Comparative analysis and ancestral sequence reconstruction of bacterial sortase family proteins generates functional ancestral mutants with different sequence specificities

by

Jordan D. Valgardson, Sarah A. Struyvenberg, Zachary R. Sailer, Isabel M. Piper, Justin E. Svendsen, D. Alex Johnson, Brandon A. Vogel, John M. Antos, Michael J. Harms, and Jeanine F. Amacher

Table of Contents

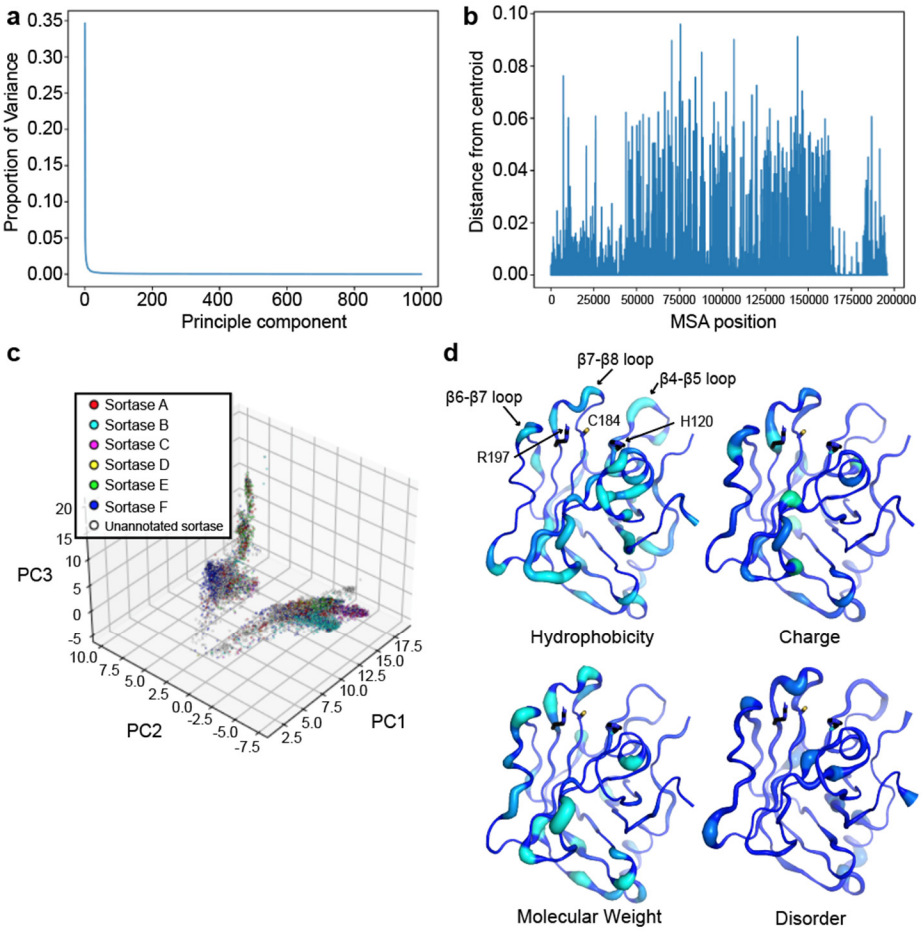

Figure S1. Principal component analysis (PCA) of sortase superfamily reveals sequence variability in structurally-conserved loops. (A) Scree plot showing the variance explained for each principle component for the first 1000 principle components. (B) Scatter plot of all 39,188 sortase proteins available from UniProt in principle component space for the first three principal components PC1, PC2, and PC3. (C) Distance from the origin for each position in the multisequence alignment for the first three principle components. (D) Variable residues for hydrophobicity, charge, molecular weight and disorder propensity highlighted on PDB 3FN5, as labeled. The *S. pyogenes* SrtA protein is shown in cartoon representation. Both color and width indicate level of variability that resulted in the PCA, with lighter colors and greater width indicating a larger degree of variability.

