

## RESEARCH ARTICLE

# The structure of nontypeable *Haemophilus influenzae* SapA in a closed conformation reveals a constricted ligand-binding cavity and a novel RNA binding motif

Petra Lukacik<sup>1,2</sup>, C. David Owen<sup>1,2</sup>, Gemma Harris<sup>2</sup>, Jani Reddy Bolla<sup>3</sup>, Sarah Picaud<sup>4</sup>, Irfan Alibay<sup>5</sup>, Joanne E. Nettleship<sup>2,6</sup>, Louise E. Bird<sup>2,6</sup>, Raymond J. Owens<sup>2,6</sup>, Philip C. Biggin<sup>5</sup>, Panagis Filippakopoulos<sup>4</sup>, Carol V. Robinson<sup>3</sup>, Martin A. Walsh<sup>1,2\*</sup>

**1** Diamond Light Source, Harwell Science & Innovation Campus, Didcot, Oxfordshire, United Kingdom, **2** Research Complex at Harwell, Harwell Science and Innovation Campus, Didcot, Oxfordshire, United Kingdom, **3** Department of Chemistry, University of Oxford, Oxford, United Kingdom, **4** Structural Genomics Consortium, University of Oxford, Oxford, United Kingdom, **5** Department of Biochemistry, University of Oxford, Oxford, United Kingdom, **6** Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom

\* [martin.walsh@diamond.ac.uk](mailto:martin.walsh@diamond.ac.uk)



## OPEN ACCESS

**Citation:** Lukacik P, Owen CD, Harris G, Bolla JR, Picaud S, Alibay I, et al. (2021) The structure of nontypeable *Haemophilus influenzae* SapA in a closed conformation reveals a constricted ligand-binding cavity and a novel RNA binding motif. PLoS ONE 16(10): e0256070. <https://doi.org/10.1371/journal.pone.0256070>

**Editor:** Annalisa Pastore, National Institute for Medical Research, Medical Research Council, London, UNITED KINGDOM

**Received:** May 6, 2021

**Accepted:** July 29, 2021

**Published:** October 15, 2021

**Copyright:** © 2021 Lukacik et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data are held or will be held in a public repository PDB <https://www.rcsb.org/> PDBIDs 6HHL 6HIM 6HJ1.

**Funding:** MW MRC grant No MR/M010937/1 PL MRC grant No MR/M010937/1 PF MRC grant MR/N010051/1 <https://mrc.ukri.org/> The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Nontypeable *Haemophilus influenzae* (*NTHi*) is a significant pathogen in respiratory disease and otitis media. Important for *NTHi* survival, colonization and persistence *in vivo* is the Sap (sensitivity to antimicrobial peptides) ABC transporter system. Current models propose a direct role for Sap in heme and antimicrobial peptide (AMP) transport. Here, the crystal structure of SapA, the periplasmic component of Sap, in a closed, ligand bound conformation, is presented. Phylogenetic and cavity volume analysis predicts that the small, hydrophobic SapA central ligand binding cavity is most likely occupied by a hydrophobic di- or tri-peptide. The cavity is of insufficient volume to accommodate heme or folded AMPs. Crystal structures of SapA have identified surface interactions with heme and dsRNA. Heme binds SapA weakly ( $K_d$  282  $\mu$ M) through a surface exposed histidine, while the dsRNA is coordinated via residues which constitute part of a conserved motif (estimated  $K_d$  4.4  $\mu$ M). The RNA affinity falls within the range observed for characterized RNA/protein complexes. Overall, we describe in molecular-detail the interactions of SapA with heme and dsRNA and propose a role for SapA in the transport of di- or tri-peptides.

## Introduction

Respiratory illness currently ranks as the third leading cause of death globally [1], with rates peaking for the very young and the elderly [2, 3]. High incidence of the gram-negative bacterium nontypeable *Haemophilus influenzae* (*NTHi*) in these illnesses makes it a strategic target for biochemical investigation. Following triggers that are not yet fully understood, *NTHi* can transition from a commensal into a pathogen, leading to opportunistic respiratory tract

**Competing interests:** NO authors have competing interests.

**Abbreviations:** *NTHi*, nontypeable *Haemophilus influenzae*; COPD, chronic obstructive pulmonary disease; OM, otitis media; Sap, sensitivity to antimicrobial peptides; AMP, antimicrobial peptide; SBP, substrate-binding protein; hBD, human Beta-Defensin; cBD, chinchilla Beta-Defensin; ITC, isothermal titration calorimetry; EMSA, electrophoretic mobility shift assay; LI, *L. lactis*; Ec, *E. coli*; Hp, *H. parasuis*.

infections such as in chronic obstructive pulmonary disease (COPD) [4], pneumonia and exacerbations of cystic fibrosis. Outside of the respiratory context, *NTHi* is also a major pathogen in meningitis and otitis media (OM), with the latter being a leading cause of specifically treated disease in children both in the UK and worldwide [5, 6]. Taken together the costs of *NTHi* related diseases represent a major economic and social burden.

While antibiotic resistance in *NTHi* is primarily effected through  $\beta$ -lactamase, the bacterium is also gaining resistance to other antibiotic classes [7, 8]. Even without this additional threat, ampicillin resistant strains now exceed 30% in some countries, which has inevitably led to higher rates of treatment failure, increased costs, and decreased availability for severely or chronically ill patients. The extracellular capsule targeted by the now routine *Hi* type b (Hib) vaccine [9] is absent in *NTHi* and, as yet, no effective vaccine exists against it. Furthermore, with the introduction of the Hib vaccine, *NTHi* has become more prevalent [10]. *NTHi* vaccine development has proven to be challenging, in part due to high heterogeneity of outer membrane protein vaccine candidates and antigenic drift present in patients with long-term *NTHi* infection [11, 12]. The development of novel treatments, antimicrobials and vaccines is therefore much needed and would be aided by an improved understanding of *NTHi* pathogenic survival strategies.

An essential system for *NTHi* survival and persistence *in vivo* is the Sap (sensitivity to antimicrobial peptides) transporter. This multi-component system uses a periplasmic protein (SapA) to bind and deliver substrates to the heterodimeric inner membrane-associated Sap permease (SapBC). The Sap ABC transporter is completed by two membrane-associated nucleotide-binding proteins (SapD & SapF) that hydrolyze ATP to provide energy for substrate transport across the bacterial inner membrane.

Studies based mainly on *NTHi sap* mutants have shown a decrease in sensitivity to host antimicrobial peptides (AMPs) both *in vitro* and *in vivo*. As part of the innate immune system, AMPs contribute to the first line of defense against infection. These small (<10 kDa), cationic or amphipathic peptides have diverse structures and mostly act by disrupting cell membranes leading to cell lysis and death. Cell killing assays that tested *sapA*, *sapD*, and *sapBC* mutant strains have shown these to be more sensitive to AMP-mediated killing. Specifically, the *sapA* mutant is more sensitive to chinchilla Beta Defensin 1 (cBD1) [13]; *sapD* to cBD1 [14], cathelicidin LL-37 and human Beta Defensin 3(hBD3) [15]; and *sapBC* to LL-37 and hBD3 [16]. Growth of *sapA* [13], *sapD* [14] and *sapF* [15] mutant strains was attenuated in a chinchilla model of OM. The *sap* genes have also been postulated to have a role in iron acquisition [17, 18], with *sapA*, *sapBC* and *sapF* mutants demonstrating an inability to utilize heme for growth in specific nutrient controlled media. Evidence has also been presented for a direct interaction of SapA protein with heme [17, 19], and with cBD1 and LL-37 [14]. Additionally, hBD1, hBD2, hBD3, LL-37, human neutrophil protein 1, and melittin were shown to displace bound heme from SapA [17]. *NTHi* [20] and *K. pneumoniae* [21] *sapA* mutants have decreased adhesion to epithelial cells and increased invasiveness, while the *sapF* mutant has altered biofilm morphology [15]. The importance of the *sap* genes in resistance to AMPs has been demonstrated in other bacterial pathogens, including *H. ducreyi* [22, 23], *S. typhimurium* [24–27] and *E. chrysanthemi* [28].

