JBC RESEARCH ARTICLE

Substrate recognition in *Bacillus anthracis* sortase B beyond its canonical pentapeptide binding motif and use in sortase-mediated ligation

Received for publication, September 19, 2024, and in revised form, February 15, 2025 Published, Papers in Press, March 4, 2025,
https://doi.org/10.1016/j.jbc.2025.108382

Sophie N. Jackson, Darren E. Lee[‡], Jadon M. Blount[‡], Kayla A. Croney[‡], Justin W. Ibershof, Caroline M. Ceravolo, Kate M. Brown, Noah J. Goodwin-Rice, Kyle M. Whitham, James McCarty, John M. Antos[*], and Jeanine F. Amacher[*]

From the Department of Chemistry, Western Washington University, Bellingham, Washington, USA

Reviewed by members of the JBC Editorial Board. Edited by Joseph Jez

Sortases are critical cysteine transpeptidases that facilitate the attachment of proteins to the cell wall in Gram-positive bacteria. These enzymes are potential targets for novel antibiotic development, and versatile tools in protein engineering applications. There are six classes of sortases recognized, yet class A sortases (SrtA) are the most widely studied and utilized. SrtA enzymes endogenously recognize the amino acid sequence LPXTG, where X = any amino acid, with additional promiscuity now recognized in multiple positions for certain SrtA enzymes. Much less is known about Class B sortases (SrtB), which target a distinct sequence, typically with an N-terminal Asn, *e.g.*, variations of NPXTG or NPQTN. Although understudied overall, two SrtB enzymes were previously shown to be specific for heme transporter proteins, and *in vitro* experiments with the catalytic domains of these enzymes reveal activities significantly worse than SrtA from the same organisms. Here, we use protein biochemistry, structural analyses, and computational simulations to better understand and characterize these enzymes, specifically investigating *Bacillus anthracis* SrtB (baSrtB) as a model SrtB protein. Structural modeling predicts a plausible enzyme-substrate complex, which is verified by mutagenesis of binding cleft residues. Furthermore, residues N- and C-terminal to the pentapeptide recognition motif are critical for observed activity. Finally, we use chimeric proteins to identify mutations that improve baSrtB activity by ∼4-fold, and demonstrate the feasibility of sortase-mediated ligation using a baSrtB enzyme variant. These studies provide insight into SrtB-target binding as well as evidence that SrtB enzymes can be modified to be of potential use in protein engineering.

Surface proteins enable bacteria to perform functions that are critical for survival. In pathogens, these proteins are frequently virulence factors, although they also include those for environmental sensing and other cellular functions (1, 2). Many surface proteins in Gram-positive bacteria are anchored by sortase enzymes (1–3). Sortases are membrane-anchored cysteine transpeptidases that attach target proteins to the cell wall *via* a ping-pong ligation reaction (1–3). The first sortase enzyme discovered was the class A sortase from *Staphylococcus aureus* (saSrtA) (3, 4). Due to their critical role in the bacterial life cycle, sortase enzymes have emerged as potential targets for novel antibiotic development (5–7). Furthermore, because of their ability to ligate two protein-derived sequences together, sortases are powerful protein engineering tools, *via* sortase-mediated ligation (SML) or *sortagging* applications (8–11). SaSrtA remains the most widely utilized enzyme in SML experiments, as well as the best studied sortase. However, >10,000 sortase sequences have been identified from over 1000 bacterial species, which are generally grouped into six classes, A-F, based on primary sequence (12–14).

The sortase catalytic mechanism proceeds in a number of distinct steps *via* the action of a conserved triad of residues (His-Cys-Arg), with recent work also suggesting a key role for a conserved Thr residue immediately preceding the active site Cys (1, 15–18). Class A sortases generally recognize and anchor proteins containing a LPXTG motif within their Cell Wall Sorting Signal (CWSS), where X = any amino acid and Leu = P4, Pro = P3, X = P2, T = P1, and G = P1' (2, 19). Upon substrate binding by the sortase enzyme, the active site Cys nucleophile cleaves between the P1/P1′ positions, resulting in an acyl-enzyme intermediate (LPXT-enzyme). To facilitate formation of this intermediate, it has been proposed that the side chain hydroxyl of the highly conserved Thr residue, as well as the backbone amides of the catalytic Cys and the residue immediately C-terminal to the conserved His, serve as an oxyanion hole to stabilize charged intermediates en route to the acyl-enzyme state (15, 17, 20, 21). Next, the acyl-enzyme is resolved by nucleophilic attack by the amine terminus (often Gly or Ala) of a second enzyme substrate, leading to release of the final ligated product, LPXT-G/A (1). Throughout this entire process, the conserved Arg likely serves to bind and position the LPXT fragment, primarily through polar contacts with the carbonyls of the P4 Leu and P3 Pro (15, 17, 20). Previous work also suggested a possible role for this Arg in stabilizing oxyanion intermediates (1). Sortases from all classes studied to date appear to follow a similar mechanism, with

[‡] These authors contributed equally to this work.
[*] For correspondence: John M. Antos, antosj@wwu.edu; Jeanine F. Amacher, amachej@wwu.edu.

© 2025 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

differing classes regulated by distinct structural features of the sortase enzyme and/or variations in the substrate sequence(s) recognized (2, 12).

Sortases contain a conserved structural core. The sortase-fold is defined as an eight-stranded β-barrel (12, 17, 22, 23). Flexible loops connect one β-strand to the next and are designated accordingly, *e.g.*, the loop connecting the β6 strand to the β7 strand is the β6-β7 loop. Several of these loops perform an important role in enzyme recognition and specificity, notably the β4-β5, β6-β7, and β7-β8 loops, of which the sequences and lengths can vary widely between sortases of the same class (13, 17, 24–27). Between classes, there are several regions of additional structural variation (1, 25). These regions include the N-terminal region before the β-barrel, the β6-β7 loop, the β7-β8 loop, and the C-terminus extending away from the β-barrel. For example, as compared to SrtA enzymes, the SrtB β6-β7 loop is markedly longer and includes an elongated α-helix; this loop is known to contribute to substrate recognition (Fig. 1) (24, 28). The catalytic domains of SrtB enzymes also contain an extended N-terminal helix, the role of which is not known (Fig. 1). Finally, the stereochemistry of substrate binding to SrtB remains unclear; although there is a structure of saSrtB bound to the peptidomimetic NPQT* where *= (2*R*,3*S*)-3-amino-4-mercapto-2-butanol, the P4 Asn is making only one contact with saSrtB, limited to a single side chain hydrogen bond with the backbone carbonyl oxygen of T177; this is in contrast to its importance in the recognition motif (24, 28). Taken together, SrtB enzymes remain an understudied class of the sortase family.

While class A sortases are often referred to as general housekeeping enzymes, class B sortases (SrtBs) are more specific. Identified SrtB enzymes include those that recognize a specific heme transporter protein, IsdC, from the iron-regulated surface determinant (Isd) family (29, 30). IsdC is the only known substrate for *S. aureus* and *Bacillus anthracis* SrtB, which recognize NPQTN and NPKTG sequences, respectively (28, 31–33). Other studied SrtB enzymes are involved in different processes, *e.g.*, *Streptococcus pyogenes* SrtB polymerizes the pilus and requires a cofactor, SipA, *Clostridioides difficile* SrtB is suspected to recognize several proteins involved in adhesion, including CD0386, and cell wall hydrolysis, and *Listeria monocytogenes* SrtB targets proteins likely involved in iron acquisition, Lmo2185 and Lmo2186, but which utilize a different pathway than *S. aureus* or *B. anthracis* (2, 34–42). Of these few SrtB enzymes studied, not all have been tested for activity *in vitro*, and those that have revealed substantially reduced activity as compared to SrtA proteins from the same organisms (24, 28, 32, 36). Overall, substrate

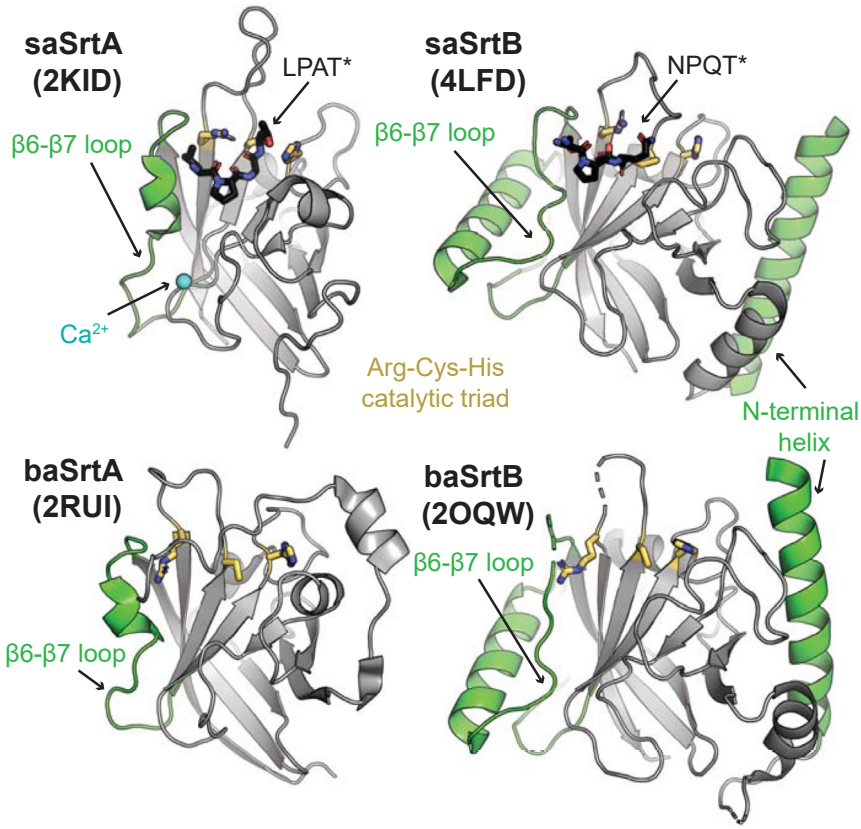

Figure 1. Class A and B sortase structures. The structures of *Staphylococcus aureus* SrtA (PDB ID 2KID) and SrtB (PDB ID 4LFD) enzymes, as well as *Bacillus anthracis* SrtA (PDB ID 2RUI) and SrtB (PDB ID 2OQW) are shown in cartoon representation and colored *gray* (23, 28, 50, 81). Structural features that differ between class A and B sortases are highlighted in *green*, which include a long N-terminal helix in SrtB and the length of the β6-β7 loop. The side chains of the Arg-Cys-His (*left*-to-*right* in this orientation) triad are shown as sticks and colored by atom (N = *blue*, S = *yellow*, C = *gold*). A calcium ion is in *blue* sphere representation in the saSrtA structure, and peptidomimetics (LPAT* and NPQT*, respectively) for saSrtA and saSrtB are in *black* sticks and colored by atom (O = *red*).

preferences for these SrtB enzymes vary, incorporating various residues at all positions: [E/N/P/Q/S/V]-[A/P/V]-[K/Q/S]-[S/T]-[G/N/S]. This is in contrast to the stringent specificity of SrtA enzymes for the LPXTG, or very similar, recognition sequences (4, 43–45). Notably, work by ourselves and others has discovered additional promiscuity in SrtA enzymes at the P1 and P1′ positions, particularly for SrtA proteins from the *Streptococcus* genera, using *in vitro* experiments (43–45). However, identified and putative cellular SrtA substrates are thought to generally follow the LPXTG motif (25, 27, 43–47).

To further our understanding of SrtB enzymology, and potentially leverage the activity of these enzymes for protein engineering, in this study we have explored the activity of SrtB enzymes from multiple organisms (*B. anthracis, C. difficile, S. aureus,* and *L. monocytogenes*) along with a series of mutants. These enzymes were chosen based on previous work, described above, as well as the pathogenic nature of the bacterial species, suggesting a deeper understanding may assist in future therapeutic efforts (48). Of these, only *B. anthracis* SrtB (baSrtB) revealed reasonable catalytic activity; therefore, we chose to use this enzyme as a model for further characterization. We interrogated the stereochemistry of SrtB substrate-binding using structural modeling with AlphaFold2 and molecular dynamics simulations, as well as mutagenesis. We also tested a variety of nucleophiles for baSrtB. We found that baSrtB requires amino acids outside of the pentapeptide NPKTG binding motif, specifically a P5 Asp and P2′ Asp for optimal activity. In addition, a chimeric baSrtB enzyme with the β7-β8 loop sequence from *L. monocytogenes* SrtB revealed an increase of ∼3-fold in relative activity, which we discovered was largely due to a single mutation, A241K. Incorporation of this residue into a A236E/A241K baSrtB double mutant increased activity to 4.4-fold higher than wild-type in a 1.8-h assay. Finally, we used a purified protein substrate with a biotinylated nucleophile to validate sortase-mediated ligation using SrtB enzymes. Overall, our work provides a number of insights into class B sortases and further highlights the potential of these enzymes as intriguing candidates for expanding the catalog of suitable SML substrates and applications.

Results

Activity assays of multiple SrtB enzymes

To better understand the relative activities of SrtB enzymes, we first identified a number of available crystal structures in the Protein Data Bank (PDB). These included SrtB enzymes from several infectious bacterial species: *B. anthracis* (baSrtB; PDB IDs 1RZ2, 2OQW) (49, 50), *C. difficile* SrtB (cdSrtB; 4UX7) (38), *Clostridium perfringens* SrtB (cpSrtB; 5B23, 5YFK) (51), *L. monocytogenes* (lmSrtB; 5JCV), *S. aureus* SrtB (saSrtB; 1NG5, 1QWZ, 1QX6, 1QXA, 4LFD) (28, 31), and *S. pyogenes* SrtB (spySrtB; 3PSQ) (36). We reasoned that available monomeric crystal structures, a few of which include either small molecules, a peptidomimetic, or a Gly3 bound, suggested that these proteins would be biochemically tractable to study. However, we chose to exclude additional studies of cpSrtB, as the substrate is not known and was previously shown not to be the heme-transport protein CPE0221 (51), and spySrtB, due to the additional requirement of the SipA protein cofactor (52). Therefore, we moved forward with recombinant expression and purification of the catalytic domains of baSrtB, cdSrtB, lmSrtB, and saSrtB, as described in the Experimental Procedures (Fig. S1*A*).