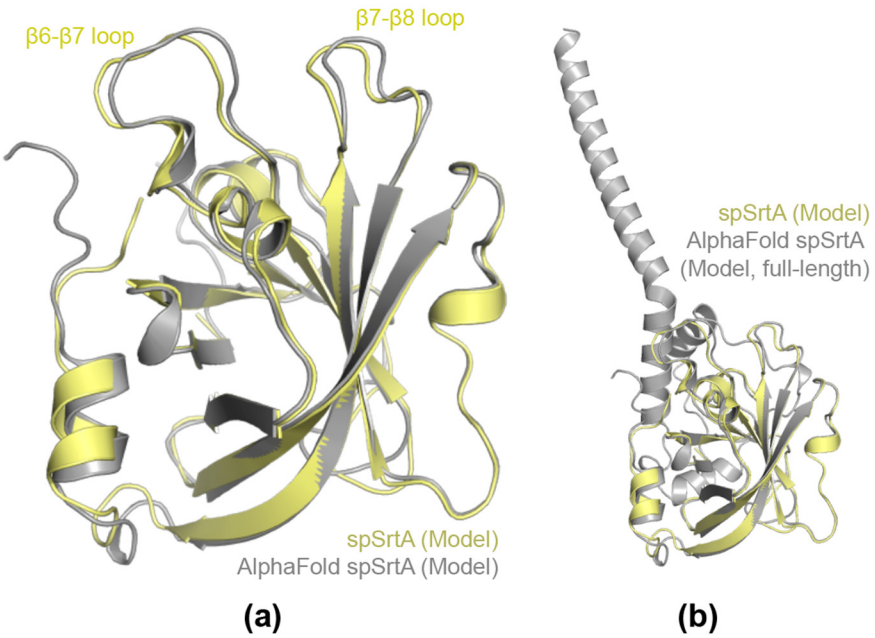

Figure S2. Structures of spSrtA, from the AlphaFold database and homology modeling. The structures of spSrtA from homology modeling (yellow) and the AlphaFold Protein Structure Database (gray) are in shown in cartoon representation. Alignment of the extracellular region of the protein **(a)** reveals an RMSD for main chain atoms of 0.501 Å (489 atoms), with the largest amount of variation in the β6-β7 and β7-β8 loops (labeled in yellow text). The full-length AlphaFold model of spSrtA is included in **(b)**, representing the intramembrane helix and N-terminal residues.

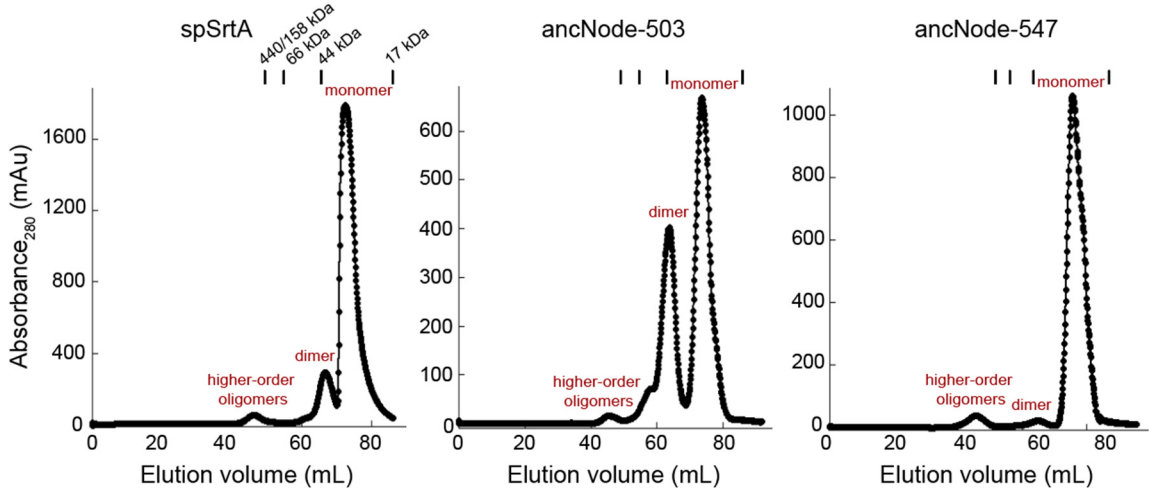

Figure S3. Size exclusion chromatography of ancestral SrtA proteins. Example chromatograms following size exclusion chromatography (Superdex S75 16/60, Cytiva) for wild-type spSrtA (left), ancNode-503 (middle), and ancNode-547 (right). During purification, fractions that correspond to the monomeric peak are selected. Previous work showed that the monomeric protein is stable over time and higher order oligomers do not reappear [1]. Size standards are based on published Cytiva documentation for the Superdex S75 16/60: from left to right, 440/158 kDa, 66 kDa, 44 kDa, and 17 kDa. The predicted molecular weights for each protein are 20.1 kDa (spSrtA), 18.3 kDa (ancNode-503), and 17.8 kDa (ancNode-547).

