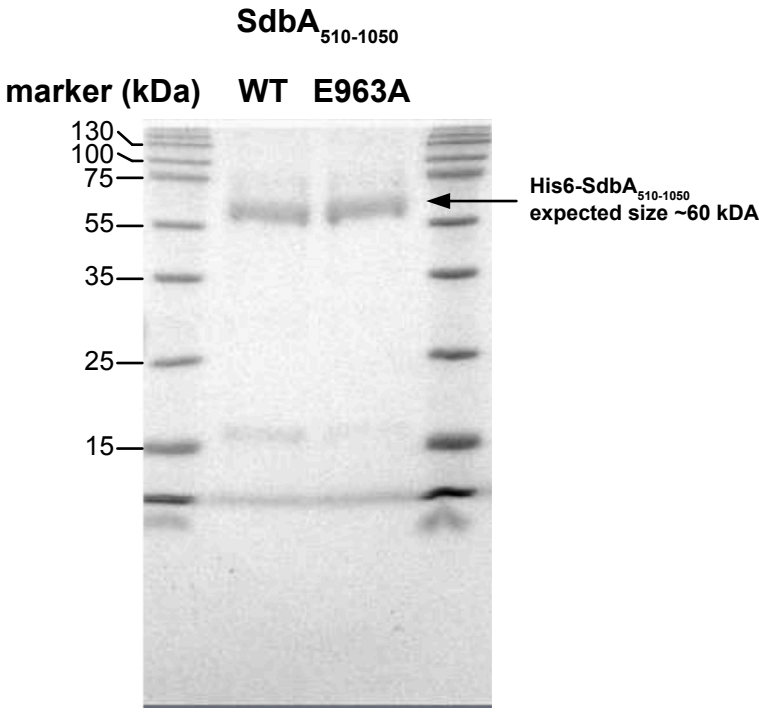


Figure S5: Purified $SdbA_{510-1050}$ wild-type and E963A.
The purity of the SdbA fragments used in the UDP-Glo Glycosyltransferase assay was assessed by SDS-PAGE and vizualized by Coomassie staining.