With purified proteins in hand, we initially determined relative activities using an established peptide-based assay. Here, putative substrate peptides were synthesized using solid-phase peptide synthesis, as described in the Experimental Procedures. Peptide sequences for each protein were as follows, with the pentapeptide recognition motifs underlined: Abz-D<u>NPKTG</u>DEK(Dnp)-NH_2 (from *B. anthracis* IsdC (baIsdC)) for baSrtB, Abz-KVE<u>NPQTN</u>AGK(Dnp)-NH_2 (from *S. aureus* IsdC (saIsdC)) for saSrtB, Abz-PV<u>PPKTG</u>DSK(Dnp)-NH_2 (from CD0386) for cdSrtB, and Abz-T<u>NPKSS</u>DSK(Dnp)-NH_2 (from Lmo2186) for lmSrtB, where Abz = 2-aminobenzoyl, Dnp = 2,4-dinitrophenol, K(Dnp) indicates a Dnp label on the ε-amine of Lys, and -NH_2 indicates a C-terminal amide. These sequences were chosen based on previous work suggesting that flanking residues beyond the pentapeptide recognition motif may be important for baSrtB recognition (37). When intact, the Dnp moiety quenches fluorescence from the Abz fluorophore *via* Förster (or fluorescence) resonance energy transfer (FRET); however, upon peptide cleavage by SrtB, the Abz fluorescence is apparent, as described in the Experimental Procedures and previously (16, 25, 27, 44, 53, 54). Activity assays revealed that baSrtB was the most active of the four enzymes, while cdSrtB and lmSrtB showed relatively little to no activity (Fig. 2). We subsequently tested Abz-K<u>NAKTN</u>DSK(Dnp)-NH_2 (from Lmo2185) activity in lmSrtB, which was also a negative result (Fig. S1*B*).

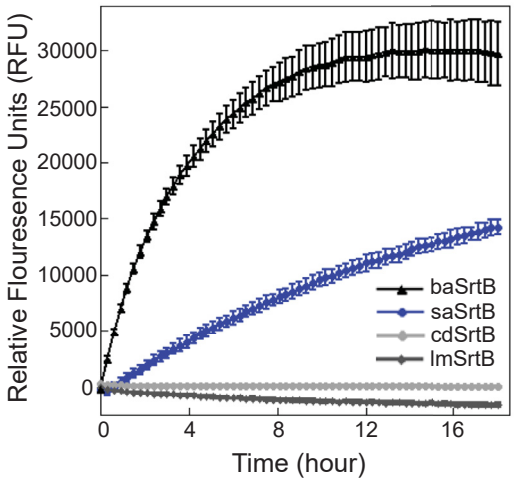

Figure 2. Relative activities of SrtB enzymes. A FRET-based reporter assay, with the general structure Abz-(SrtB recognition sequence)-K(Dnp), was used to monitor relative activity of SrtB enzymes from *Bacillus anthracis* (baSrtB), *Staphylococcus aureus* (saSrtB), *Clostridioides difficile* (cdSrtB), and *Listeria monocytogenes* (lmSrtB) organisms. Sequences tested were as follows: Abz-D<u>NPKTG</u>DEK(Dnp)-NH_2 (from *B. anthracis* IsdC (baIsdC)) for baSrtB, Abz-KVE<u>NPQTN</u>AGK(Dnp)-NH_2 (from *S. aureus* IsdC (saIsdC)) for saSrtB, Abz-PV<u>PPKTG</u>DSK(Dnp)-NH_2 (from CD0386) for cdSrtB, and Abz-T<u>NPKSS</u>DSK(Dnp)-NH_2 (from Lmo2186) for lmSrtB. Assays were run in triplicate technical replicates, and the standard deviations are shown.

Our initial assays were run using 50 μM SrtB and 175 to 200 μM of each respective peptide. This is reflective of the relatively low activities previously observed of SrtB enzymes; in contrast, many of our previous SrtA assays were run with 10 μM enzyme and 50 μM substrate (25, 27). Because of the long reaction time of these initial SrtB assays as well as the relatively high concentration of the fluorescent peptide, we discovered that we likely encountered *inner filter effects* in our assays, whereby fluorescence is suppressed by a high concentration of fluorophore (55). These effects were previously found to affect Abz fluorescence at concentrations as low as 50 μM and we found that the effect become more pronounced over the long timeframe (18 h) of these initial assays, likely due to evaporation of the reaction mixture (53). Therefore, we confirmed our relative activity results for baSrtB, saSrtB, and cdSrtB using an HPLC-based assay over 20 to 48 h (Fig. S1, *C–E*), and moving forward, we will restrict our reported fluorescence resonance energy transfer (FRET)-based assay results to 1 to 2 h, a timeframe which limits these effects. Notably, all our enzymes included a 6xHis tag (followed by a Tobacco Etch Virus (TEV) protease cleavage site), for use in protein purification. As a control experiment, we compared the relative activities of baSrtB with and without the 6xHis present (Fig. S2, *A* and *B*); interestingly, the cleaved enzyme revealed almost doubled relative activity at the 2-h timepoint. Moving forward, we chose to keep the 6xHis tag on all variants tested to maximize SrtB relative activity. As a final control, we also evaluated a baSrtB catalyzed reaction involving a peptide substrate lacking the Abz/Dnp labels (Ac-DNPKTGDE-NH_2). Based on LC-MS analysis (Fig. S1*F*), we observed that this substrate could be recognized by baSrtB and ligated to a model triglycine nucleophile, thereby demonstrating that the Abz/Dnp labels were not strictly required for enzyme activity.

Sequence-, pH-, and nucleophile selectivity of baSrtB

Based on the results of our initial activity assays, we chose to further investigate baSrtB as a model for this class of enzymes. We first wanted to validate previous reports that baSrtB was most active between pH 8.0 to 9.0, as well as that it recognizes amino acids beyond the NPKTG pentapeptide motif (32). We utilized our FRET-based activity assays to test baSrtB with the Abz-DNPKTGDEK(Dnp)-NH_2 peptide in the presence of reaction buffers that ranged in pH from 7.0 to 9.0 in 0.5 pH unit steps (Fig. 3*A*). In a 1.8 h assay, our results were similar but in slight contrast to the previously reported results; baSrtB was most active at pH 7.5 and 8.0, with pH 8.5 showing very similar results. We chose to use pH 8.0 for subsequent assays.

We next wanted to test the nucleophile-selectivity of baSrtB, as there is known diversity of the amine nucleophile that serves as the ligation partner for SrtB enzymes. Depending on the organism, this nucleophile may be an N-terminal pentaglycine strand, dialanine, or *meso*-diaminopimelic acid, amongst other possibilities (56–58). When a suitable nucleophile is present, then acyl-enzyme formation (the original cleavage reaction) is thought to be the rate-limiting step (59). However, in the absence of a suitable nucleophile, the apo-enzyme is restored through hydrolysis. In this case, hydrolysis is expected to become the rate-limiting step and the overall reaction rate is reduced (59). We therefore wondered if a poor nucleophile might be a contributing factor to low baSrtB *in vitro* activity. The endogenous nucleophile for baSrtB is thought to be *m*-diaminopimelic acid, though baSrtB may not recognize *m*-diaminopimelic acid *in vitro* (32).

We tested the following nucleophiles with baSrtB and the substrate Abz-DNPKTGDEK(Dnp)-NH_2: triglycine, Gly-NH_2, Ala-NH_2, D-Ala-NH_2, and D-Ala, as well as the strong nucleophile hydroxylamine (Fig. S2*C*). For these studies, the D-amino acid nucleophiles D-Ala-NH_2 and D-Ala were selected based on structural and chemical similarity to the presumed nucleophilic portion of *m*-diaminopimelic acid, and L-amino Ala-NH_2 was selected to investigate any stereochemical preference that might be displayed by the enzyme. Glycine nucleophiles were selected based on previously published experiments demonstrating the ability of baSrtB to catalyze ligation reactions to the nucleophile GGGK-Biotin (32). Hydroxylamine was previously shown to serve as a suitable nucleophile for several sortase homologs by ourselves and others; this was used as the nucleophile for the wild-type activity assays (Fig. 2) (4, 17, 25, 43). All reactions progressed at seemingly similar rates in a 1.6 h activity assay regardless of nucleophile, and were only slightly elevated above the activity levels seen in the hydrolysis control (Fig. 3*B*). Products formed were confirmed by HPLC-based assays using mass spectrometry (Table S1). Taken together, we concluded that our choice of nucleophile was not dramatically affecting the relatively weak baSrtB activity. In addition, our results showed that baSrtB can use L- and D-alanine as a nucleophile, a result not previously tested.

Finally, to investigate the extended sequence motifs for baSrtB, we evaluated the following peptide sequences (NPKTG pentapeptide is underlined, all contain Abz and K(Dnp-NH_2), which are omitted here for clarity): DNPKTGDE, NPKTGDE, DNPKTGD, NPKTGD, and NPKTGG (Fig. 3, *C* and *D*). Due to the nature of the FRET-based relationship between the fluorophore/quencher pair (Abz/Dnp) used in our assays, the background fluorescence was dependent on peptide length; therefore, we normalized to relevant control peptides (Fig. S2*B*). Here, we tested baSrtB enzyme with (Fig. 3*C*) and without (Fig. 3*D*) the 6xHis tag, as we reasoned the positively-charged 6xHis tag may contribute to activity because of these negatively-charged flanking residues in our target sequence. Our results confirm that baSrtB required the P5 Asp (DNPKTGDE), P2′ Asp (DNPKTGDE), and P3′ Glu (DNPKTGDE) for full activity; however, we also confirmed that the +6xHis baSrtB enzyme may be more sensitive to the presence of the P3′ Glu (DNPKTGDE). While the activity of both (±His) enzymes required all three additional residues for full activity, removal of the P3′ Glu did not affect relative -6xHis baSrtB activity for the NPTKGD peptide as compared to a peptide (NPKTGDE) with this amino acid included (Fig. 3, *C* and *D*). Furthermore, the relative fluorescence of +6xHis baSrtB with the DNPKTGD peptide at the 2 h timepoint was

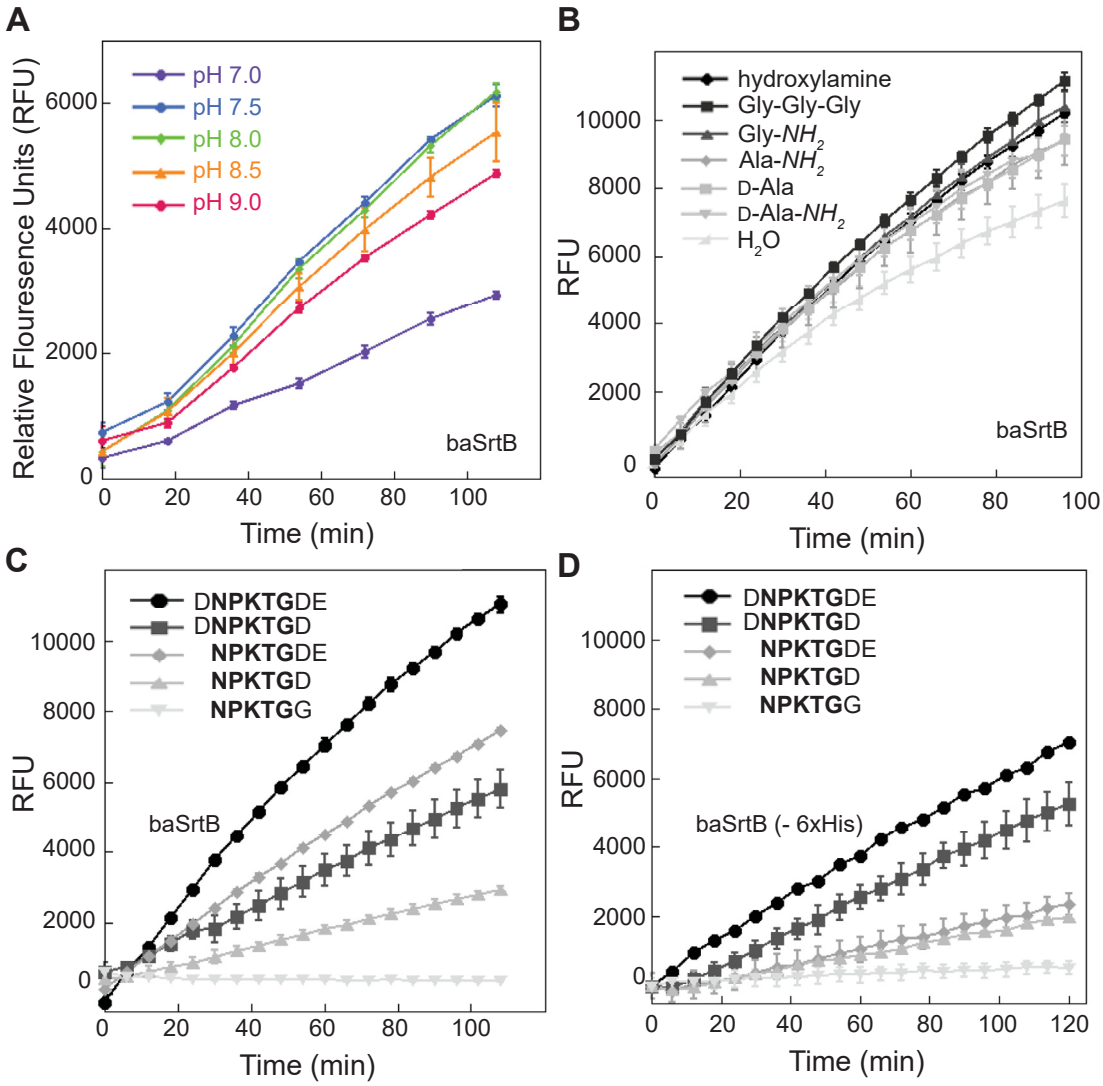

Figure 3. Additional variables that affect relative catalytic activity of baSrtB. Having established baSrtB as the most active in this assay, additional experiments characterized the pH dependence (*A*), preferred nucleophile (*B*), and position-specific recognition of substrate (*C* and *D*). Our investigation into the effects of varying peptide lengths on baSrtB activity included the enzyme with (*C*) and without (*D*) a 6xHis tag included. Wild-type saSrtB and pH-dependence assays were performed in duplicate and all others were run in at least triplicate technical replicates; all curves represented are averages measured in relative fluorescence units (RFU) following subtraction of relative negative control (minus SrtB) background fluorescence (Fig. S2*B*), and including the standard deviation of replicate experiments.

equivalent to -6xHis baSrtB with the same peptide, 5346 ± 37 *versus* 5274 ± 620, respectively, minimizing the observed His-tag effect (Fig. 3, *C* and *D*). For both enzymes (±6xHis), we saw the largest reductions in activity with the loss of the P5 Asp and P2′ Asp. Notably, the loss of all three residues failed to show any activity in either assay (Fig. 3, *C* and *D*).