Recombinant protein sequences used in this study. Underlined amino acids are those included in purification and protease cleavage tags. All ancestral proteins contain a 6x His-tag in combination with a TEV protease cleavage site. The wild-type saSrtA and spSrtA sequences were previously published [2].

>*Staphylococcus aureus* Δ59-sortase A (saSrtA)
MGSSHHHHHHSSGLVPRGSHMQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNRGVSFAEE
NESLDDQNISIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRDVKPTDVGVLDEQKG
KDKQLTLITCDDYNEKTGVWEKRKIFVATEVK

>*Streptococcus pneumoniae* Δ80-sortase A (spSrtA)
MESSHHHHHHAVLTSQWDAQKLPVIGGIAIPELEMNLPIFKGLDNVNLFYGAGTMKREQVMGEGNY
SLASHHIFGVDNANKMLFSPLDNAKNGMKIYLTDKNKVYTYEIREVKRVTPDRVDEVDDRDGVNEITL
VTCEDLAATERIIVKGDLKETKDYSQTSDEILTAFNQPYKQFY

>ancStaphSrtA
MESSHHHHHHENLYFQSQKPPEIPKDKSKMAGYISVPDADIKEPVYPGPATPEQLNRGVSFAEEDESLD
DQNISIAGHTFTDRPHYQFTNLKAAKKGSKVYFKVGNETRKYKMTSIRDVNPDDVEVLDEQGEKNQLT
LITCDDYNEQTGVWEKRKIFVAEQVK

>ancStrepSrtA
MESSHHHHHHENLYFQSISLVAQAQSNNLPVIGGIAIPELGINLPIFKGVGNTSLLYGAGTMKEDQVM
GEGNYALASHHIFGVTASDMLFSPLERAKNGMKIYLTDKDNVYTYTITSVEVVTPERVDVIDDTEGKKE
ITLVTCTDYEATQRIIVKGELEETTPYNEASEDILNAFNQSYNQF

>ancNode-408
MESSHHHHHHENLYFQSIQPPSLSAKVDKSAIGQIAIPSVGLNLPIFKGTTNENLLAGAGTMSPDQKMG
EGNYVLAGHHMREDLLFGPLMKVKKGDKIYLTDQNEVYTYKVTETKVVHETDTSVLDDTGEPRLTLIT
CDTDTDQRFVVTAELVEKEPMKEESQEVKYQQKNQFILFLLL

>ancNode-503
MESSHHHHHHENLYFQSEPPSLASAKMDKQVIGQIAIPSVNINLPILKGTTNENLLAGAATMKPDQKM
GKGNYVLAGHHMREDLLFSPLHNVKKGDKIYLTDNKHVYTYKVTETKVVDPTETDVLDDTGEPQITLI
TCDNTDKRLVVTGELVETTPFEEEQVK

>ancNode-547
MESSHHHHHHENLYFQSSLSLAKARMDDLHVIGAIAIPSVNMNLPILKGVSNENLAVGAGTMKPDQK
MGKGNYALAGHHMNNPNLLFSPLHRVKKGDKIYLTDMKHVYTYKVTSTKVVDPTEVDVIDDTGEPLI
TLITCDDDGTNRLIVQGELVETTPFDA

Supplemental References

1. Piper, I. M.; Struyvenberg, S. A.; Valgardson, J. D.; Johnson, D. A.; Gao, M.; Johnston, K.; Svendsen, J. E.; Kodama, H. M.; Hvorecny, K. L.; Antos, J. M.; Amacher, J. F. Sequence variation in the β7-β8 loop of bacterial class A sortase enzymes alters substrate selectivity. *J. Biol. Chem.* **2021**, *297*, 100981.
2. Nikghalb, K. D.; Horvath, N. M.; Prelesnik, J. L.; Banks, O. G. B.; Filipov, P. A.; Row, R. D.; Roark, T. J.; Antos, J. M. Expanding the Scope of Sortase-Mediated Ligations by Using Sortase Homologues. *Chembiochem* **2018**, *19*, 185–195.