Based on these studies a model has been proposed [17] for a multifunctional Sap transporter where SapA shuttles heme and AMPs to the Sap ABC transporter which moves them across the cytoplasmic membrane. Within the cytoplasm, AMPs are degraded and heme is utilized in a nutritional context. The disruption of the *sap* genes clearly has dramatic effects but the precise molecular interactions that are responsible for these effects are poorly understood. To shed light on these issues we have determined the crystal structure of *NTHi* SapA and characterized its interactions with a number of chemically distinct ligands.

## Results

### SapA crystallized in open and closed conformations

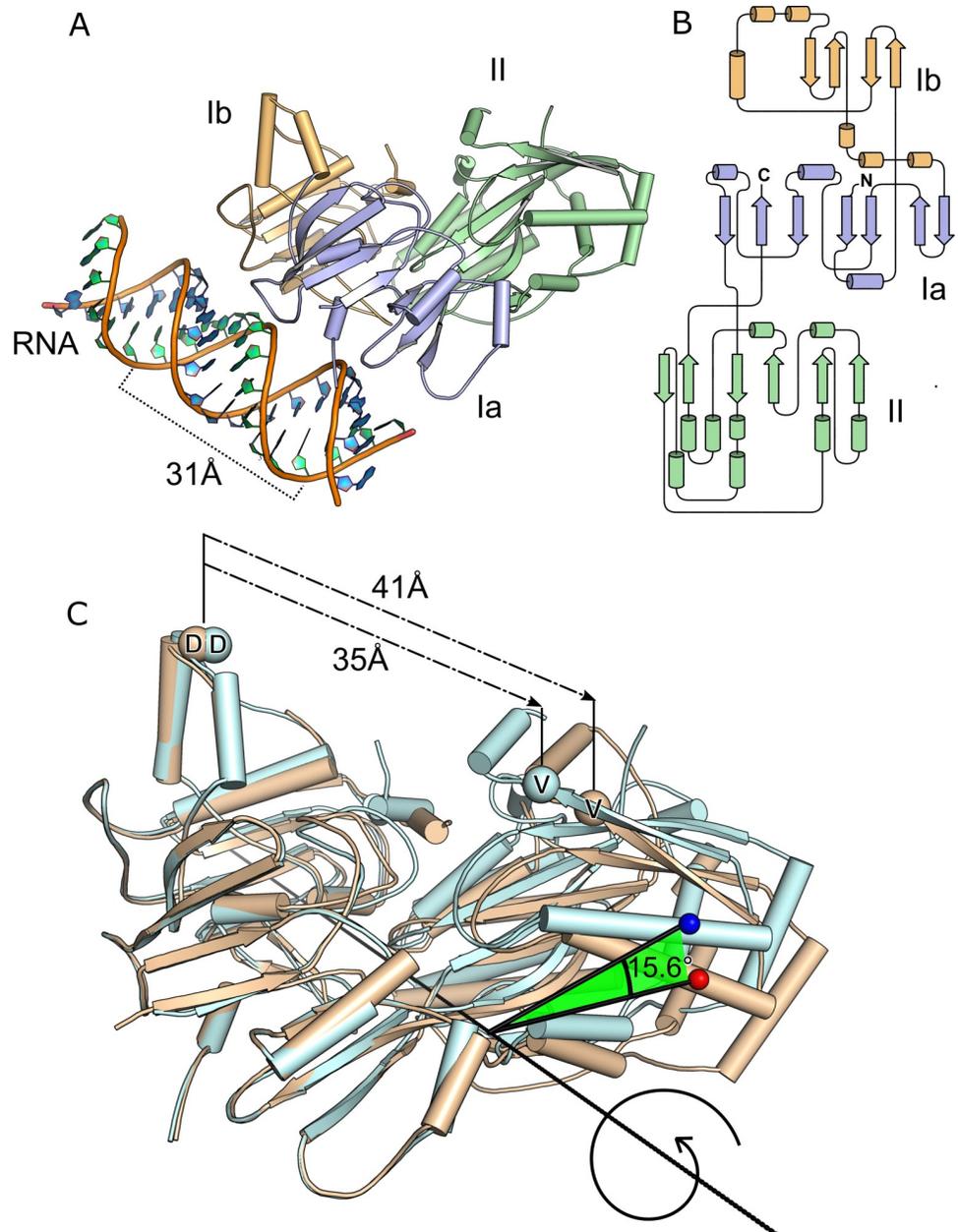
Despite numerous optimization attempts, the expression of SapA was poor, with a final yield of  $\approx 5$  mg pure protein per 50 g of *E. coli* cells. Crystals were nevertheless obtained and the structure of the 60 kDa SapA protein was solved to a resolution of 2.6 Å. The structure confirmed the classification of SapA into the Substrate Binding Protein (SBP) superfamily, as was predicted from sequence analysis. SBPs are a class of proteins that are often associated with membrane protein complexes, in particular, ABC transporters [29]. Within the SBP superfamily, SapA can be further categorized as a Class II SBP [30] based on  $\beta$ -sheet topology, and to SBP Cluster C by the atypical presence of an additional domain of approximately 15 kDa [29, 31]. In the SapA crystal structure this additional domain, Ib, associates with domain Ia to form one lobe, whilst domain II forms the second lobe (Fig 1A). Domain Ia encompasses residues 35:68, 230:312, 532:561; residues 69:229 constitute domain Ib, and domain II is formed by residues 313:531. All three domains present a mixed  $\alpha/\beta$  architecture with central  $\beta$ -sheets surrounded by  $\alpha$ -helices (Fig 1B). The polypeptide chain is for the most part in well-defined electron density, except for residues 139–162 which form a loop that could not be modeled.

We crystallized SapA in several forms (Table 1, S1A–S1C Fig): the first, denoted as SapA<sub>closed</sub> (PDB = 7OFZ), contains a single SapA chain with its two lobes in close contact with each other. The second, denoted as SapA<sub>mixed</sub> (PDB = 7OG0), contains two copies of SapA, with chain A adopting a similar conformation to SapA<sub>closed</sub> and chain B showing substantially greater separation between the lobes, this conformation is denoted SapA<sub>open</sub>. As a comparison, the distance between Asp 222 and Val 435, residues that respectively reside near to the termini of lobes I and II, are 35 Å apart in SapA<sub>closed</sub> and over 40 Å apart in SapA<sub>open</sub>. A transition from the open to closed state requires a 15.6° rotation around the axis bisecting domains I and II (Fig 1C). Conformational shifts such as these are characteristic of the SBP superfamily upon binding/release of their cognate ligand [29] and have been described as a “Venus flytrap” mechanism [32]. The structure of SapA<sub>closed</sub> aligns well with the structures of its homologs in their ligand-occupied states (Fig 2). Specifically, the alignment rmsds of SapA<sub>closed</sub> with ligand occupied *H. parasuis* HbpA (HpHbpA), *E. coli* DppA (EcDppA) and *Pseudoaltermonas* sp. SM9913 DppA (PsDppA) are 1.71 Å, 2.33 Å and 2.27 Å, respectively. The observation that the SapA<sub>closed</sub> structure was in the same conformation as its ligand-bound homologs was unexpected as no extraneous ligands were added to the purified protein or crystals.

### SapA binds an endogenous ligand within its small and hydrophobic/neutral binding cavity

In SapA<sub>closed</sub>, the lobes enclose a cavity with two narrow openings to the protein surface. Difference electron density ( $F_o - F_c$ ) became apparent within this cavity during the later stages of refinement (Fig 3A). This additional density was not observed in SapA<sub>open</sub>. Since the location of the SapA<sub>closed</sub> cavity approximately coincides with the ligand binding cavities of other SBPs (Fig 2), we inferred that this density might represent an endogenous ligand likely derived from the cell lysate. The presence of an endogenous ligand rationalizes the observed closed conformation of SapA<sub>closed</sub>. The amino acids lining the ligand-binding cavity of SapA are predominantly hydrophobic or neutral (Table 2), the only exceptions are Glu 63, Lys 68 and Asp 71; however, these residues are located at the opening of the ligand binding cavity and may be classed as surface exposed.