Stereochemistry of substrate binding to SrtB enzymes

To better understand the results described above and to characterize baSrtB, we wanted to investigate the stereochemistry of substrate binding. We first attempted to use X-ray crystallography to determine a complex structure of catalytically inactive C233A baSrtB protein, expressed and purified as described in the Experimental Procedures, bound to a peptide matching the DNPKTGDE sequence. Using previously determined crystallization conditions, as well as positive hits from a number of conditions in commercially available crystallization screens, we were able to generate quality crystals that diffracted with resolution <2 Å (Fig. S3*A*). However, data processing, molecular replacement using the apo baSrtB as a model (PDB ID 2OQW), and initial rounds of refinement confirmed that the peptide was not present in our crystal lattice. This was true for crystals grown under multiple salt and precipitant conditions. We chose not to finish refinement of the apo structure, as it was very similar to already available data (PDB ID 2OQW) (50).

We therefore decided to model substrate binding to baSrtB using Alphafold2 (60). As an input, we used a single sequence matching the full-length *B. anthracis* IsdC, followed by a linker $(G_4S)_2$, then the full-length baSrtB without the initiator Met (see Supporting Information for full sequence used). Although

there are no substrate-bound structures of baSrtB in the PDB, we reasoned that the substrate may be situated into the peptide-binding cleft in the predicted folded structure, based on other known sortase structures. Indeed, the output AlphaFold2 models with the five highest confidence scores position the NPKTG recognition sequence in the active site of baSrtB (Fig. S3*B*). The baIsdC protein contains a long, disordered linker, approximately residue N148 to R213, the residue immediately preceding the transmembrane helix; however, the location of baIsdC in the 5 models does not substantially vary, with the exception of the model ranked_3, although it is not fixed in one location (Fig. S3*B*). Notably, baIsdC is processed by a signal peptidase during membrane translocation, therefore, we also used AlphaFold2 to model this complex using the baIsdC sequence starting at A30. This resulted in an equivalent model for the full-length baSrtB protein (RMSD $<$ 0.2 Å), therefore, we will exclude these residues moving forward in order to better represent the mature folded protein (33). Alignment of the main chain atoms of the recognition motif sequence (DNPKTGDE), baIsdC transmembrane helix, and baSrtB revealed RMSD values of $<$ 0.6 Å for all models aligned to the model labeled as ranked_0. We will use ranked_0 as our structural model here due to this high degree of similarity (Fig. 4*A*). Subsequently, upon its release and during the preparation of this work, we used AlphaFold3 to model the baSrtB-DNPKTGDE complex (61). AlphaFold3 generated very similar models to AlphaFold2 (RMSD = 0.268 Å over 732 main chain atoms) (Fig. S3, *C* and *D*); thus, we are confident with our AlphaFold2 modeling approach and results presented below.

Our AlphaFold2 model of baIsdC-baSrtB allowed us to make a number of predictions about this binding interaction. Although our analyses will be primarily focused on DNPKTGDE binding to baSrtB, we also observed three potential interactions worth further analyses: (1) the folded domain of baIsdC may be forming intramolecular contacts with DNPKTGDE or (2) intermolecular electrostatic contacts with SrtB, and (3) the transmembrane helices may dimerize in the membrane (Fig. 4*B*). Any of these interactions could enhance reaction rates and may have a substantial effect on endogenous activity, which has not been measured to our knowledge. With respect to DNPKTGDE recognition by the catalytic domain of baSrtB, the peptide is well-positioned in the active site, in a conformation that we predict is competent for activity (Fig. 4*C*). For example, the thiol group of C233 is 3.6 Å from the carbonyl carbon of the P1 Thr in the ranked_0 model, a compatible distance for nucleophilic attack by this rotamer or others. In the other four models, the distances range from 3.5 to 3.8 Å (Fig. 4*C*). The other catalytic residues, H140 and R243 are in a consistent position with respect to the peptide as compared to a previously reported spySrtA-LPATA structure (PDB ID 7S51) (Fig. 4*D*) (17).

Investigating specific contacts at peptide positions revealed differences compared to a published saSrtB-NPQT* structure (PDB ID 4LFD) (28). Main chain atoms (641 total) from the A-protomer from saSrtB-NPQT* aligned with an overall RMSD of 0.67 Å (Fig. S4*A*). As mentioned above, the P4 Asn in PDB ID 4LFD is forming a single hydrogen bond with the backbone carbonyl oxygen of T177 in saSrtB (Fig. S4*B*). In contrast, our model revealed specific interactions with three baSrtB residues, the side chain hydroxyl of Y191 in the β6-β7 loop, the side chain hydroxyl of S231 in the β7 strand, and the side chain guanidino group of the catalytic R243, a characteristic consistent with previous observations by ourselves and others that the P4 residue stabilizes the ligand in SrtA enzymes (Fig. 5*A*) (15, 17, 20). We also see differences in the position of the P3 Pro; whereas, in saSrtB-NPQT*, this residue is primarily stabilized by I182, in our model of baSrtB-DNPKTGDE, there are hydrophobic interactions observed with side chain atoms of L106, F121, and Y138 (Figs. 5*B*, S4*C*). Notably, all four of these SrtB residues (Ile, Leu, Phe, and Tyr) are structurally conserved in the SrtB sequences we analyzed.

Additional residues in the binding motif are challenging to compare directly, as the catalytic Cys thiol groups are shifted 6 Å with respect to each other, despite otherwise good alignment in the sortase domains. However, in our baSrtB-DNPKTGDE model, we see electrostatic interactions between the P2 Lys and E240 in the β7-β8 loop (Fig. 5*C*), the P1 Thr positioned similarly to that of our spySrtA-LPATA structure and stabilized by the Cys amide (Fig. 5*D*), and interactions with the P2′ Asp and side chain atoms of Y39 in the N-terminal helix (unique to SrtB enzymes) (Fig. 5*E*) and K143 in the β4-β5 loop. We also observed similar stabilizing backbone interactions between the guanidino group of the catalytic R243 residue and the carbonyl oxygens of the P3 Pro and P2 Lys (Fig. 5*F*), consistent with the spySrtA-LPATA complex structure (17).

We did not observe any specific interactions with baSrtB which indicated why the P5 Asp nor P3′ Glu in the DNPKTGDE (positions underlined) sequence had an effect on activity (Fig. 2*C*). We initially reasoned that this may be due to the nature of the AlphaFold2 model generated, namely the presence of the full-length baIsdC protein, of which both residues may be at the interface for potential intramolecular interactions in baIsdC. Importantly, we did not include the 6xHis tag in our AlphaFold2 models (as a result of using the full-length sequence of baSrtB); however, while an AlphaFold3 model of 6xHis-baSrtB with the DNPKTGDE ligand confirms that the P3′ Glu is in the vicinity of the 6xHis tag, these residues are not interacting in any of the five output models (Fig. S4, *D* and *E*). Overall, we wanted to investigate the stereochemistry of these positions further, as well as the stability of the entire recognition motif. Therefore, we ran triplicate molecular dynamics (MD) simulations of the peptide fragment with baSrtB. In addition, we ran a triplicate MD simulation with baSrtB and a control poly-Ala peptide, which confirmed that this nonspecific ligand is not stable in the binding pocket (Fig. S5).

Our MD results confirmed that the baSrtB protein and DNPTKTGDE peptide are relatively stable over the course of the three 1000 ns simulations, visualized as the average root mean squared deviation (RMSD) of Cα atoms over time (Figs. 6*A*, S6*A*). For peptide amino acids, there is a clear difference in relative flexibility for the pentapeptide motif,

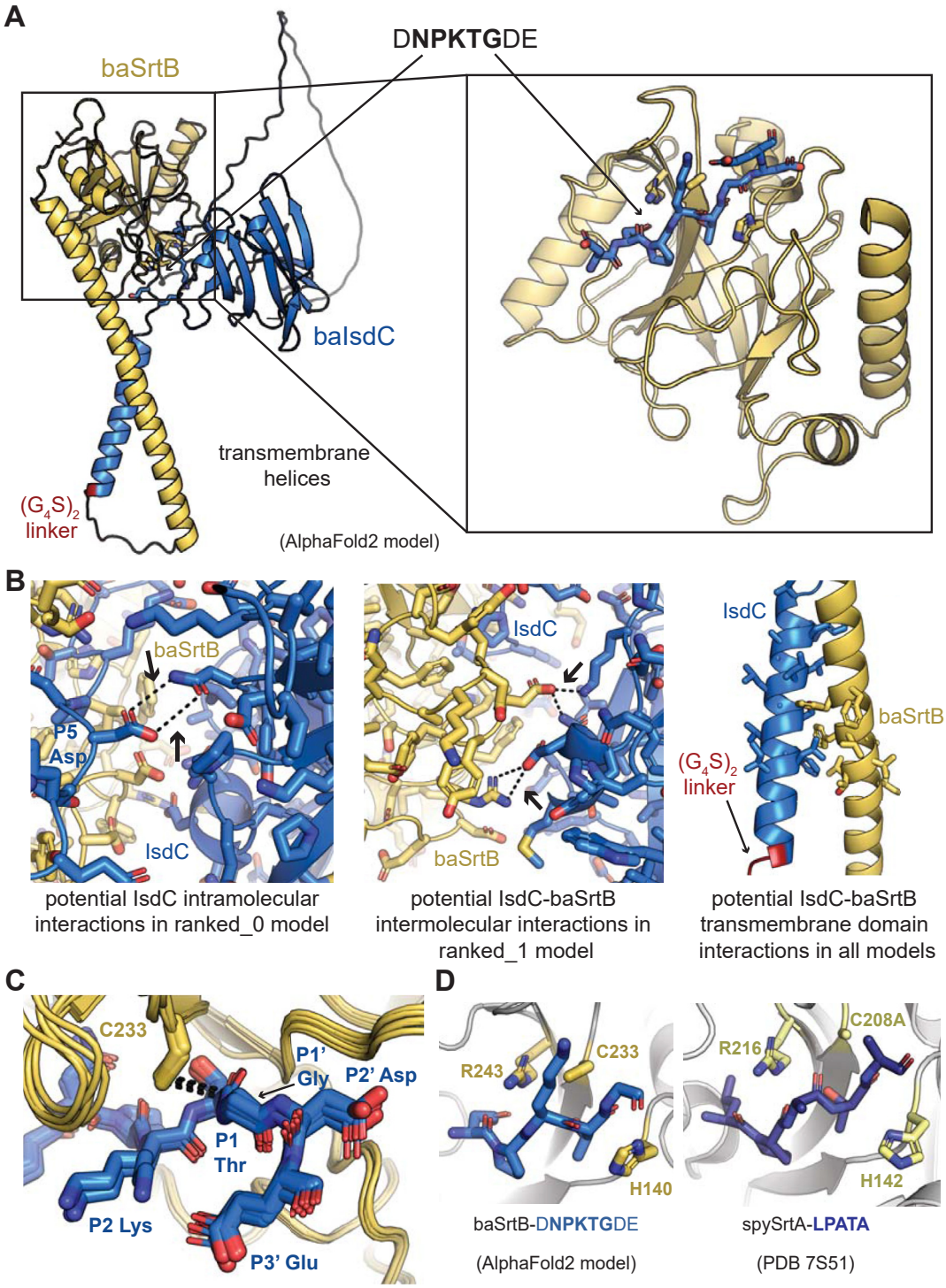

Figure 4. An AlphaFold2 model of baSrtB-substrate binding. *A*, an AlphaFold2 model of baIsdC binding baSrtB was generated using the input of a single polypeptide sequence containing both full-length proteins, and a $(G_4S)_2$ linker (labeled). The zoomed-in catalytic domain is also shown of baSrtB bound to the DNPKTGDE recognition sequence from baIsdC. *B*, the 5 AlphaFold2 output models (ranked_0–4) indicated a number of potential interactions which may impact baSrtB activity, as highlighted here and described in the main text. *C*, all 5 AlphaFold2 models show the recognition sequence from baIsdC in a catalytically-competent conformation with respect to baSrtB. *D*, BaSrtB-substrate binding is very similar to an experimentally-determined structure of *Streptococcus pyogenes* SrtA (spySrtA) bound to the LPATA peptide (PDB ID 7S51), with respect to orientation of substrate and location of catalytic residues (17). For all, protein structures are shown in cartoon representation with side chains modeled as sticks, where applicable, and colored by atom (N = *blue*, O = *red*, S = *yellow*, C = as labeled by protein). BaIsdC, or its substrate sequence, DNPKTGDE, is in *blue*, baSrtB is in *golden yellow*, and in (*D*), spySrtA is in *yellow-green*, and its peptide substrate, LPATA, in *purple-blue*.

NPKTG, as compared to the P5 Asp and P2′/P3′ D/E residues (Fig. S6*B*). While interacting residues for the pentapeptide motif agree with those discussed above, based on the ranked_0 model alone, analyses of electrostatic interactions over the simulations suggests that the P5 Asp most commonly interacts with: N102, Y124, R125, and Y190, the P2′ Asp with: Y39, R116, H140 (catalytic His), K143, and Y235, and the P3′ Glu with: Y39, R116, R141, K143, and K152. There appear to be two dominant positively-charged "pockets" for P2′ and P3′ binding, formed by R116/K143 and R141/K152, respectively. An electrostatic potential surface map of baSrtB confirms that the S2′ and S3′ binding sockets form the most-positively charged surface, consistent with the observed preference for addition of the P2′/P3′ D/E substrate residues (Figs. 3*D* and 5*B*). While the P3′ Glu appears to also interact with these binding pockets, our biochemical data suggests that interactions with the 6xHis tag may also contribute to an additional preference for a P3′ Glu in the +6xHis baSrtB enzyme (Fig. 3, *C* and *D*).