Analysis of the ligand binding cavities across a range of peptide binding proteins structurally homologous to SapA carried out by [33] showed that cavity volumes correlate well with



**Fig 1. Overall SapA structure.** A, the overall structure of SapA<sub>closed</sub> coloured by domain. Each SapA monomer was co-crystallised with a single-stranded A-form RNA molecule. The antisense strand (turquoise rings) completing the duplex is contributed by a symmetry mate SapA<sub>closed</sub> molecule not shown in the diagram. The pitch distance of the RNA helix is indicated. B, topology diagram of SapA. C, comparison of the closed (light cyan) and open (wheat) conformations of SapA. Rotation angle around axis was calculated and visualised with the program *ProSMART*. The distances between the C $\alpha$  atoms of residues Asp 222 and Valine 435 are indicated in the closed and open conformations.

<https://doi.org/10.1371/journal.pone.0256070.g001>

the size of peptides accepted by the binding site. For instance, *L. lactis* OppA (*LI*OppA), which can bind peptides of up to 35 residues, has a very large cavity (4900 Å<sup>3</sup>), whereas dipeptide binding *Ec*DppA has a much smaller cavity (700 Å<sup>3</sup>). The SapA cavity volume is even smaller again at approximately 400 Å<sup>3</sup>, which implies that SapA can only bind a small ligand such as a single amino acid or dipeptide (Fig 3B). That SapA is likely to bind a ligand of this type is also

Table 1. Crystallographic data collection and refinement statistics.

Crystal form	SapA <sub>closed</sub>	SapA <sub>mixed</sub>	SapA <sub>heme</sub>
PDB ID	7OFZ	7OG0	7OFW
<b>Data Collection</b>			
Wavelength (Å)	0.9200	0.9174	0.9763
Resolution range (Å) <sup>a</sup>	72.2–2.62 (2.71–2.62)	65.74–2.61 (2.70–2.61)	71.36–3.15 (3.26–3.15)
Space group	P 4 21 2	P 21 21 2	P 4 21 2
Unit cell (Å, °)	144.4 144.4 62.0 90 90 90	143.3 148.0 59.6 90 90 90	142.7 142.7 60.6 90 90 90
Total reflections <sup>a</sup>	38352 (3840)	144902 (14854)	116349 (11497)
Unique reflections <sup>a</sup>	19270 (1929)	38143 (3848)	11330 (1111)
Multiplicity <sup>a</sup>	2.0 (2.0)	3.8 (3.9)	10.3 (10.3)
Completeness (%) <sup>a</sup>	94.9 (96.9)	96.9 (98.7)	99.9 (99.8)
Mean I/σI <sup>a</sup>	18.2 (1.1)	11.2 (2.1)	16.9 (3.3)
R-merge <sup>a</sup>	0.031 (0.600)	0.089 (0.623)	0.128 (1.029)
R-meas <sup>a</sup>	0.044 (0.848)	0.104 (0.721)	0.135 (1.084)
CC1/2 <sup>a</sup>	0.999 (0.571)	0.997 (0.698)	0.998 (0.845)
<b>Refinement</b>			
Reflections used in refinement <sup>a</sup>	19267 (1929)	38142 (3848)	11326 (1110)
R-work <sup>a</sup>	0.218 (0.318)	0.199 (0.286)	0.239 (0.341)
R-free <sup>a</sup>	0.263 (0.381)	0.249 (0.308)	0.276 (0.366)
Number of non-hydrogen atoms	4527	8917	4395
Protein	4005	8030	3963
RNA/ions	407	728	387
heme	-	-	43
solvent	114	157	2
RMSD bond lengths (Å)	0.014	0.014	0.01
RMSD bond angles (°)	1.8	1.4	1.5
Ramachandran favored (%)	96.1	97.3	96.5
Ramachandran allowed (%)	3.7	2.0	3.1
Ramachandran outliers (%)	0.2	0.7	0.4
Clashscore	3.3	4.8	4.6
<b>Average B-factors (Å<sup>2</sup>)</b>			
all	69.6	61.4	98.2
macromolecules	65.3	55.7	94.1
RNA/ions	116.0	129.4	138.9
heme	-	-	116
solvent	54.1	35.9	41.2

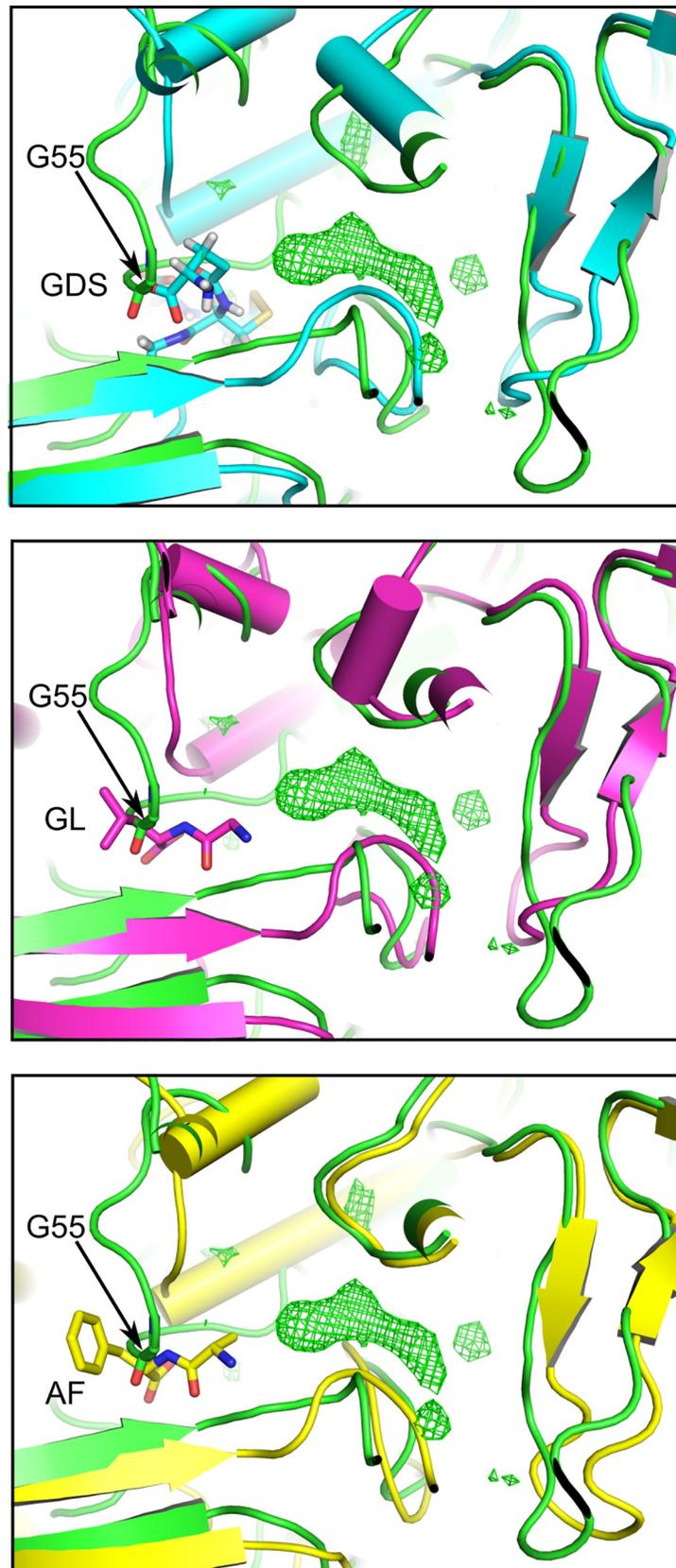
<sup>a</sup>Values in parentheses refer to the highest resolution shell.

<https://doi.org/10.1371/journal.pone.0256070.t001>

predicted by phylogenetic analysis of structurally characterized SBP Cluster C members. Protein sequences of Cluster C members as described by [31], together with the sequences of NTHiSapA and PsDppA, were used to construct a cladogram of sequence relationships (Fig 4). SapA falls within a clade of proteins functionally described to bind either dipeptides or glutathione (a tripeptide).

### SapA binds heme with low affinity

Initial qualitative assessment of hemin/SapA interaction was carried out by Native PAGE assay (Fig 5A). The results are strongly suggestive of complex formation at high hemin

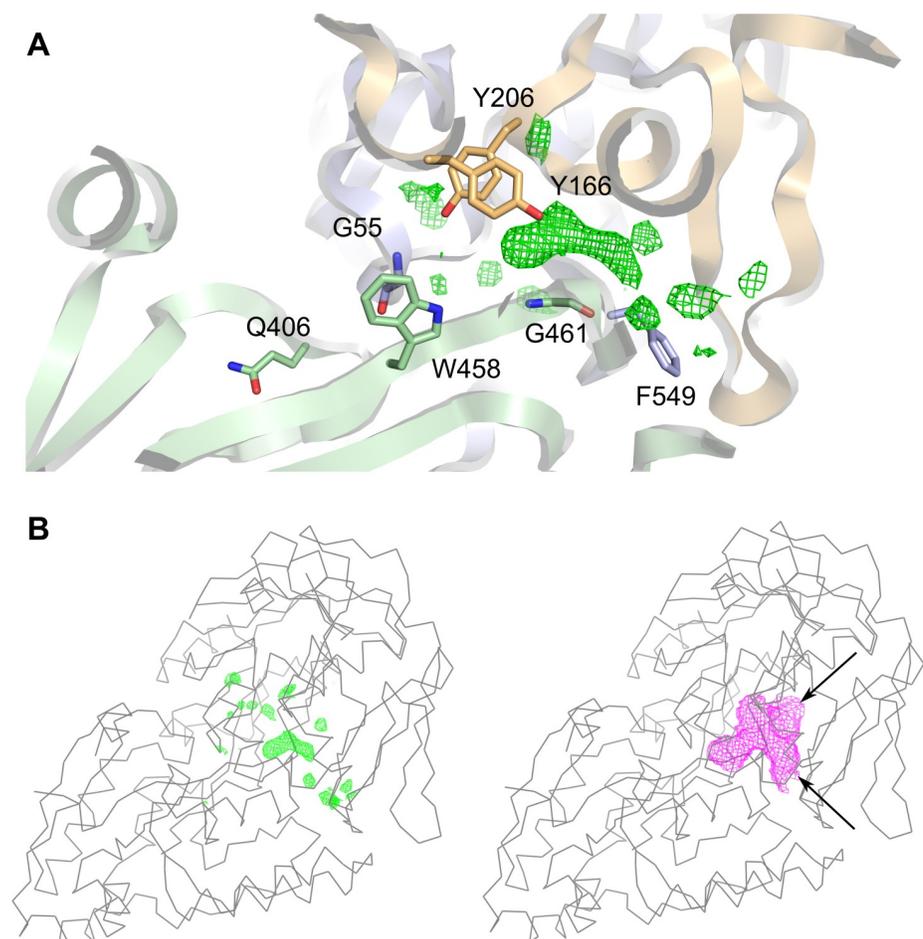


**Fig 2. Superimposition of the NTHi SapA<sub>closed</sub> structure (green) with its closest structural homologs.** A, *HpHbpA* (cyan, PDB = 3M8U); B, *EcDppA* (magenta, PDB = 1DPP); C, *PsDppA* (yellow, PDB = 4QFL). The dipeptide (GL,AF) and oxidised glutathione disulphide (GDS) ligands are shown in stick representation. The green mesh represents the difference density *F<sub>o</sub>-F<sub>c</sub>* within the ligand-binding cavity of NTHi SapA<sub>closed</sub> (contoured at 3 $\sigma$ ).

<https://doi.org/10.1371/journal.pone.0256070.g002>

concentration. Under the conditions of the assay, the interaction did not appear to be saturated before the limit of the heme solubility was reached.

Hemin was also soaked into SapA crystals. From these a structure was obtained, SapA<sub>heme</sub> (PDB = 7OFW). As with SapA<sub>closed</sub>, SapA<sub>heme</sub> displayed difference density in the central ligand binding pocket. However, the pocket remains too small (approximately 400 Å<sup>3</sup>) to accommodate a substantial ligand such as heme, which occupies a volume closer to 800 Å<sup>3</sup>. Furthermore, additional difference density, corresponding to heme, was present on the protein surface. The bound heme forms a crystal contact between two symmetry-related SapA molecules, each coordinating the ligand via a surface exposed histidine (His 306 in both cases) (Fig 5B).



**Fig 3. The SapA ligand-binding cavity.** A, a detailed view of the SapA<sub>closed</sub> ligand binding cavity. Aromatic residues lining the cavity are shown in stick representation together with the other residues discussed in the text. Unexplained *F<sub>o</sub>-F<sub>c</sub>* residual density is represented as a green mesh and is contoured at 3 $\sigma$ . B, left: The same *F<sub>o</sub>-F<sub>c</sub>* density highlighted in the context of the overall view of SapA<sub>closed</sub> ( $\alpha$  trace). Right: The ligand-binding cavity of SapA<sub>closed</sub> as calculated by the program *Voidoo* (magenta mesh). Arrows indicate cavity openings to the protein surface.

<https://doi.org/10.1371/journal.pone.0256070.g003>

**Table 2. Residues in the putative ligand-binding cavity.** Neutral or hydrophobic residues are underlined.

Domain	Residue Range	Cavity residues
domain 1b	(loop 55:68)	<u>gly55</u> , <u>ser57</u> , <u>met58</u> , <u>asn59</u> , <u>val60</u> , glu63, lys68
domain 1a	(loop 71:72)	asp71, <u>ile72</u>
domain 1a	(loop 166)	<u>tyr166</u>
domain 1a	(loop 200:206)	<u>ser200</u> , <u>ala203</u> , <u>ser204</u> , <u>gln205</u> , <u>tyr206</u>
domain 2	(loop 458:463)	<u>trp458</u> , <u>leu459</u> , <u>ala460</u> , <u>gly461</u> , <u>asn462</u> , <u>leu463</u>
domain 1b	(loop 547:552)	<u>thr547</u> , <u>phe549</u> , <u>gly550</u> , <u>ser551</u> , <u>leu552</u>

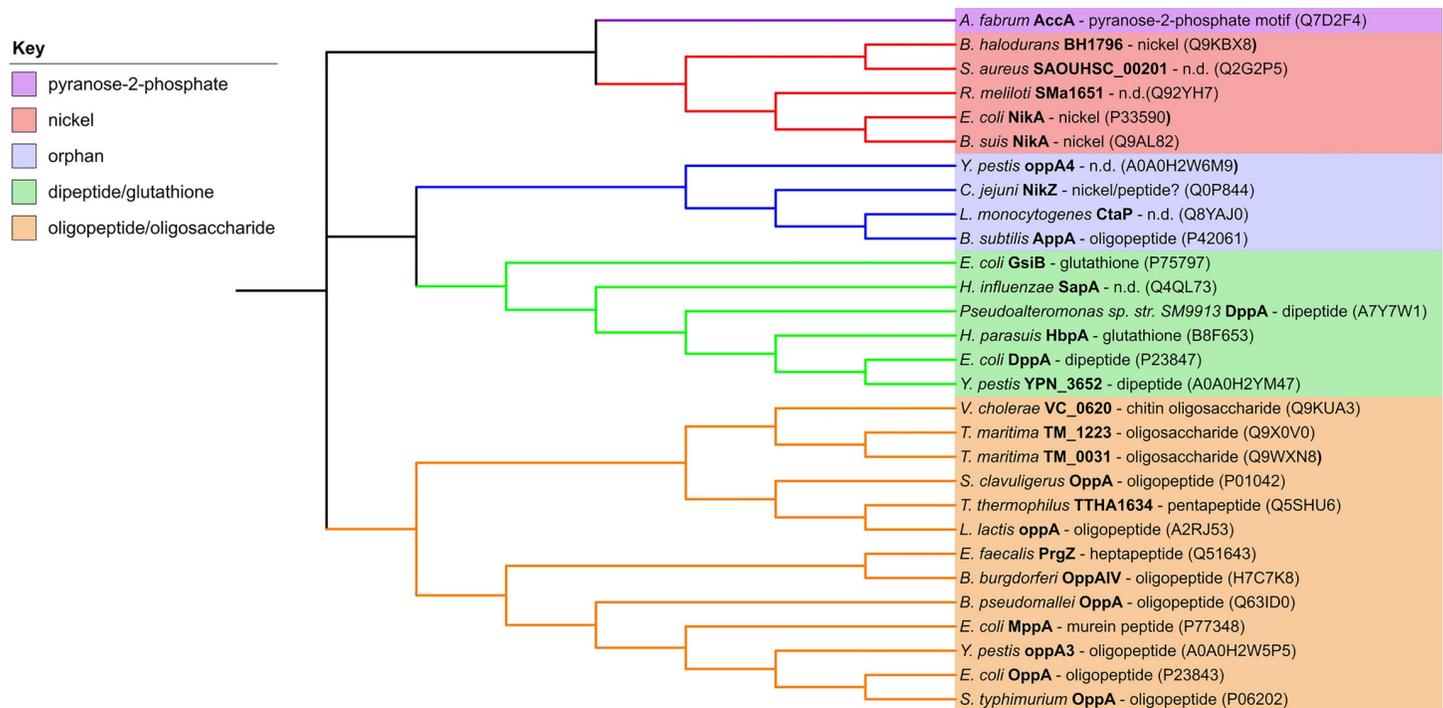
<https://doi.org/10.1371/journal.pone.0256070.t002>