Although relatively inactive in our hands, we also wanted to investigate the stereochemistry of ligand binding to the saSrtB and lmSrtB proteins. Therefore, we created substrate-bound models using AlphaFold2 for saSrtB-VENPQTNAG, lmSrtB-EKNAKTNDS, and lmSrtB-VTNPKSSDS using a similar protocol as that with baSrtB. Specifically, we modeled single polypeptide sequences of saIsdC-$(G_4S)_2$-saSrtB, Lmo2185-$(G_4S)_2$-lmSrtB, and Lmo2186-$(G_4S)_2$-lmSrtB (Fig. 6*C*). The structural models containing saIsdC were treated equivalently as those with baIsdC and again, models of the full-length saSrtB with or without the saIsdC signal sequence included were equivalent (RMSD < 0.1 Å); here, the mature protein is predicted to start at A29 (33). We also attempted to use this approach for the cdSrtB protein, but the peptide was not properly docked near the active site. All input sequences used in AlphaFold2, including for IsdC-baSrtB, are in the Supporting Information. We also extracted peptide-bound models of these complexes and ran 1000 ns triplicate MD simulations. As with baSrtB, ligand-bound saSrtB and lmSrtB are stable over the course of our trajectories with a well-positioned pentapeptide motif, consistent with the clear non-covalent interactions that stabilize peptide binding in our baSrtB-DNPKTGDE model (Figs. 6*C*, S6, *C*–*E*).

This result was particularly interesting for saSrtB, as there is an available structure bound to a peptide mimetic, NPQT* where T* = (2*R*,3*S*)-3-amino-4-mercapto-2-butanol (PDB ID 4LFD), as mentioned above (28). Whereas none of the four protomers of the crystal structure show specific binding of the P4 Asn to saSrtB residues, our model suggests that the P4 Asn may interact with the side chain atoms of Y181 and S221, analogous residues as those predicted to stabilize P4 Asn binding to baSrtB (Y191 and S231, respectively) (Fig. 6*D*). Overall, peptide binding is quite different in PDB ID 4LFD *versus* our AlphaFold2 model (Fig. 6*E*). Notably, the NE2 atom of the catalytic His (H130) side chain is 8.9 Å away from the P1 Thr carbon that would be the site of initial nucleophilic attack in the experimental structure (PDB ID 4LFD). Because ligand binding to saSrtB in our model was consistent with that for baSrtB, we wanted to use mutagenesis to further investigate these observations and directly assay the validity of our baSrtB AlphaFold2 and MD results.

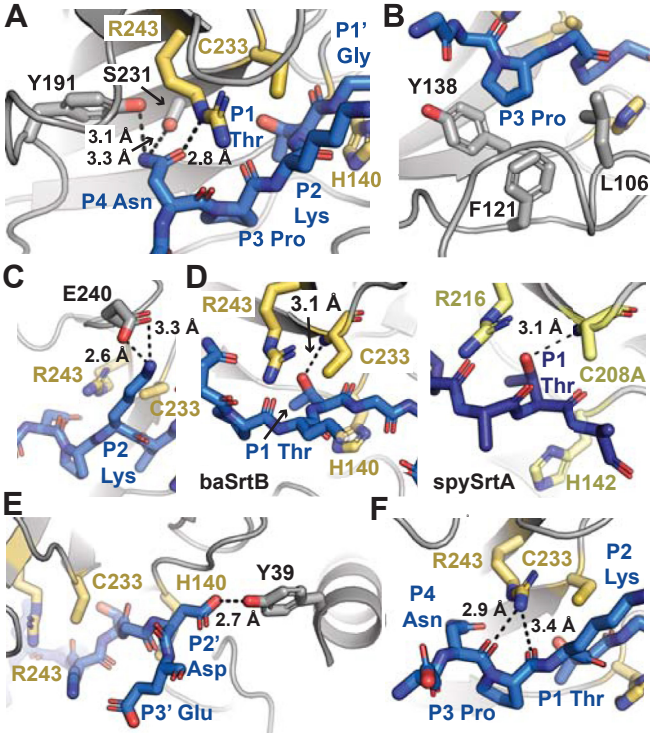

Figure 5. Position-specific recognition of DNPKTGDE peptide substrate by baSrtB. Position-specific recognition of the peptide substrate by baSrtB is highlighted for the (*A*) P4 Asn (DNPKTGDE), (*B*) P3 Pro (DNPKTGDE), (*C*) P2 Lys (DNPKTGDE), (*D*) P1 Thr (DNPKTGDE), and (*E*) P2′ Asp (DNPKTGDE) residues. Backbone interactions between the catalytic Arg and peptide are in (*F*). For all, baSrtB is in *gray* cartoon, with relevant side chain atoms shown as sticks and colored by atom (N = *blue*, O = *red*, S = *yellow*, C = as labeled by protein). The catalytic residues of baSrtB, H140, C233, and R243, are in *golden yellow*. The peptide substrate is in *blue* and colored by atom. In (*D*), spySrtA (PDB 7S51) is in *yellow-green*, and its peptide substrate, LPATA, in *purple-blue* for comparison with baSrtB. Measurements are in dashed *black* lines, with distances labeled.

Mutagenesis supports the model of ligand binding to baSrtB

To biochemically interrogate our substrate-bound model of baSrtB, we tested 15 mutations and truncations. These included: N-terminal truncations aimed to investigate the role of the additional α-helix in SrtB enzymes, as compared to SrtA ($baSrtB_{42\text{-}254}$ and $baSrtB_{65\text{-}254}$), N-terminal α-helix point mutations (D38A, Y39A, Y39F), mutations in catalytic residues (H140A, R243A), and several hypothesized to interact with ligand residues (L106A, R116A, Y191A, Y191F, S231A, Y235A, and E240A) (Fig. 7*A*). All proteins were recombinantly expressed, purified, and tested as previously described and in the Experimental Procedures.

Our mutagenesis data agreed with the baSrtB-DNPKTGDE model. We saw a total loss in activity in the catalytic residue mutations, those predicted to interact with the P4 Asn, Y191 and S231, and L106A, a residue that appears to interact with and stabilize the P3 Pro (Figs. 5*B*, 7, *B* and *C*). In addition, both N-terminal α-helix truncations dramatically reduced activity

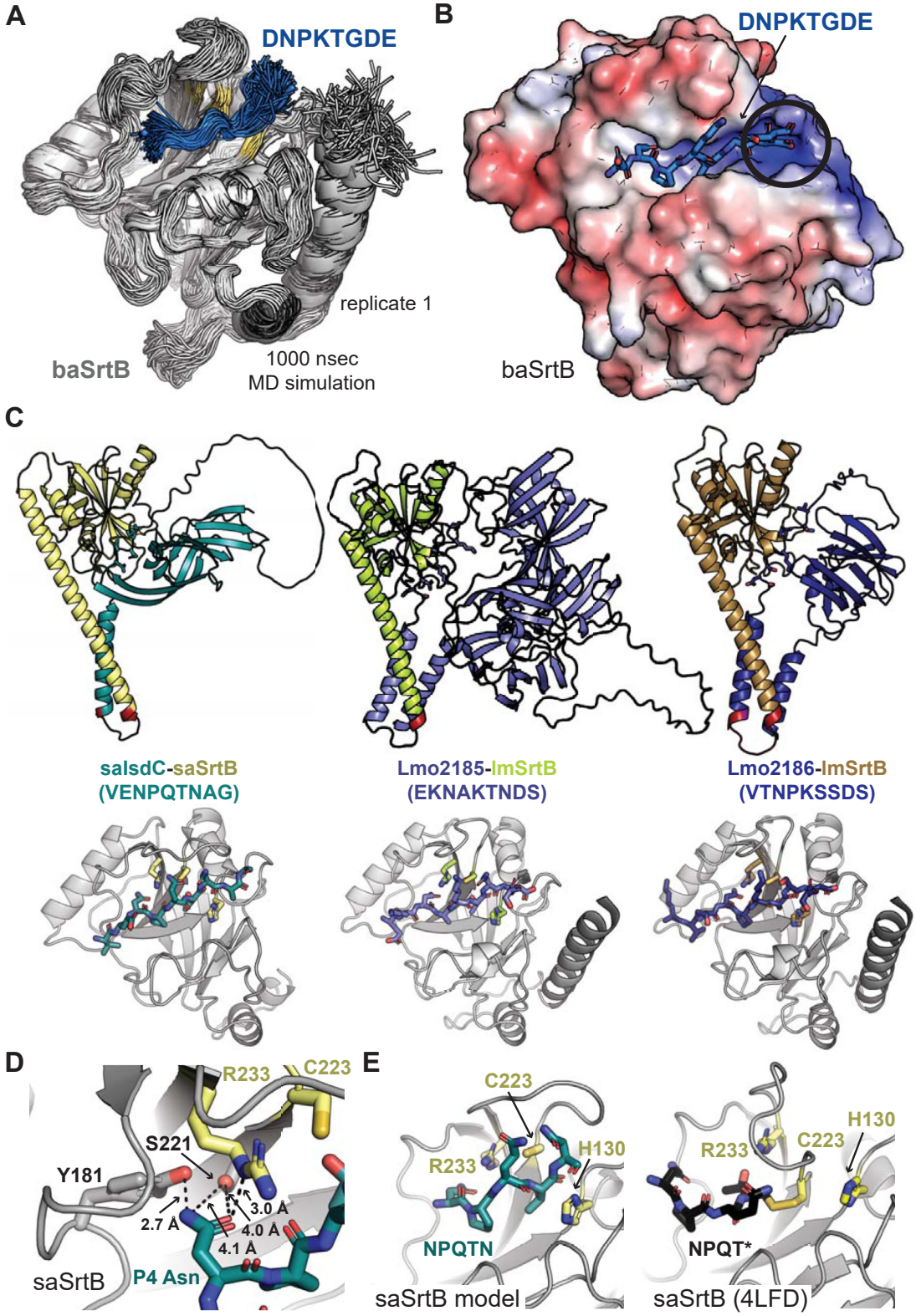

Figure 6. Molecular dynamics simulation of baSrtB-substrate, and AlphaFold2 models of other SrtB enzymes. *A*, alignment of 200 states (every 5 nsec of simulation time) during one replicate of our baSrtB-DNPKTGDE molecular dynamics simulation. The baSrtB enzyme is in *gray* cartoon, with catalytic residues highlighted in *golden yellow*, and the peptide substrate is in the *blue* cartoon. *B*, electrostatic potential surface map (±5 eV, on a *red* (negative) to *blue* (positive) scale) of the baSrtB catalytic domain with the DNPKTGDE peptide shown as *blue* sticks and colored by atom (N = *blue*, O = *red*). A positively-charged pocket where we predict the P2′ Asp and P3′ Glu residues interact is circled. *C*, alphaFold2 models of salsdC-saSrtB (*left*), Lmo2185-lmSrtB (*middle*), and Lmo2186-lmSrtB (*right*) are shown in cartoon representation and colored as labeled. The zoomed in catalytic domains bound to peptide substrate (sequences in parentheses) are below each model in cartoon representation, with the catalytic residue side chains and peptide substrates as sticks and colored by atom. *D*, predicted interactions between the P4 Asn in the saSrtB substrate, NPQTN, and enzyme. Distances are shown as black dashed lines and labeled. *E*, comparison of the saSrtB model generated here (*left*) and NPQT* peptidomimetic binding in the experimentally-determined structure, PDB ID 4LFD (28). Structures are rendered as elsewhere in the figure.