Isothermal titration calorimetry (ITC) was performed to quantitatively assess the binding affinity. A sigmoidal binding curve for this reaction was not observed for this reaction due to the binding affinity. Therefore, the binding stoichiometry of the reaction was fixed to 1, allowing a  $K_d$  of  $282 \pm 18.3 \mu\text{M}$  to be proposed (Fig 5C). Although the heme bound in the SapA<sub>heme</sub> crystal structure is present at a crystal contact, providing a stoichiometry of 0.5, there are no further interactions between the protein molecules. Therefore, there is no reasonable basis to suggest that the 2 SapA:1 heme complex shown in Fig 5B is a physiologically relevant dimer. Control experiments indicated that the measured heats were not caused by dilution or hemin aggregation.

### SapA interacts with dsRNA

SapA co-crystallized with a strand of endogenously acquired RNA coordinated on the opposite face to the binding cleft. Identical RNA binding was observed at the same site in both SapA<sub>closed</sub> and SapA<sub>open</sub> and therefore did not appear to affect the overall conformation of the protein. The RNA crystallized as a double helix with parameters consistent for standard A-form dsRNA [34], with a radius of  $\sim 12 \text{ \AA}$  and pitch of  $31 \text{ \AA}$  with 11 bp per turn and a rise per bp of  $2.8 \text{ \AA}$  ( $31 \text{ \AA}/11 \text{ bp}$ ). The antisense strand is provided by a SapA symmetry mate. Furthermore, although definitive electron density is only present for a 19 bp stretch RNA, inspection of the symmetry neighbors indicates that the RNA strands extend across asymmetric units, forming a matrix within the crystal. The RNA backbone is coordinated to positively-charged chemical groups on the SapA surface. Specifically, the guanidino group of Arg 101 forms a hydrogen bond with the ribose 2' hydroxyl and the 3' oxygen. Additionally, Gln 85 interacts with the O4 ring oxygen and 2' hydroxyl of neighboring ribose residues (Fig 6A). APBS electrostatics analysis [35] of the SapA surface revealed that Arg 101 contributes to a wider positively charged patch formed by lysine residues—Lys 73, 258, 259, 262, 542 and 545 (Fig 6B). This area provides a counter-charge to the negatively charged RNA molecule. Interestingly, both Gln 85 and Arg 101 are located within the conserved SBP family 5 motif—(LIVM) AX<sub>2</sub>(W)X<sub>1</sub> or <sub>2</sub>(SN)(KE)DX<sub>4</sub>T(FY)X(LIV)RX<sub>3</sub>K - which in NTHi SapA encompasses residues 84–106 of domain 1b. This motif represents a signature sequence within a system of classification based on sequence similarity where most SBPs can be grouped into 8 families or clusters. Family 5 nominally includes binding proteins for peptides and nickel [36] such as *E. coli* oligopeptide, dipeptide, murein, and nickel binding proteins OppA, DppA, MppA, and NikA. Besides SapA in *H. influenzae* it includes the periplasmic lipoprotein HbpA.

The ribonucleotides were modeled as C-G pairs because at  $2.6 \text{ \AA}$  resolution it was not possible to assign the exact sequence of individual bases. The density accounting for the RNA showed signs of possible density averaging between the two stands suggesting that heterogeneous RNAs bind this site.



**Fig 4. Cladogram of sequence relationships between structurally characterized SBP cluster C members.** Clades organize by distinct functions that are colored according to the key.

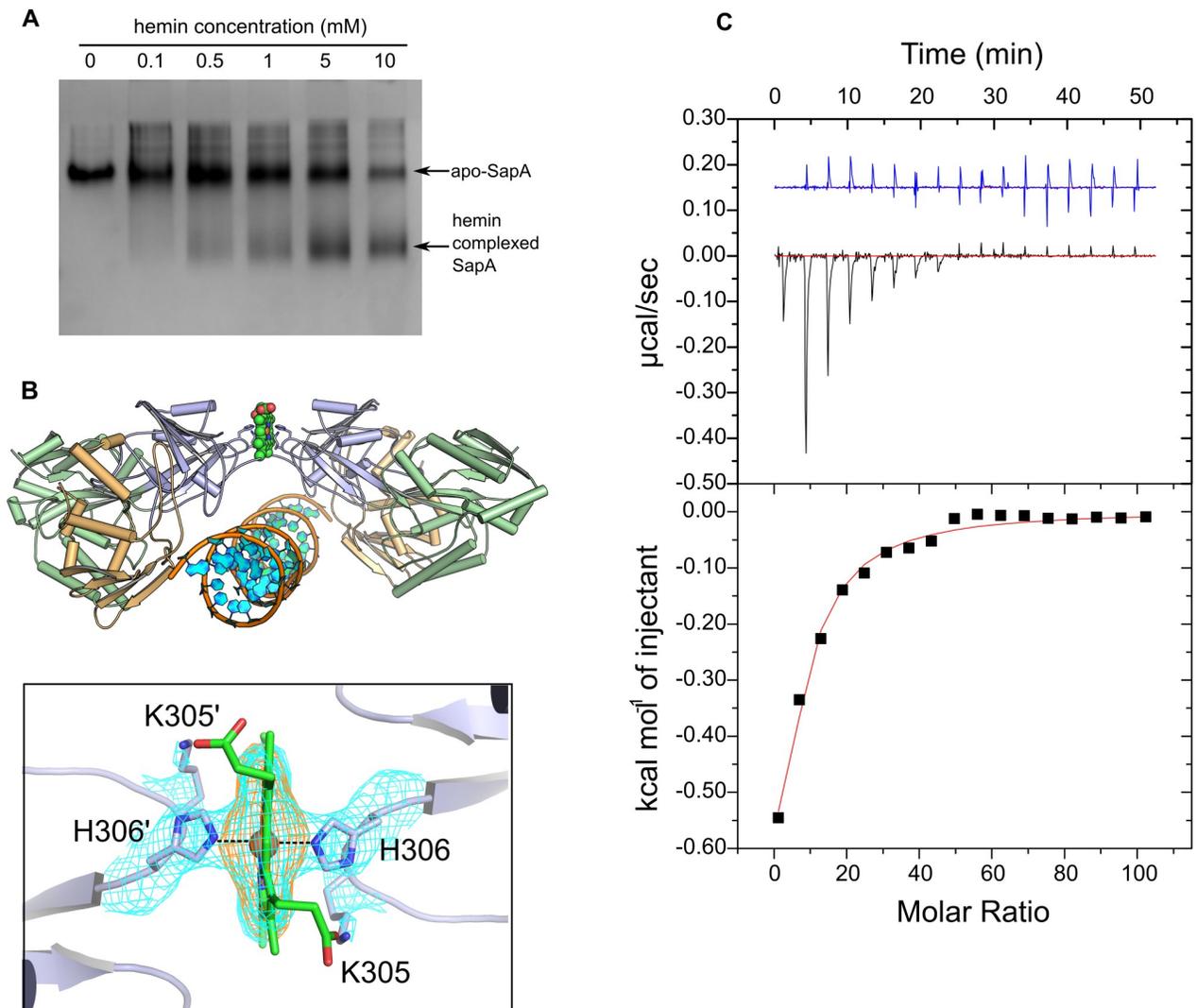
<https://doi.org/10.1371/journal.pone.0256070.g004>

To determine the specificity of Arg 101 and Gln 85 in binding RNA, single (Arg101Ser) and double (Gln85Ser and Arg101Ser) serine mutants were generated and tested by agarose gel electrophoretic shift assay (EMSA). Here the wild-type and mutant proteins were assayed for interaction with DH5 $\alpha$  *E. coli* ribosomal RNA, as measured by the decrease in intensity of 23S and 16S rRNA bands at increasing protein concentrations. The double mutant showed the greatest reduction in binding relative to the wild-type, whilst the single mutant showed an intermediate effect (Fig 6C, S2A Fig). Dissociation constants ( $K_d$ ) could be estimated from these experiments as 4.4  $\mu$ M for wild-type, 7.7  $\mu$ M for single and 23.0  $\mu$ M for double mutant SapAs.