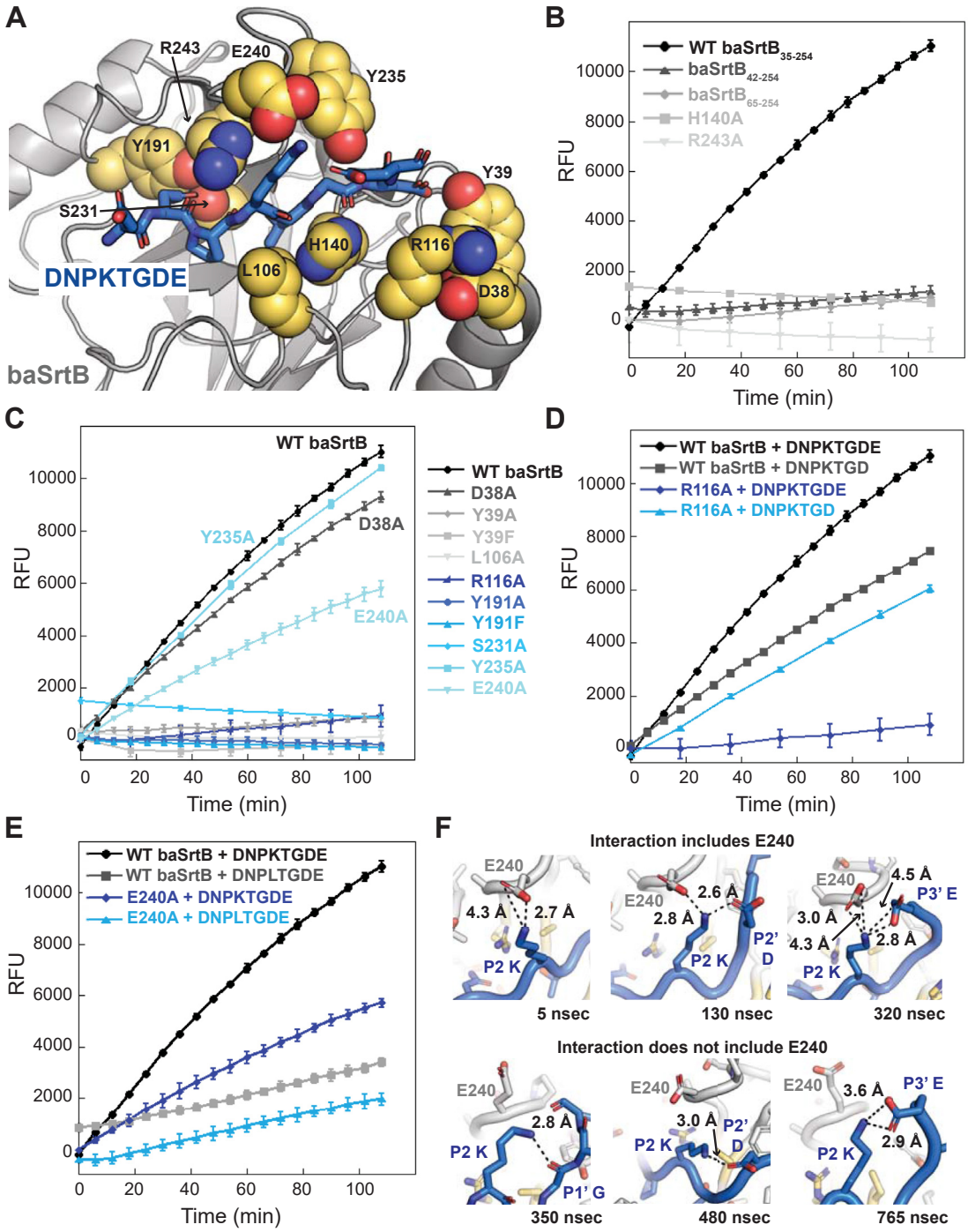

Figure 7. Mutagenesis of baSrtB confirms the validity of the AlphaFold2 model. *A*, the baSrtB enzyme is in *gray* cartoon and DNPKTGDE peptide substrate in *blue* sticks and colored by atom (N = *blue*, O = *red*). Mutations tested in this work are shown as *golden yellow* spheres, colored by atom, and labeled. *B* and *C*, FRET-based activity assays for baSrtB variants are shown as the averages of at least triplicate technical replicates, with standard deviation. The N-terminal helix truncations and catalytic residue mutations, H140A and R243A, are in (*B*) and all other mutations in (*C*–*E*). The wild-type data is in all graphs for comparison. *D*, comparison of relative activity levels between WT baSrtB and R116A baSrtB with the DNPKTGDE and DNPKTGD peptides, as labeled. Activity is mostly restored in the R116A baSrtB variant when the peptide lacks a P3′ Glu. *E*, comparison of relative activity levels between WT baSrtB and E240A with the DNPKTGDE and DNPLTGDE peptides, as labeled. *F*, representative states from one of the replicate MD simulations of WT baSrtB with DNPKTGDE reveals interactions between the peptide P2 Lys and E240 or intrapeptide atoms. Structures are colored as in (*A*) with the peptide main chain shown as cartoon and side chain atoms shown as sticks in all cases but one (*bottom left*), where the peptide main chain is also in stick representation.

and point mutations suggested this is predominantly due to interactions with a single residue, Y39 (Fig. 7, *B* and *C*). Both Y39A and Y39F baSrtB proteins revealed similar levels of activity, consistent with an interaction between the Tyr side chain hydroxyl group and the P2′ Asp residue that is suggested by our model (Figs. 5*E* and 7*C*). These findings suggest that the extended N-terminal α-helix is important for baSrtB catalysis. The activity of the D38A and Y235A proteins were relatively

similar to wild-type (Fig. 7*C*). The R116A baSrtB protein revealed a decrease in activity consistent with our structural observations of a positive P2′, P3′ binding pocket, suggesting this residue may additionally interact with the P3′ Glu-interacting residue (Figs. 6*B* and 7*D*). Consistent with this, the activity of R116A baSrtB (containing the 6xHis tag) with the DNPKTGD peptide, which lacks the P3′ Glu, restored activity to near wild-type levels for this peptide (Fig. 7*D*). We did not explicitly test R141, K143, and K152, which we predicted may also contribute to this positive binding pocket for the P2′ Asp residue.

In addition, the E240A baSrtB protein showed a loss of activity as compared to wild type baSrtB (Fig. 7*E*). Our modeling suggests the E240 residue forms electrostatic interactions with the P2 Lys (Fig. 5*C*). Consistent with this, we see a marked reduction in activity for wild-type baSrtB, and similar levels of activity between wild-type baSrtB and E240A, with the substrate peptide DNPLTGDE, which substitutes a Leu for the P2 Lys (Fig. 7*E*). However, E240A activity is also reduced with the DNPLTGDE substrate compared to DNPKTGDE activity, indicating that the P2 Lys may form additional interactions that facilitate activity (Fig. 7*E*). Indeed, we visualized several states in one of our representative MD simulations which contained intrapeptide interactions of the P2 Lys with the main chain carbonyl of the P1′ Gly, or side chain atoms of P2′ Asp and P3′ Glu residues (Fig. 7*F*). These interactions occurred both in the presence and absence of E240 (Fig. 7*F*). Taken together, these results support our substrate-bound baSrtB structure as a good model of the stereochemistry of this complex, specifically for the core NPKTG residues; furthermore, we hypothesize our AlphaFold2 models represent the likely mode of ligand binding for other SrtB enzymes, *e.g.*, saSrtB and lmSrtB.

Chimeric proteins reveal interactions that dramatically improve baSrtB activity

Previously, we studied the β7-β8 loop of SrtA enzymes in great detail, discovering many interactions that affect activity and specificity (25, 27). We were therefore curious if residues in this loop have a similar effect on SrtB enzymes. To this end, we created a number of chimeric enzymes, with the β7-β8 loop sequences of saSrtB, lmSrtB, and cdSrtB engineered into baSrtB. We recombinantly expressed, purified, and tested these variants as previously described, and as in the Experimental Procedures. We will refer to these proteins as: baSrtB$_{aureus}$ (β7-β8 loop sequence: C**ED**A**YSETTK**R), baSrtB$_{monocytogenes}$ (C**DTEK**D**YEK**GR), and baSrtB$_{difficile}$ (C**TYEF**D–**DA**R), where the catalytic Cys and Arg residues are underlined, bold indicates mutations as compared to baSrtB, and dashes indicate changes in loop length. For reference, the baSrtB β7-β8 loop sequence is: CDYALDPEAGR.

All three chimeric enzymes were active, although to varying degrees (Fig. 8*A*). Whereas the baSrtB$_{aureus}$ and baSrtB$_{dificile}$ chimeras revealed reduced activity as compared to wild-type baSrtB, we saw a dramatic increase with baSrtB$_{monocytogenes}$; here, the t = 1.8 h fluorescence was 2.8-fold higher in the β7-β8 loop variant as compared to wild-type baSrtB (Fig. 8*A*). We made single point mutations to interrogate this result further, changing one baSrtB β7-β8 loop residue (sequence: D**YALD**P**EA**G) at a time to the corresponding amino acid in the lmSrtB loop (sequence: D**TEKD**Y**EK**G, differences underlined): Y235T, A236E, L237K, P239Y, and A241K (note: we are using baSrtB numbering because of the consistency in loop lengths between these enzymes) (Fig. 8*B*). All proteins were recombinantly expressed, purified, and tested as previously reported and as in the Experimental Procedures. Of these mutant proteins, P239Y baSrtB and Y235T baSrtB had relatively equivalent activity, and A236E, L237K, and A241K baSrtB revealed increased activity over wild-type baSrtB (Fig. 8*B*). Specifically, the increase in activity of A241K baSrtB recapitulated, and even surpassed (3.6-fold over wild-type at t = 1.8 h), that seen in baSrtB$_{monocytogenes}$ (Fig. 8*B*). A double mutant, A236E/A241K baSrtB, increased this effect further, with a 4.4-fold increase in activity as compared to wild-type at t = 1.8 h (Fig. 8*B*).

AlphaFold2 models generated of baIsdC-$(G_4S)_2$-A241K and baIsdC-$(G_4S)_2$-A236EA241K baSrtB indicated that the observed effect may be due to stabilizing interactions between A236E with R131 and/or A241K with D188, respectively (Fig. 8*C*). While the β7-β8 loop backbone atoms of the three proteins (wild-type, A241K, and A236E/A241K) are very similar, we predict that these stabilizing interactions may affect conformational dynamics during catalysis. We expressed, purified, and tested the double mutant D188A/A241K baSrtB, which revealed ∼2-fold decreased activity as compared to A241K baSrtB alone, although it is still 2.1-fold higher than wild-type (Fig. 8*B*). Taken together, our results revealed single (A241K) and double (A236E/A241K) mutations that have a large effect on baSrtB activity.

Sortase-mediated ligation (SML) using baSrtB mutants

Finally, we wanted to explore baSrtB for use in a sortase-mediated ligation (SML) application. To this end, we designed a fluorescent substrate protein consisting of the cyan fluorescent protein variant mTurq linked to a DNPKTGDEGGGG motif at its C-terminus (Fig. 9). This was paired with a synthetic triglycine nucleophile decorated with a biotin affinity tag (GGG-*biotin*). In the presence of A236E/A241K baSrtB, we observed formation of the expected C-terminal ligation product by LC-MS (Fig. 9). As is typical for sortase ligations, an excess of the GGG-*biotin* was utilized to drive the reaction toward the desired modified protein product (62). After a 21 h incubation at room temperature, ∼47% of the mTurq-DNPKTGDEGGGG substrate was converted to the biotinylated product as estimated by LC-MS. We utilized the 6xHis version of this enzyme, based on our previous results that inclusion of the 6xHis tag boosts overall activity of baSrtB (Fig. 3, *C* and *D*). We reasoned that this experimental boost may enhance the usefulness of baSrtB as an SML tool.

Interestingly, mass spectrometry also revealed low levels of a higher molecular weight species with a molecular weight formed after the initial reaction of A236E/A241K baSrtB and

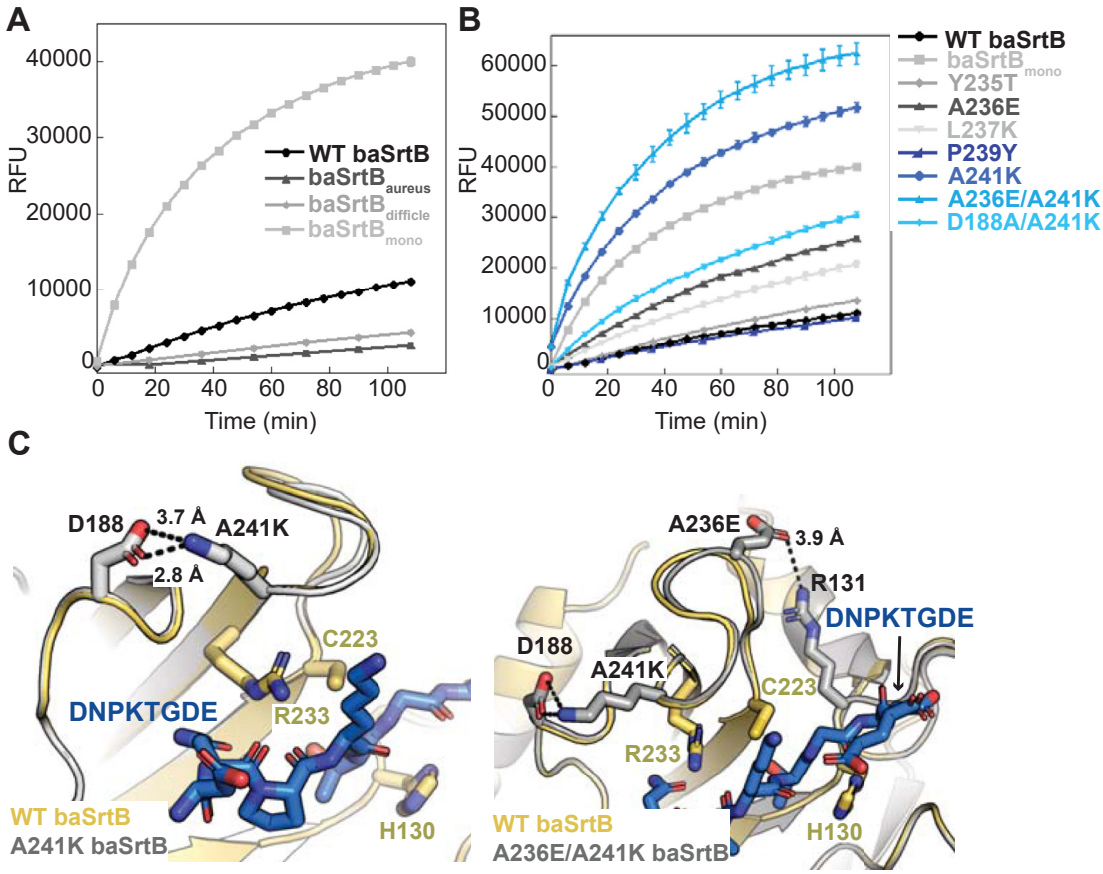

Figure 8. Loop-swapped chimeras elucidate baSrtB mutants with higher relative activities. *A*, BaSrtB chimeras were created by inserting the β7-β8 loop sequences from saSrtB ($baSrtB_{aureus}$), cdSrtB ($baSrtB_{difficile}$), and lmSrtB ($baSrtB_{mono}$), and FRET-based activity assays were run in at least triplicate technical replicates. The averaged values are shown, along with the standard deviation. *B*, single mutations were created in baSrtB to replicate amino acids in the β7-β8 loop of lmSrtB. The results of averaged replicate experiments are shown, with standard deviation. The wild-type baSrtB and $baSrtB_{mono}$ data is also shown for comparison. Double mutants were also tested to assay additive effects (A236E/A241K) or to investigate the basis of the activity difference (D188A/A241K). *C*, hypothesized interactions that increase the relative activity of A241K and A236E/A241K baSrtB as compared to wild-type are shown using AlphaFold2 models created of the variant proteins. The wild-type baSrtB protein is in *golden* cartoon, with the A241K (*left*) or A236E/A241K (*right*) baSrtB proteins in *gray* cartoon, as labeled. Relevant side chain sticks are shown as *gray* sticks and colored by atom (N = *blue*, O = *red*). The DNPKTGDE peptide substrate is in *blue* sticks and colored by atom. Measurements are shown as *black* dashed lines, and distances labeled.

the mTurq-DNKPTGDEGGGG substrate (Fig. S8*A*). This product was also evident by SDS-PAGE, showing a slow accumulation over the course of the reaction (Fig. S8*B*), and was also observed in control reactions lacking GGG-*biotin* (Fig. S8*C*). While we initially speculated that this higher molecular weight adduct corresponded to the thioester-linked intermediate, subsequent experiments using wild-type baSrtB and an Abz-DNPKTGDEK(Dnp)-NH_2 peptide revealed that the substrate was likely adding to the N-terminus of the enzyme (Fig. S9). This was supported by the fact that the observed enzyme-substrate adducts were susceptible to alkylation with N-ethylmaleimide (Fig. S9*A*), suggesting the presence of a free catalytic cysteine residue. In addition, removal of the N-terminus in enzyme-substrate adducts using an existing TEV cleavage site generated fragments that were consistent with N-terminal modification (Fig. S9*B*). Notably, this phenomenon is not unique to SrtB, and acylation at residues other than the catalytic cysteine has been suggested previously in the case of *S. aureus* sortase A (63).

Overall, these findings demonstrate that A236E/A241K baSrtB is a viable enzyme for use in SML applications. While this approach does not currently match the efficiency of SML using sortase A variants (64), it does offer the advantage of a ligation site motif (DNPKTGDE) that is unique from the canonical LPXTG sequence. In addition, further optimization of ligation conditions and continued engineering of the baSrtB enzyme itself are likely to improve ligation yields, and have the potential to make baSrt variants a valuable addition to the SML toolkit.

Discussion

Despite 25 years of research into bacterial sortases, class A sortases remain the best-studied of this enzyme family. Of this work, the vast majority investigates the characteristics, mechanism, and SML utilization of *S. aureus* SrtA, which arguably contains some unique traits, *e.g.*, dependence on allosteric activation by calcium and strict specificity for the LPXTG binding motif (1, 4, 23, 43–45). While *S. aureus* SrtA has relatively high activity with respect to other studied sortases, previous directed evolution experiments revealed that a small number of mutations, *e.g.*, a pentamutant, can have dramatic

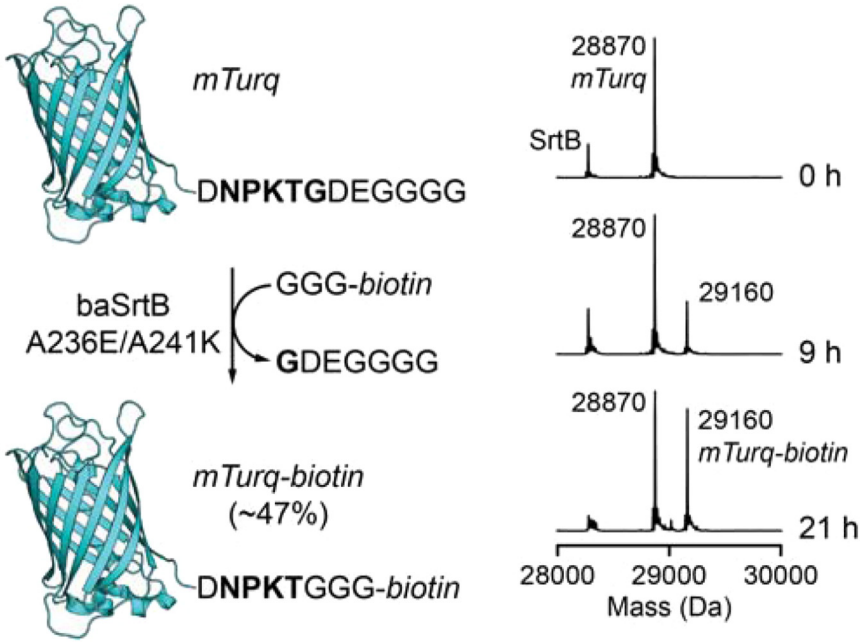

Figure 9. Sortase-mediated ligation catalyzed by mutant baSrtB. A variant of mTurq containing a C-terminal DNPKTGDEGGGG extension (*mTurq*, 200 μM) was ligated to a biotinylated triglycine nucleophile (GGG-*biotin*, 1 mM) in the presence of 50 μM A236E/A241K baSrtB at room temperature. LC-MS analysis (*right*) revealed formation of the desired biotinylated product (*mTurq-biotin*), reaching ∼47% product formation at 21 h as estimated by peak heights in the deconvoluted mass spectrum (calculated MW for *mTurq* = 28,870 Da, calculated MW for *mTurq-biotin* = 29,159 Da, calculated/observed MW for A236E/A241K baSrtB = 28,274/28,275 Da). Full synthesis and characterization details for preparing GGG-*biotin* are provided in Supporting Information and Fig. S7.

effects on overall catalytic efficiency, increasing it by > 100-fold (65). In addition, biochemical and structural studies by ourselves in recent years investigated the structurally-conserved β7-β8 loop of sortases, specifically amongst *Streptococcus* SrtA proteins, finding that residues in this loop can have a major impact on the specificity and activity of a given enzyme (13, 17, 25, 27, 54). Here, we chose to turn our attention to the understudied class B sortases, which had previously been shown to have relatively poor activity as compared to class A sortases from the same organisms (24, 28, 32, 36). Based on initial experiments with 4 enzymes, we chose *B. anthracis* SrtB to use as a model enzyme to investigate general properties of this sortase class and to test if our previous approach of engineering β7-β8 loop chimeras would elucidate a better understanding of class B sortases.

Our results revealed a number of new insights into baSrtB and class B sortases in general. As described here, baSrtB recognizes amino acids beyond the pentapeptide motif, here NPKTG, is typically used to describe sortase recognition. Specifically, a P2′ Asp is required for baSrtB activity, while the P5 Asp residue, and also P3′ Glu to a lesser extent, also contribute (Fig. 3*D*). Furthermore, a single mutation in the N-terminal helix, a structural feature that differentiates class B and class A sortase catalytic domains, Y39, can dramatically reduce activity confirming it is necessary for full baSrtB activity; we hypothesize this is due to interactions with the P2′ residue (Figs. 5*E*, 7, *B* and *C*). Due to a lack of general activity, we were unable to test if there are similar non-covalent interactions formed in the other SrtB enzymes studied and their respective P2′ target residues.

We also used AlphaFold2 (and subsequently AlphaFold3) to model the substrate-sortase interaction for several SrtB enzymes by utilizing a single polypeptide sequence combining the full-length primary sequences of substrate and SrtB with a Gly-Ser linker. Although additional tools to predict the structures of complexes that utilize AlphaFold have been introduced in recent years, *e.g.*, ColabFold or AlphaFold – Multimer, this approach was able to take advantage of freely available AlphaFold2 servers (here, the European Galaxy Server) and quickly predict the sortase-substrate structure (60, 66–68). Furthermore, our models suggested possible interactions between the transmembrane domains of both proteins, which we are currently testing. We hypothesize that this additional site of substrate recognition could increase reaction rates substantially in bacteria, and may prove useful for SML applications.

Our models largely agree with the binding conformation of available SrtA-peptide structures and seem to be catalytically competent based on what is known about the general sortase reaction mechanism (1, 21). While the only available structure of a class B sortase, *S. aureus* SrtB, bound to a peptidomimetic, NPQT*, raised questions about the lack of specific contacts with the P4 Asn in the binding motif, our models for 3 SrtB enzymes suggested that this amino acid is stabilized by non-covalent bonds with conserved Tyr (β6-β7 loop) and Ser (β7 strand) residues, as well as the catalytic Arg (Fig. 5*A*). We used mutagenesis to verify that knocking out any one of these residues abrogates binding for baSrtB and a DNPKTGDE peptide (Asn is underlined) (Fig. 7, *B* and *C*). These models also support earlier studies by ourselves and others that the

catalytic Arg likely stabilizes binding of the peptide substrate as opposed to direct contact with tetrahedral oxyanion intermediates, as previously proposed (15, 17, 23). Instead, we argue that a highly conserved Thr residue, immediately preceding the catalytic Cys, as well as backbone amides, stabilize the acyl-enzyme intermediate, which is consistent with the structural models presented here as well (17).

Finally, we identified more active baSrtB variants *via* our β7-β8 loop chimera experiments, which allowed for targeted mutagenesis of specific β7-β8 loop residues (Fig. 8). Specifically, we produced single (A241K) and double (A236E/A241K) mutant baSrtB variants with 4-fold increased activity over a ∼2-h period in our peptide activity assays. Importantly, all these variants included a 6xHis tag, which we discovered boosts the overall activity of the enzyme, perhaps due to an interaction with the P3′ Glu. The latter of these (A236E/A241K) was then evaluated for its potential use in sortase-mediated ligation (SML). We successfully showed that A236E/A241K baSrtB can indeed facilitate the modification of a model mTurq substrate *in vitro* (Fig. 9), which enables the use of an alternate SML recognition site (DNPKTGDE). In addition, these findings suggest that a more extensive mutagenesis screen, *e.g.*, using directed evolution as was previously done with SrtA, could uncover SrtB variants with even better catalytic efficiencies relatively easily. Considering the varied recognition sequences between SrtB enzymes, between each other and as compared to class A sortases, this may prove incredibly useful for SML applications. Taken together, our work provides additional insight into class B sortases and acts as a proof of principle that understudied sortases, or classes of sortases, may hold great potential for the next generation of sortase-based protein engineering tools.

Experimental procedures

Expression and purification of recombinant proteins

Accession numbers of SrtB catalytic domains used for expression and purification include: $saSrtB_{30\text{-}244}$ (UniProt SRTB_STAA8), $baSrtB_{35\text{-}254}$ (UniProt A0A6L8PZR0_BACAN), $cdSrtB_{26\text{-}255}$ (NCBI WP_167414662.1), and $lmSrtB_{26\text{-}246}$ (UniProt SRTB_LISMO). All baSrtB variants were based on this wild-type sequence. Proteins were expressed and purified largely as previously described for SrtA (17, 25, 27). Briefly, His-tagged sequences, which also contained a TEV protease cleavage site, were inserted into pET28a(+) plasmids (Genscript) and transformed into BL21 (DE3) *Escherichia coli* cells. Following growth and overexpression of the protein of interest using IPTG, cells were resuspended in lysis buffer [0.05 M Tris pH 7.5–8.0, 0.15 M NaCl, 0.5 mM ethylenediaminetetraacetic acid (EDTA)] and lysed by sonication. Purified protein was isolated by immobilized metal affinity chromatography (IMAC) using a 5 ml HisTrap HP column (Cytiva) with wash buffer [0.05 M Tris pH 7.5–8.0, 0.15 M NaCl, 0.02 M imidazole pH 7.5–8.0, and 0.001 M tris(2-carboxyethyl)phosphine (TCEP)] and elution buffer [wash buffer with 0.3 M imidazole pH 7.5–8.0]. Eluted protein was characterized by SDS-PAGE, and pure fractions were concentrated using Amicon Ultra 10K Centrifugal Filters. Final purification was achieved by size exclusion chromatography (SEC) with running buffer [0.05 M Tris pH 7.5, 0.15 M NaCl, 0.001 M TCEP for initial assays in Figs. 2 and 3*A*, S1, *C–E* or 0.1 M Tris pH 8.0, 0.3 M NaCl]. Protein concentrations were determined using theoretical extinction coefficients calculated using ExPASy ProtParam (69). Purified proteins were flash-frozen using liquid nitrogen and stored at −80 °C. Selected variants were monitored for relative stability (as oligomeric state) and a similar radius of gyration using SEC (Cytiva S75 10/300 column) (Fig. S1*A*).

The His tag of baSrtB protein used for crystallography and cleaved baSrtB assays was cleaved off using TEV protease. BaSrtB and TEV protease were incubated at a 1:100 ratio at 4 °C overnight. This was followed by a second IMAC purification step, with the buffers described above, prior to SEC. Crystallization attempts consisted of 1 to 1.5 mM C233A baSrtB protein in a 1:1 ratio with *Ac*-DNPKTGDE-NH_2 peptide. Following incubation and crystallization by hanging drop vapor diffusion using a mixture of 2 μl protein + 2 μl well solution, crystals appeared in a few days to weeks under several conditions. Examples included, 1) 0.225 M potassium acetate, 18% (*w/v*) PEG 3350; 2) 0.175 M sodium acetate, 20% (*w/v*) PEG 3350; 3) 0.21 sodium chloride, 22% (*w/v*) PEG 3350, 0.1 M Hepes pH 7; 4) 0.175 M sodium bromide, 22% (*w/v*) PEG 3350. However, in all cases, cryo X-ray diffraction data revealed that the peptide was not bound, as described in the main text.

The mTurq-DNPKTGDEGGGG sequence was also inserted into the pET28a(+) plasmid (Genscript) and included N-terminal His_6 and TEV cleavage site purification tags, as well as a flexible spacer (GSGCGGGGSGGGGS) between the mTurq domain and the DNPKTGDEGGGG extension. Expression and purification of mTurq-DNPKTGDEGGGG followed a similar protocol as that described above, with the exception that the SEC running buffer was 0.05 M Tris pH 8.0, 0.15 M NaCl, 0.001 M TCEP.

Peptide synthesis and sortase activity assays

Unless indicated otherwise, all substrate peptides were synthesized using manual Fmoc solid phase peptide synthesis (SPPS) and contained N-terminal 2-aminobenzoyl (Abz) and C-terminal 2,4-dinitrophenyl lysine (K(Dnp)) moieties, as previously described (17, 25, 27, 43). An additional, unlabeled control peptide (Ac-DNPKTGDE-NH_2, Ac = acetyl) was synthesized using analogous procedures. FRET activity assay reactions using Abz/Dnp labeled peptides were performed in black, flat-bottom Costar 96-well plates. Activity assay reactions were prepared with 0.05 mM SrtB unless noted below, 0.175 to 0.2 mM peptide model substrate, unless noted below, 5 mM nucleophile (*e.g.*, hydroxylamine or the others tested), 0.05 M Tris-HCl pH 7.0 to 9.0 (baSrtB = pH 7.5 then pH 8.0 based on pH experiment; saSrtB = pH 8.5; cdSrtB = 7.5; lmSrtB = pH 8.0), and 0.15 M NaCl. For the data in Fig. S1, *C–E*, 0.01 mM baSrtB and saSrtB and 0.02 mM cdSrtB were used, with 0.05 mM peptide model substrate. For the data in Figure 3*A*, 0.03 mM baSrtB was used, with 0.055 mM *Abz-*

DNPKTGDEK(Dnp)-NH_2. For the data in Fig. S1*F*, 0.05 mM baSrtB was used along with 0.2 mM Ac-DNPKTGDE-NH_2 and 5 mM triglycine (GGG) at pH 8.0. For all assays, peptide stocks often contained DMSO to aid solubility. Residual DMSO in reaction mixtures was ≤1.5%, which was confirmed to not affect baSrtB activity. Reaction monitoring *via* FRET was achieved by measuring fluorescence intensity (λ_{ex} = 320 nm, λ_{em} = 420 nm) using a Biotek Synergy H1 plate reader. All reactions were performed at least in triplicate technical replicates, unless indicated otherwise. Background-subtracted fluorescence (measured in relative fluorescence units, RFU) was obtained by subtracting the RFU of negative controls lacking SrtB. Data was plotted as line graphs using KaleidaGraph, v5.01 (Synergy Software).