The mutants were also tested for DH5 $\alpha$  *E. coli* genomic DNA binding. The interaction was only observed at the highest protein concentrations tested (41  $\mu$ M) with no significant difference between the wild-type and mutant proteins (S2B Fig). Overall, these results indicate that Arg 101 and Gln85 are responsible for the specific binding to RNA rather than contributing to a general electrostatic effect by providing a counter charge to the anionic nucleic acids.

## Discussion

SapA was successfully crystallized in an open (no ligand) and closed conformation with an endogenous ligand in the binding cavity. The cavity volume of SapA<sub>closed</sub> and the unexplained electron density within it can potentially accommodate a small ligand such as a short peptide or even an extended polypeptide chain that could protrude out of the narrow openings of the SapA ligand-binding cavity. However, there is not enough cavity volume to accommodate AMPs in their folded state. Additionally unfolding of BDs would require disruption of disulfide bridges. SapA<sub>closed</sub> has a cavity volume of 400  $\text{\AA}^3$  whereas LL-37 and hBDs have volumes ranging from 3000  $\text{\AA}^3$  to 8500  $\text{\AA}^3$ . As discussed previously, the SapA binding cavity is formed predominantly by hydrophobic or neutral residues, providing no obvious countercharges for

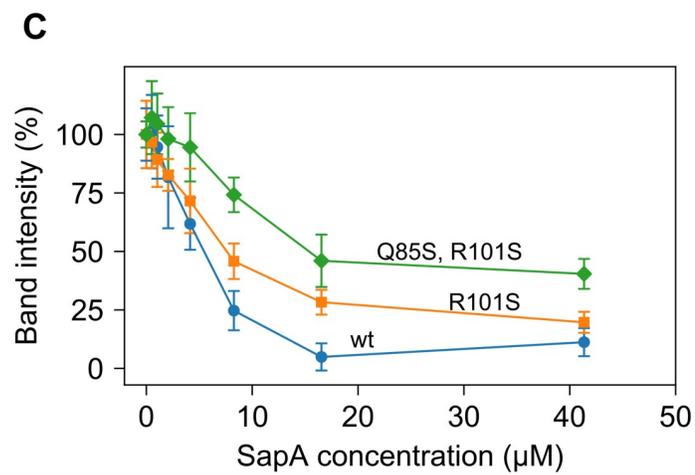
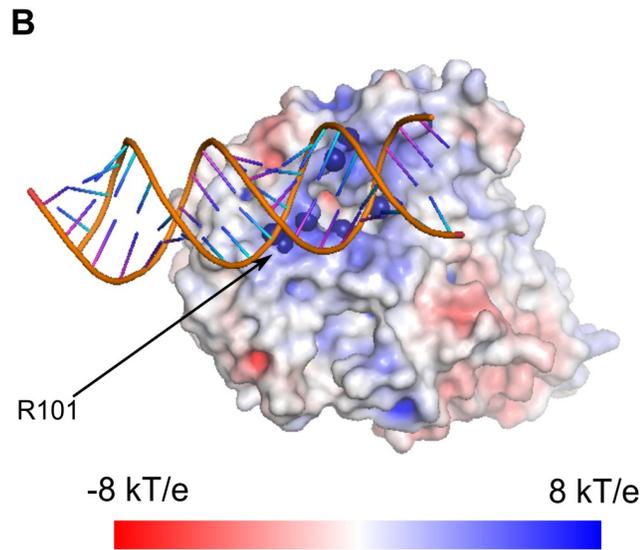
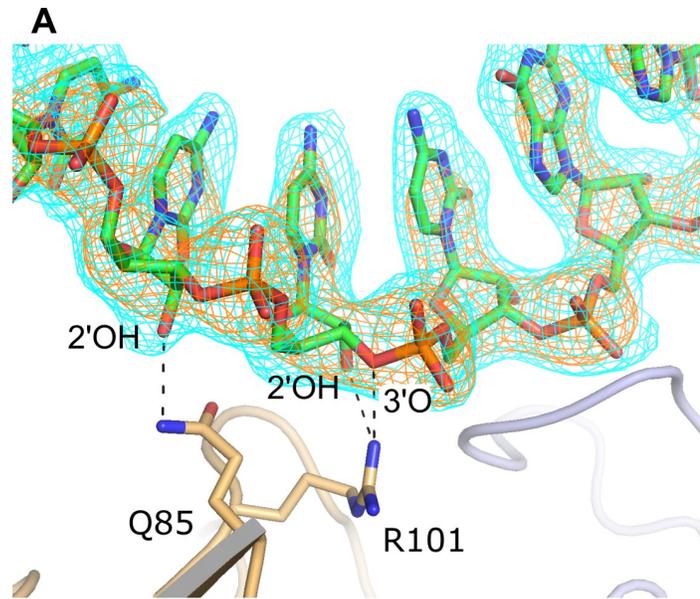


**Fig 5. SapA interaction with heme.** A, native gel assay for heme interaction with SapA: Lanes 1–6 correspond to increasing concentrations of Hemin incubated with 5 µg of SapA. B, Soaking of heme into SapA crystals. Top panel: a spherical representation of a heme molecule coordinated on the interface of two symmetry-related copies of the SapA<sub>heme</sub> monomer. Bottom panel: the heme iron (orange sphere) is coordinated by His 306 (dashed line) from two symmetry mates. The meshes represent 2*Fo*-*Fc* (cyan) and *Fo*-*Fc* (orange) maps contoured at 1 and 3.5σ respectively. C, isothermal titration calorimetry; Top panel: The experimental trace (black) obtained upon injection of heme into SapA and into buffer control (blue). The baseline is indicated in red. Bottom panel: The integrated heats upon injection (black squares) and the data fit (red line) after subtraction of the control data.

<https://doi.org/10.1371/journal.pone.0256070.g005>

the reported cationic AMP ligands of SapA (Fig 3A, Table 2). We found no evidence for hBD1, hBD2, hBD3 and LL-37 binding to highly purified SapA protein by crystallographic or biophysical methods (ITC, thermofluor; S1 File Supporting methods).

We suggest the most likely ligand occupying the binding cavity of SapA<sub>closed</sub> is a hydrophobic di/tripeptide based on overlap predictions from cavity size volume, the hydrophobic nature of the cavity and phylogenetic analysis. We attempted many experiments to confirm binding of dipeptides, however were unable to do so, probably due to some of our sample already being in a ligand occupied state preventing further interaction. Our efforts to devise a denature/refold protocol followed by ligand screening were unsuccessful and additionally complicated by low SapA expression levels.



**Fig 6. SapA interaction with RNA.** A, residues in the Ia domain of SapA involved in hydrogen bonding (black dashed lines) to the RNA helix backbone. The meshes represent 2Fo-Fc (cyan) and Fo-Fc (orange) maps contoured at 1 and 2.5 $\sigma$  respectively. B, the surface of SapA coloured by electrostatic potential. NH1 and NH2 atoms of Arg 101 and NZ atoms of Lysines 73, 258, 259, 262, 542 and 545 are visualised as blue spheres. Electrostatic potentials (units of kT/e from -8 to 8) were calculated by APBS and visualized on the solvent-accessible surface by the program PyMOL (Version 1.5.0 Schrödinger, LLC). C, Agarose gel electrophoretic mobility shift assays: 0.5  $\mu$ g of genomic *E. coli* DH5 $\alpha$  RNA was incubated with increasing concentrations of wt SapA (blue line), single mutant SapA R101S (orange line) and double mutant SapA R101S, Q85S (green line). The intensities of the 23S and 16S rRNA bands were measured at each protein concentration. (See S2A Fig for typical gel data).

<https://doi.org/10.1371/journal.pone.0256070.g006>

It should be noted that while NTHiSapA organizes into a functional clade with di/tripeptide binding SPBs, there is poor conservation of ligand-binding residues with its homologs *PsDppA* (33% identity), *EcDppA* (30% identity) and *HpHbpA* (31% identity) (S3 Fig). In the DppA SBP family, strongly conserved binding site aspartate and arginine residues (Asp 436, Arg 383 in *EcDppA*, Asp 432, Arg 379 in *HpHbpA*) confer specificity for peptides by coordinating the dipeptide terminal carboxyl and amino moiety respectively. However, in SapA, neither are conserved, with the aspartate replaced by Gly 461 and the arginine replaced by Gln 406, with the Gln 406 side chain oriented out of the binding site (Fig 3A). In SapA, the tip of the loop centered on Gly 55 replaces the approximate volume occupied by the ligands in DppA and HbpA, effectively bisecting the binding cleft (Fig 2). As a result, the endogenous SapA ligand lies shifted along the cleft relative to the DppA and HbpA ligands, in a cavity that is much more reduced in volume. The possibility that SapA binds something other than a short peptide therefore cannot be ruled out. Cluster C SBPs, albeit more phylogenetically distant ones, also bind nickel, longer oligopeptides, oligosaccharides and pyranose-2-phosphate containing compounds (Fig 4).

We can confirm that NTHiSapA directly binds heme in agreement with other studies [17, 19]. In addition to native gel and ITC evidence we provide structural evidence for heme binding on the surface of the SapA molecule. This surface location is far from the ligand binding cavity and out of the context of any defined pocket and may therefore cast some doubt on the physiological relevance of this interaction. Nevertheless, at the very minimum, we can conclude that a surface residue His 306 has the capacity to coordinate heme to the SapA molecule. It is the first molecular-detail experimental structural evidence for heme binding to SapA or indeed any Cluster C SBP. SapA homologs *EcDppA*, *EcMppA* and NTHiHbpA have also been reported to bind heme but without direct structural evidence. Consequently, heme site predictions have relied on computational docking [17–19, 37]. Our observation that the heme site is physically separate from the ligand binding cavity in NTHiSapA correlates well with the results of a SPR competition assay carried out with NTHiOppA [19]. Here the authors showed that heme does not directly compete with peptide in the substrate-binding pocket. Their second prediction that, while separate, the two sites are in close proximity is not supported by our data. This discrepancy may be due to the limitations of computational docking and in the case of SapA-heme modelling [18] exacerbated by the previous unavailability of a crystal structure.

Questions remain surrounding the physiological relevance of RNA binding to SapA. The  $K_d$  of the RNA interaction is approximately 4.4  $\mu$ M, 60 fold tighter than the interaction measured here with heme. This  $K_d$  falls within the range of measured binding affinities reported for 73 structurally characterized RNA-protein complexes [38]. Additionally, the RNA interaction is within a conserved sequence motif, which can be indicative of a functional role [39]. While there are no reports of RNA specifically localizing to the periplasm, RNA may conceivably exist in this compartment during processes such as viral infection and during the recently described bacterial secretion of small non-coding RNAs [40]. Under these circumstances, periplasmic SapA could encounter RNA. We can only speculate on what would be the function of

such an interaction. Perhaps RNA binding motifs present on bacterial periplasmic proteins can act as "sticky patches" to prevent or slow down the entry of viral RNA as a mechanism of bacterial resistance to viral infection.

RNA contamination of the SapA protein may explain previous observations regarding AMP binding. Cationic AMPs interact with a variety of anionic macromolecules including nucleic acids [41–45]. Therefore, interactions previously observed between SapA and AMPs [14] may in fact be via interactions with contaminating RNA bound to the SapA sample, with the RNA acting as the bridging molecule. This model would explain the results of a competition assay [17] where the ability of different AMPs to displace heme from SapA was correlated to their overall positive charge. Such observations are consistent with charge-charge interactions between molecules rather than a sequence-specific interaction characteristic of peptide/protein binding. Such effects were not apparent in our experiments as we ensured that our SapA samples avoided RNA contamination. The  $A_{260}/A_{280}$  ratio of purified SapA samples was monitored; and a measured value of ~0.6 was taken to be consistent with a nucleic acid free sample. In this work, we utilized a number of techniques using highly purified components to characterize SapA and the molecular interactions with its proposed ligands. The weight of our evidence favors the involvement of SapA, and its cognate transporter, in the transport of dipeptides or tripeptides, even though further research will be required to confirm this on a functional level. Research should take two distinct directions: one, to confirm the true substrate of the transporter and two, to elucidate the mechanisms of AMP resistance. That AMP sensitivity may not be mediated through direct molecular interactions with Sap needs to be considered, given that *sapA* sensitivity to AMPs is not conserved across species nor is it confined to a particular set of AMPs. For instance, the *H. ducreyi sapA* mutant was more sensitive to LL-37 but not human defensins [22] while resistance to LL-37 in *E. coli* was unaffected by  $\Delta sapBCDF$  deletion [46]. In closing, the data presented here provide the basis for the community to cast a wider net and consider other mechanisms and pathways that could explain the observed heme requirement and AMP sensitivity of NTHi *sap* genetic mutants in cell assays and *in vivo*.

## Materials and methods

### NTHi SapA protein expression and purification

Purified NTHi 86-028NP genomic DNA (gift of Kevin Mason, Nationwide Children's Hospital, Columbus, Ohio, USA) served as a template for PCR amplification of the SapA protein-coding region. The In-Fusion® cloning method was used to insert the PCR product into a pOPINF plasmid vector [47]. The construct was designed to exclude the predicted signal peptide region and encompassed residues 33 to 560 with a S3C cleavage site. The recombinant plasmid was transformed into *E. coli* Rosetta™ (DE3) expression strain (Novagen). A single colony was used to inoculate 100 ml of Power Prime Broth™ media (Molecular Dimensions) supplemented with 50 µg/ml carbenicillin, 35 µg/ml chloramphenicol and grown overnight at 37°C, 230 rpm. 10 ml of the overnight culture was then used to inoculate 1 liter of Overnight Express™ Instant TB autoinduction media (Novagen). The cultures were initially grown at 37°C for 5–6 h, followed by prolonged growth at lowered temperatures (230 rpm, 20 h, 25°C). The cells were harvested by centrifugation and stored at -80°C. Approximately 50 g of cells were resuspended in lysis buffer (50 mM Hepes pH 8, 500 mM NaCl, 30 mM imidazole pH 8, 0.2% Tween-20, 5% (w/v) glycerol) supplemented with DNase I (10–20 µg/ml) and a cOmplete™ EDTA-free Protease Inhibitor Cocktail tablet (Roche), and lysed using a Constant Systems Ltd. cell disrupter (3 passes, 30 kpsi, 4°C). The crude extract was centrifuged (50000 g, 1hr, 4°C) and filtered with a 0.22 µm filter to remove insoluble components.

The soluble lysate was loaded onto a 1 ml nickel sepharose HisTrap FF column (GE Healthcare) equilibrated in wash buffer (50 mM Hepes pH 8, 500 mM NaCl, 30 mM imidazole pH 8, 5% (w/v) glycerol). The bound protein was washed using 20 column volumes of wash buffer and eluted using an Imidazole step gradient using the wash buffer containing 500 mM Imidazole.