Characterization of peptide activity assays using liquid chromatography mass spectrometry (LC-MS)

BaSrtB, saSrtB, and cdSrtB peptide activity assays were monitored over 20 to 48 h through RP-HPLC using a Dionex Ultimate 3000 HPLC system and a Phenomenex Kinetex 2.6 μM C18 100 Å column (100 × 2.1 mm) [aqueous (95% water, 5% MeCN, 0.1% formic acid)/MeCN (0.1% formic acid) mobile phase at 0.3 ml/min, method: hold 10% MeCN 0.0–1.0 min, linear gradient of 10–90% MeCN 1.0–7.0 min, hold 90% MeCN 7.0–9.0 min, linear gradient of 90–10% MeCN 9.0–9.1 min, re-equilibrate at 10% MeCN 9.1–13.4 min)]. The HPLC system was interfaced with an Advion CMS expression^L mass spectrometer in order to confirm substrate and product identity through electrospray ionization mass spectroscopy (ESI-MS).

For mass spectrometry characterization of baSrtB nucleophile activity assay products, reaction solutions were collected from the 96-well plate after approximately 20 h of incubation time. Protein was removed using Amicon Ultra 3K Centrifugal Filters and flow through was analyzed by LC-MS. For these assays, separation was achieved using a Dionex Ultimate 3000 HPLC system and a Phenomenex Kinetex 2.6 μM Polar C18 100 Å column (100 × 2.1 mm) [aqueous (95% water, 5% MeCN, 0.1% formic acid)/MeCN (0.1% formic acid) mobile phase at 0.3 ml/min, method: hold 0% MeCN 0.0–1.0 min, linear gradient of 0–90% MeCN 1.0–7.0 min, hold 90% MeCN 7.0–9.0 min, linear gradient of 90–0% MeCN 9.0–9.1 min, re-equilibrate at 0% MeCN 9.1–13.4 min)]. As noted above, the HPLC system was interfaced with an Advion CMS expression^L mass spectrometer for identification of reaction products by ESI-MS.

The model ligation between Ac-DNPKTGDE-NH_2 and triglycine (GGG) (Fig. S1*F*) was characterized by LC-MS after a 25 h incubation at room temperature. For this reaction, separation was performed using a Dionex Ultimate 3000 HPLC system and a Phenomenex Kinetex 2.6 μM Polar C18 100 Å column (100 × 2.1 mm) [water (0.1% formic acid)/MeCN (0.1% formic acid) mobile phase at 0.3 ml/min, method: hold 2% MeCN 0.0–1.0 min, linear gradient of 2–90% MeCN 1.0–7.0 min, hold 90% MeCN 7.0–9.0 min, linear gradient of 90–2% MeCN 9.0–9.1 min, re-equilibrate at 2% MeCN 9.1–13.4 min)]. Analysis of reaction products by ESI-MS was achieved using the inline Advion CMS expression^L mass spectrometer described above.

Biotinylation of mTurq using baSrtB catalyzed sortase-mediated ligation

Modification of mTurq was achieved by combining 200 μM mTurq-DNPKTGDEGGGG with 1 mM GGG-*biotin* (see Supporting Information and Fig. S7 for synthesis and characterization details) in the presence of 50 μM A236E/A241K baSrtB (70, 71). Running buffer [0.1 M Tris pH 8.0, 0.3 M NaCl] was added to reach a final buffer concentration of 0.05 M Tris pH 8.0, 0.15 M NaCl. Reactions also contained residual DMSO (2.1% v/v) from the GGG-*biotin* stock solution. Reactions were incubated at room temperature for 21 to 24 h. At various timepoints, aliquots of the reaction mixture were quenched using SDS sample buffer and subjected to analysis using SDS-PAGE. Reactions were also analyzed by LC-MS using an Agilent 6545XT AdvanceBio Q-TOF system interfaced with an Agilent 1290 HPLC system. Separations upstream of the Q-TOF were achieved with a Phenomenex Aeris 3.6 mM WIDEPORE C4 200 Å column (100 × 2.1 mm) [H_2O (0.1% formic acid)/MeCN (0.1% formic acid) mobile phase at 0.3 ml/min, method: hold 10% MeCN 0.0–1.0 min, linear gradient of 10–90% MeCN 1.0–9.0 min, hold 90% MeCN 9.0–11.0 min, linear gradient of 90–10% MeCN 11.0–11.1 min, re-equilibrate at 10% MeCN 11.1–15.0 min]. Deconvolution of protein charge ladders was achieved using Agilent MassHunter BioConfirm software (version 10.0). The extent of biotinylated product formation was estimated by comparing peak heights for the unreacted mTurq-DNPKTGDEGGGG substrate and the biotinylated mTurq product in the deconvoluted mass spectra.

Preparation and characterization of enzyme-substrate adducts

Formation of enzyme-substrate adducts was achieved by incubating wild-type baSrtB (50 μM) with Abz-DNPKTGDEK(Dnp)-NH_2 peptide (200 μM) in 0.05 M Tris pH 8.0, 0.15 M NaCl buffer for 24 or 27 h at room temperature. Reaction mixtures were then treated with either N-ethylmaleimide (NEM, 1 mM final concentration) for 1 h at room temperature, or TEV protease (1:25 M ratio relative to baSrtB) for 1 h at room temperature. All reactions, both before and after NEM/TEV treatment, were analyzed by LC-MS using the same procedures and instrumentation described above for analyzing the biotinylation of mTurq.

Molecular dynamics simulations

Sortase catalytic domains, equivalent to those expressed and purified for *in vitro* experiments, and the substrate residues indicated as the binding substrate were extracted from determined AlphaFold2 models. The peptide substrate was

reidentified as a separate chain. For the control simulation of baSrtB with polyAla (Fig. S5), the peptide was created using the mutagenesis wizard in PyMOL using DNPKTGDE as the starting sequence. Molecular dynamics simulations were performed in triplicate using GROMACS 2022.4 with the AMBER99SB-ILDN protein, nucleic AMBER94 force fields (72–76). PyMOL was used to cap the peptide and protein N-terminus with an acyl group, while the protein C-terminus was capped with N-methyl. The system was solvated with TIP3P water molecules and centered within a cubic box with periodic boundary conditions and a minimum of 1 nm from between the protein and the edge of the box. Ions were added to a 0.15 M physiological ion concentration with a neutral net charge balanced with sodium cations and chloride ions. The steepest descent energy minimization was performed on the solvated system with a maximum force tolerance of 1000 kJ/mol/nm for all structures. Long-range electrostatic interactions were treated with the particle mesh Ewald (PME) algorithm using a grid spacing of 0.16 nm and a 1.0 nm cutoff for Coulombic and Lennard-Jones interactions (77). All systems and replicates were equilibrated separately in an NVT ensemble for 100 picoseconds with an integration time step of 2 fs at a reference temperature of 300K with position restraints on all protein heavy atoms with velocities assigned randomly from a Maxwell-Boltzmann distribution. We used the stochastic velocity rescaling thermostat. All systems and replicates were equilibrated separately in an NPT ensemble for 5 nanoseconds without position restraints at a reference pressure of 1.0 bar using the Isotropic Parrinello-Rahman barostat (78). All bonds to hydrogen atoms were constrained using the LINCS algorithm (79). Sufficient equilibration was verified by a converged protein RMSD and system density. All systems and replicates were simulated in a separate production run in an NVT ensemble for 1000 nanoseconds at a reference temperature of 300K. Molecular dynamics simulations were viewed using PyMOL (Schrödinger software).

Structural modeling using AlphaFold2 (or AlphaFold3) and software used for structural analyses

Substrate sequences used for modeling from UniProt include: saIsdC (ISDC_STAA8), baIsdC (UniProt A0A2A7DBZ7_BACAN), Lmo2185 (HBP2_LISMO), and Lmo2186 (HBP1_LISMO). The accession numbers of the full-length SrtB sequences are the same as above. Structural models were determined with AlphaFold2 on the European Galaxy server, using default settings (60, 68, 80). Input sequences are in the Supporting Information. The output structures are ranked according to the predicted long-distance difference test (pLDDT) (60). Output structures were analyzed using PyMOL (Schrödinger software), and unless otherwise noted, the ranked_0 output was used as the reference structure. Subsequent modeling was done using the AlphaFold3 server (61). For models generated using AlphaFold3, the PAE plots, ipTM, and pTM scores are reported. The electrostatic potential surface map of baSrtB was determined using the APBS plugin in PyMOL.

Data availability

All data are contained in the article and supporting information.

Supporting information—This article contains supporting Information (70, 71).

Acknowledgment—We want to thank all additional members of the Amacher and Antos labs, as well as Dr Sierra Cullati, for helpful discussion and research support. We also gratefully acknowledge the support of Sarina Kiesser in running mass spectrometry assays.

Author contributions—D. E. L., J. M., J. M. A., J. M. B., and K. A. C. writing–review & editing; D. E. L., J. M. A., J. F. A., S. N. J., J. M. B., and K. A. C. visualization; D. E. L., S. N. J., J. M. B., and K. A. C. validation; D. E. L., J. W. I., K. W., C. M. C., K. M. B., N. J. G., S. N. J., J. M. B., and K. A. C. investigation; D. E. L., J. W. I., J. M., C. M. C., K. M. B., N. J. G., S. N. J., J. M. B., and K. A. C. formal analysis; J. M., J. F. A., and J. M. A. resources; J. M., J. M. A., and J. M. B. funding acquisition. J. M. A. supervision; J. M. A., J. M. A., and S. N. J. conceptualization; J. F. A. and S. N. J. writing–original draft.

Funding and additional information—The QTOF system was acquired via an NSF MRI grant (DBI-1920340) to J.M. Antos. This work was supported by NIH 1R15GM154315-01 to J.F. Amacher, J.M. Antos, and J. McCarty. It was additionally supported by NSF CHE-2044958 and a Cottrell Scholar Award from the Research Corporation for Science Advancement to J.F. Amacher, and NSF CHE-2102189 to J. McCarty. J.M. Blount was supported by a Cottrell Postbac Award. Although we were unable to determine a peptide-bound structure of baSrtB bound to a target sequence, we want to thank the Berkeley Center for Structural Biology (BCSB) for crystallography resources. The BCSB is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The Pilatus detector was funded under NIH grant S10OD021832.

Conflicts of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: baSrtB, *Bacillus anthracis* sortase B; CWSS, Cell Wall Sorting Signal; MD, molecular dynamics; saSrtA, *Staphylococcus aureus* sortase A; SEC, size exclusion chromatography; SML, sortase-mediated ligation; SrtA, class A sortase; SrtB, Class B sortase; TEV, Tobacco Etch Virus.

References

1. Jacobitz, A. W., Kattke, M. D., Wereszczynski, J., and Clubb, R. T. (2017) Sortase transpeptidases: structural biology and catalytic mechanism. *Adv. Protein Chem. Struct. Biol.* **109**, 223–264
2. Spirig, T., Weiner, E. M., and Clubb, R. T. (2011) Sortase enzymes in Gram-positive bacteria. *Mol. Microbiol.* **82**, 1044–1059
3. Mazmanian, S. K., Liu, G., Ton-That, H., and Schneewind, O. (1999) *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **285**, 760–763
4. Ton-That, H., Liu, G., Mazmanian, S. K., Faull, K. F., and Schneewind, O. (1999) Purification and characterization of sortase, the transpeptidase that

cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 12424–12429
5. Sapra, R., Rajora, A. K., Kumar, P., Maurya, G. P., Pant, N., and Haridas, V. (2021) Chemical biology of sortase A inhibition: a gateway to anti-infective therapeutic agents. *J. Med. Chem.* **64**, 13097–13130
6. Clancy, K. W., Melvin, J. A., and McCafferty, D. G. (2010) Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. *Biopolymers* **94**, 385–396
7. Zrelovs, N., Kurbatska, V., Rudevica, Z., Leonchiks, A., and Fridmanis, D. (2021) Sorting out the superbugs: potential of sortase A inhibitors among other antimicrobial strategies to tackle the problem of antibiotic resistance. *Antibiotics (Basel)* **10**, 164
8. Mao, H., Hart, S. A., Schink, A., and Pollok, B. A. (2004) Sortase-mediated protein ligation: a new method for protein engineering. *J. Am. Chem. Soc.* **126**, 2670–2671
9. Antos, J. M., Truttmann, M. C., and Ploegh, H. L. (2016) Recent advances in sortase-catalyzed ligation methodology. *Curr. Opin. Struct. Biol.* **38**, 111–118
10. Tsukiji, S., and Nagamune, T. (2009) Sortase-mediated ligation: a gift from Gram-positive bacteria to protein engineering. *Chembiochem* **10**, 787–798
11. Popp, M. W., Antos, J. M., Grotenbreg, G. M., Spooner, E., and Ploegh, H. L. (2007) Sortagging: a versatile method for protein labeling. *Nat. Chem. Biol.* **3**, 707–708
12. Malik, A., and Kim, S. B. (2019) A comprehensive in silico analysis of sortase superfamily. *J. Microbiol.* **57**, 431–443
13. Valgardson, J. D., Struyvenberg, S. A., Sailer, Z. R., Piper, I. M., Svendsen, J. E., Johnson, D. A., *et al.* (2022) Comparative analysis and ancestral sequence reconstruction of bacterial sortase family proteins generates functional ancestral mutants with different sequence specificities. *Bacteria* **1**, 121–135
14. Malik, A., Subramaniyam, S., Kim, C.-B., and Manavalan, B. (2022) SortPred: the first machine learning based predictor to identify bacterial sortases and their classes using sequence-derived information. *Comput. Struct. Biotechnol. J.* **20**, 165–174
15. Tian, B.-X., and Eriksson, L. A. (2011) Catalytic mechanism and roles of Arg197 and Thr183 in the *Staphylococcus aureus* sortase A enzyme. *J. Phys. Chem. B* **115**, 13003–13011
16. Ton-That, H., Mazmanian, S. K., Alksne, L., and Schneewind, O. (2002) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Cysteine 184 and histidine 120 of sortase form a thiolate-imidazolium ion pair for catalysis. *J. Biol. Chem.* **277**, 7447–7452
17. Johnson, D. A., Piper, I. M., Vogel, B. A., Jackson, S. N., Svendsen, J. E., Kodama, H. M., *et al.* (2022) Structures of *Streptococcus pyogenes* class A sortase in complex with substrate and product mimics provide key details of target recognition. *J. Biol. Chem.* **298**, 102446
18. Frankel, B. A., Tong, Y., Bentley, M. L., Fitzgerald, M. C., and McCafferty, D. G. (2007) Mutational analysis of active site residues in the *Staphylococcus aureus* transpeptidase SrtA. *Biochemistry* **46**, 7269–7278
19. Bradshaw, W. J., Davies, A. H., Chambers, C. J., Roberts, A. K., Shone, C. C., and Acharya, K. R. (2015) Molecular features of the sortase enzyme family. *FEBS J.* **282**, 2097–2114
20. Chen, J.-L., Wang, X., Yang, F., Li, B., Otting, G., and Su, X.-C. (2023) 3D structure of the transient intermediate of the enzyme–substrate complex of sortase A reveals how calcium binding and substrate recognition cooperate in substrate activation. *ACS Catal.* **13**, 11610–11624
21. Amacher, J. F., and Antos, J. M. (2024) Sortases: structure, mechanism, and implications for protein engineering. *Trends Biochem. Sci.* **49**, 596–610
22. Race, P. R., Bentley, M. L., Melvin, J. A., Crow, A., Hughes, R. K., Smith, W. D., *et al.* (2009) Crystal structure of *Streptococcus pyogenes* sortase A: implications for sortase mechanism. *J. Biol. Chem.* **284**, 6924–6933
23. Suree, N., Liew, C. K., Villareal, V. A., Thieu, W., Fadeev, E. A., Clemens, J. J., *et al.* (2009) The structure of the Staphylococcus aureus sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized. *J. Biol. Chem.* **284**, 24465–24477
24. Bentley, M. L., Gaweska, H., Kielec, J. M., and McCafferty, D. G. (2007) Engineering the substrate specificity of *Staphylococcus aureus* Sortase A. The beta6/beta7 loop from SrtB confers NPQTN recognition to SrtA. *J. Biol. Chem.* **282**, 6571–6581
25. Piper, I. M., Struyvenberg, S. A., Valgardson, J. D., Johnson, D. A., Gao, M., Johnston, K., *et al.* (2021) Sequence variation in the β7-β8 loop of bacterial class A sortase enzymes alters substrate selectivity. *J. Biol. Chem.* **297**, 100981
26. Wójcik, M., Szala, K., van Merkerk, R., Quax, W. J., and Boersma, Y. L. (2020) Engineering the specificity of *Streptococcus pyogenes* sortase A by loop grafting. *Proteins* **88**, 1394–1400
27. Gao, M., Johnson, D. A., Piper, I. M., Kodama, H. M., Svendsen, J. E., Tahti, E., *et al.* (2022) Structural and biochemical analyses of selectivity determinants in chimeric *Streptococcus* Class A sortase enzymes. *Protein Sci.* **31**, 701–715
28. Jacobitz, A. W., Wereszczynski, J., Yi, S. W., Amer, B. R., Huang, G. L., Nguyen, A. V., *et al.* (2014) Structural and computational studies of the *Staphylococcus aureus* sortase B-substrate complex reveal a substrate-stabilized oxyanion hole. *J. Biol. Chem.* **289**, 8891–8902
29. Mazmanian, S. K., Ton-That, H., Su, K., and Schneewind, O. (2002) An iron-regulated sortase anchors a class of surface protein during *Staphylococcus aureus* pathogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 2293–2298
30. Mazmanian, S. K., Skaar, E. P., Gaspar, A. H., Humayun, M., Gornicki, P., Jelenska, J., *et al.* (2003) Passage of heme-iron across the envelope of *Staphylococcus aureus*. *Science* **299**, 906–909
31. Zong, Y., Mazmanian, S. K., Schneewind, O., and Narayana, S. V. L. (2004) The structure of sortase B, a cysteine transpeptidase that tethers surface protein to the *Staphylococcus aureus* cell wall. *Structure* **12**, 105–112
32. Puorger, C., Di Girolamo, S., and Lipps, G. (2017) Elucidation of the recognition sequence of sortase B from *Bacillus anthracis* by using a newly developed liquid chromatography-mass spectrometry-based method. *Biochemistry* **56**, 2641–2650
33. Maresso, A. W., Chapa, T. J., and Schneewind, O. (2006) Surface protein IsdC and Sortase B are required for heme-iron scavenging of *Bacillus anthracis*. *J. Bacteriol.* **188**, 8145–8152
34. Mariscotti, J. F., García-del Portillo, F., and Pucciarelli, M. G. (2009) The *Listeria monocytogenes* sortase-B recognizes varied amino acids at position 2 of the sorting motif. *J. Biol. Chem.* **284**, 6140–6146
35. Pucciarelli, M. G., Calvo, E., Sabet, C., Bierne, H., Cossart, P., and García-del Portillo, F. (2005) Identification of substrates of the *Listeria monocytogenes* sortases A and B by a non-gel proteomic analysis. *Proteomics* **5**, 4808–4817
36. Kang, H. J., Coulibaly, F., Proft, T., and Baker, E. N. (2011) Crystal structure of Spy0129, a *Streptococcus pyogenes* class B sortase involved in pilus assembly. *PLoS One* **6**, e15969
37. Donahue, E. H., Dawson, L. F., Valiente, E., Firth-Clark, S., Major, M. R., Littler, E., *et al.* (2014) *Clostridium difficile* has a single sortase, SrtB, that can be inhibited by small-molecule inhibitors. *BMC Microbiol.* **14**, 219
38. Chambers, C. J., Roberts, A. K., Shone, C. C., and Acharya, K. R. (2015) Structure and function of a *Clostridium difficile* sortase enzyme. *Sci. Rep.* **5**, 9449
39. Kang, C.-Y., Huang, I.-H., Chou, C.-C., Wu, T.-Y., Chang, J.-C., Hsiao, Y.-Y., *et al.* (2020) Functional analysis of *Clostridium difficile* sortase B reveals key residues for catalytic activity and substrate specificity. *J. Biol. Chem.* **295**, 3734–3745
40. Yin, J.-C., Fei, C.-H., Lo, Y.-C., Hsiao, Y.-Y., Chang, J.-C., Nix, J. C., *et al.* (2016) Structural insights into substrate recognition by *Clostridium difficile* sortase. *Front. Cell Infect. Microbiol.* **6**, 160
41. van Leeuwen, H. C., Klychnikov, O. I., Menks, M. A. C., Kuijper, E. J., Drijfhout, J. W., and Hensbergen, P. J. (2014) *Clostridium difficile* sortase recognizes a (S/P)PXTG sequence motif and can accommodate diaminopimelic acid as a substrate for transpeptidation. *FEBS Lett.* **588**, 4325–4333
42. Xiao, Q., Jiang, X., Moore, K. J., Shao, Y., Pi, H., Dubail, I., *et al.* (2011) Sortase independent and dependent systems for acquisition of haem and haemoglobin in *Listeria monocytogenes*. *Mol. Microbiol.* **80**, 1581–1597

43. Nikghalb, K. D., Horvath, N. M., Prelesnik, J. L., Banks, O. G. B., Filipov, P. A., Row, R. D., *et al.* (2018) Expanding the scope of sortase-mediated ligations by using sortase Homologues. *Chembiochem* **19**, 185–195
44. Schmohl, L., Bierlmeier, J., von Kügelgen, N., Kurz, L., Reis, P., Barthels, F., *et al.* (2017) Identification of sortase substrates by specificity profiling. *Bioorg. Med. Chem.* **25**, 5002–5007
45. Kruger, R. G., Otvos, B., Frankel, B. A., Bentley, M., Dostal, P., and McCafferty, D. G. (2004) Analysis of the substrate specificity of the *Staphylococcus aureus* sortase transpeptidase SrtA. *Biochemistry* **43**, 1541–1551
46. Matsumoto, T., Takase, R., Tanaka, T., Fukuda, H., and Kondo, A. (2012) Site-specific protein labeling with amine-containing molecules using Lactobacillus plantarum sortase. *Biotechnol. J.* **7**, 642–648
47. Boekhorst, J., de Been, M. W. H. J., Kleerebezem, M., and Siezen, R. J. (2005) Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs. *J. Bacteriol.* **187**, 4928–4934
48. Sizar, O., Leslie, S. W., and Unakal, C. G. (2024) *StatPearls*, StatPearls publishing, Treasure Island (FL)
49. Zhang, R., Wu, R., Joachimiak, G., Mazmanian, S. K., Missiakas, D. M., Gornicki, P., *et al.* (2004) Structures of sortase B from *Staphylococcus aureus* and *Bacillus anthracis* reveal catalytic amino acid triad in the active site. *Structure* **12**, 1147–1156
50. Maresso, A. W., Wu, R., Kern, J. W., Zhang, R., Janik, D., Missiakas, D. M., *et al.* (2007) Activation of inhibitors by sortase triggers irreversible modification of the active site. *J. Biol. Chem.* **282**, 23129–23139
51. Tamai, E., Sekiya, H., Maki, J., Nariya, H., Yoshida, H., and Kamitori, S. (2017) X-ray structure of *Clostridium perfringens* sortase B cysteine transpeptidase. *Biochem. Biophys. Res. Commun.* **493**, 1267–1272
52. Zähner, D., and Scott, J. R. (2008) SipA is required for pilus formation in *Streptococcus pyogenes* serotype M3. *J. Bacteriol.* **190**, 527–535
53. Kruger, R. G., Dostal, P., and McCafferty, D. G. (2004) Development of a high-performance liquid chromatography assay and revision of kinetic parameters for the *Staphylococcus aureus* sortase transpeptidase SrtA. *Anal. Biochem.* **326**, 42–48
54. Vogel, B. A., Blount, J. M., Kodama, H. M., Goodwin-Rice, N. J., Andaluz, D. J., Jackson, S. N., *et al.* (2023) A unique binding mode of P1′ Leu-containing target sequences for *Streptococcus pyogenes* sortase A results in alternative cleavage. *RSC Chem. Biol.* **5**, 30–40
55. Weitner, T., Friganović, T., and Šakić, D. (2022) Inner filter effect correction for fluorescence measurements in microplates using variable vertical axis focus. *Anal. Chem.* **94**, 7107–7114
56. Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O., and Clubb, R. T. (2001) Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 6056–6061
57. Dramsi, S., Magnet, S., Davison, S., and Arthur, M. (2008) Covalent attachment of proteins to peptidoglycan. *FEMS Microbiol. Rev.* **32**, 307–320
58. Freund, C., and Schwarzer, D. (2021) Engineered sortases in peptide and protein chemistry. *Chembiochem* **22**, 1347–1356
59. Frankel, B. A., Kruger, R. G., Robinson, D. E., Kelleher, N. L., and McCafferty, D. G. (2005) *Staphylococcus aureus* sortase transpeptidase SrtA: insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. *Biochemistry* **44**, 11188–11200
60. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589
61. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., *et al.* (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500
62. Dai, X., Böker, A., and Glebe, U. (2019) Broadening the scope of sortagging. *RSC Adv.* **9**, 4700–4721
63. Aulabaugh, A., Ding, W., Kapoor, B., Tabei, K., Alksne, L., Dushin, R., *et al.* (2007) Development of an HPLC assay for *Staphylococcus aureus* sortase: evidence for the formation of the kinetically competent acyl enzyme intermediate. *Anal. Biochem.* **360**, 14–22
64. Zou, Z., Ji, Y., and Schwaneberg, U. (2023) Empowering site-specific bioconjugations in vitro and in vivo: advances in sortase engineering and sortase-mediated ligation. *Angew. Chem. Int. Ed.* **63**, e202310910
65. Chen, I., Dorr, B. M., and Liu, D. R. (2011) A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11399–11404
66. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022) ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682
67. [preprint] Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A. W., Green, T., *et al.* (2021) Protein complex prediction with AlphaFold-multimer. *BioRxiv*. https://doi.org/10.1101/2021.10.04.463034
68. Galaxy Community (2022) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res.* **50**, W345–W351
69. Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., *et al.* (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* **112**, 531–552
70. Pichlo, C., Juetten, L., Wojtalla, F., Schacherl, M., Diaz, D., and Baumann, U. (2019) Molecular determinants of the mechanism and substrate specificity of Clostridium difficile proline-proline endopeptidase-1. *J. Biol. Chem.* **294**, 11525–11535
71. Carmona, A. K., Schwager, S. L., Juliano, M. A., Juliano, L., and Sturrock, E. D. (2006) A continuous fluorescence resonance energy transfer angiotensin I-converting enzyme assay. *Nat. Protoc.* **1**, 1971–1976
72. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., *et al.* (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958
73. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005) GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718
74. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., *et al.* (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25
75. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., *et al.* (2013) Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854
76. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174
77. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577
78. Parrinello, M. (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182
79. Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472
80. Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., *et al.* (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **46**, W537–W544
81. Chan, A. H., Yi, S. W., Terwilliger, A. L., Maresso, A. W., Jung, M. E., and Clubb, R. T. (2015) Structure of the Bacillus anthracis sortase A enzyme bound to its sorting signal: a flexible amino-terminal appendage modulates substrate access. *J. Biol. Chem.* **290**, 25461–25474