The major peak from Ni-affinity purification was collected and applied to a HiLoad 16/600 Superdex 200 pg size-exclusion column (GE Healthcare) equilibrated in 20 mM Hepes pH 8, 500 mM NaCl and 5% (w/v) glycerol. Post size-exclusion, the protein was subjected to (His)<sub>6</sub> tag cleavage by incubation with 50% (w/w) Human Rhinovirus B 3C protease (2 h at 20°C then 14 h at 4°C), followed by reverse His-tag purification.

Purity at all stages was monitored by SDS-PAGE. High purity samples were pooled and concentrated to 10 mg/ml for crystallization using a 10 kDa MWCO Amicon-Ultra centrifugal filter unit (Merck). In parallel, the buffer was exchanged on the unit (final buffer: 20 mM Hepes pH 8, 500 mM NaCl). The molecular weight of the purified protein was confirmed by intact mass spectrometry (Predicted Mass / Experimental Mass = 60460.64 / 60458.29 Da).

### Crystallisation, data collection and structure determination

Diffraction quality crystals (SapA<sub>closed</sub>) were obtained from 200 nl sitting drops where the protein was mixed 1:1 with 22% (w/v) PEG 3350, 0.25 M NaBr, 0.1 M Bis-Tris Propane pH 7.5.

In an attempt to obtain RNA-free crystals the above procedure was modified by adding 0.5 ml of RNase Cocktail™ Enzyme Mix (Ambion) to the clarified and filtered cell lysate. This was incubated on ice for 1 hour prior to Nickel sepharose purification. A second crystal form, SapA<sub>mixed</sub>, was subsequently obtained from this protein preparation, in a crystallization condition containing 10% (w/v) PEG 8000, 0.1 M imidazole pH 8.0, 0.2 M Ca(OAc)<sub>2</sub>. Heme derivative SapA<sub>heme</sub> crystals were obtained by soaking closed-form crystals for 48 h in a mother liquor solution containing Hemin added to a theoretical concentration of 25 mM. Hemin was added from a freshly prepared unfiltered suspension generated by the addition of hemin to 0.5 M NaOH and subsequent neutralization to pH 8. A theoretical concentration of 25 mM hemin was also maintained in the cryoprotectant solution.

The crystals were then rapidly transferred from the sitting droplet into a cryoprotectant solution, the composition of which was based on the original crystal growth condition, and either 20% (v/v) glycerol or ethylene glycol. The crystals were then immediately plunged into liquid nitrogen. X-ray diffraction data were collected at 100 K at beamline I04 and I04-1, Diamond Light Source (Didcot, UK). Data were processed using the xia2 automated reduction pipeline [48] which makes use of Mosflm [49], Pointless [50], CCP4 [51] and XDS [52].

The structure of SapA was solved by molecular replacement by the programs BALBES [53] and Phaser [54] in the CCP4 suite of programs. The molecular replacement pipeline BALBES identified the crystallographic structure 3M8U as the best search model. Initial model auto-building and refinement was carried out in PHENIX [55]. Further rounds of manual building and refinement were carried out using the programs *Coot* [56], Refmac5 [57], BUSTER [58] and the PDB-REDO server [59]. The quality of the final model was assessed with the PDB Validation Services and MolProbity [60]. Domain movements were analyzed using ProSMART [61] by aligning the structure of SapA<sub>closed</sub> with chain B of SapA<sub>mixed</sub>. Structural alignment with SapA homologs was carried out using the SSM superpose tool within the program *Coot*. Sap cavity volume calculations were carried out using the program Voidoo [62] using the same parameters as described by [33]. Identification of residues lining the cavity was carried out with the same program.

## Phylogenetic analysis

The phylogenetic analysis used structurally characterised SBP cluster C members as described in Scheepers *et al* with NTHiSapA and PsDppA added to the set. UniProt derived sequences for this set were aligned with Clustal Omega [63]. Phylogenetic neighbour joining tree data were obtained from within Clustal Omega and the corresponding cladogram was visualised using iTOL [64] software.

## Native PAGE gel shift assays

Samples for native PAGE were prepared by mixing 5 µg of purified SapA protein with 0, 0.1, 0.5, 1, 5, 10 mM hemin from a freshly prepared concentrated hemin stock. The volumes were adjusted to a final volume of 20 µl by the addition of 20 mM Hepes pH 8, 300 mM NaCl buffer. The samples were then incubated for 45 min at room temperature, prior to separation by native PAGE (10% gel, running buffer pH 8.8) as previously described [65, 66].

## Isothermal titration calorimetry

Isothermal titration calorimetric measurements were carried out using a MicroCal iTC200 microcalorimeter (Malvern Instruments UK) at 25°C. 200 µL of SapA (20 µM) in was placed in the cell and 40 µL of hemin (10 mM) in the syringe. The buffer used in ITC was 20 mM Hepes pH 8, 300 mM NaCl. The concentration of SapA and hemin solutions were determined using spectrophotometry at 280 and 385 nm, and the extinction coefficients used were  $\epsilon_{280} = 77490 \text{ M}^{-1}\text{cm}^{-1}$  and  $\epsilon_{385} = 58440 \text{ M}^{-1}\text{cm}^{-1}$ , respectively. Sixteen 2.4 µL injections were performed at an injection speed of 0.5 µL/sec, with a pre-injection of 0.5 µL, a three-minute interval between injections and a stirrer speed of 750 rpm. To establish the heat of dilution, a control experiment was performed where hemin (10 mM) was injected into the ITC buffer using identical experimental conditions. This was then subtracted from the main experiment. Data were analyzed using MicroCal Origin software (version 7) fitting to a single site binding model. The binding enthalpy ( $\Delta H$ ) and association constant ( $K_a$ ), were permitted to float during the least-squares minimization process and taken as the best-fit values. The binding stoichiometry ( $n$ ) was fixed to one.

## RNA and DNA electrophoretic mobility shift assay (EMSA)

SapA R101S and SapA Q85S R101S mutants were generated for the purpose of testing in EMSAs. The double-stranded DNA encoding the mutant proteins was purchased as gBlocks® Gene Fragments (Integrated DNA Technologies) and cloned into pOPINF plasmids by the In-Fusion® method. Purification was carried out as for the wild-type protein.

0.5 µg of genomic total RNA from *E. coli* DH5α (Ambion) was incubated with increasing amounts of wild-type SapA, SapA R101S and SapA Q85S R101S mutants in a total volume of 20 µl of TBE at room temperature for 1 hour. Following incubation, the samples were loaded on a 1% (w/v) agarose/TBE gel containing SYBR™ Safe Gel Stain (Invitrogen) at 1:10000 dilution. The samples were then separated at 60 V for 1 hour in 1X TBE buffer. The gels were then imaged and bands were quantified using the ImageQuant TL 1D v8.1 software. The background subtracted intensities for rRNA 16S and 23S bands were summated, normalized against the average intensity at zero concentration and plotted. The experiments were repeated 4 or 5 times. The  $K_d$  of an interaction was estimated by least squares linear regression of the normalized reduction in rRNA band intensity using Origin (version 7) fitting to the

equation:

$$PR = \frac{Kd + [P] + [R] - \sqrt{(Kd + [P] + [R])^2 - 4[P][R]}}{2}$$

where PR is the normalized reduction in rRNA band intensity, and [P] and [R] are the concentration of the SapA protein and the rRNA, respectively.

Total genomic DNA was isolated from Subcloning Efficiency™ DH5α™ Competent Cells (Invitrogen) using the GenElute™ Bacterial Genomic DNA Kit (Sigma) according to the manufacturer's instructions. The genomic DNA was then concentrated by vacuum concentration (to approximately 200 ng/μl) and dialyzed into water using a 3500 MWCO Slide-A-Lyzer MINI Dialysis Device (Pierce). The assays were carried out as above with the following modifications: Protein samples were diluted in TAE and incubated with 0.5 μg of gDNA for 1 hour. Electrophoresis was carried out at 75 V for 30 minutes on a 0.8% (w/v) agarose/TAE gel.

## Supporting information

### S1 File. Supporting methods: Expression and purification of human beta defensins and isothermal titration calorimetry of AMPs.

(DOCX)

**S1 Fig. Surface representations of SapA structures.** A, SapA<sub>closed</sub>. B, SapA<sub>mixed</sub>. C, SapA<sub>heme</sub>. RNA and heme are shown in sphere representation and protein chains A and B are colored gray and yellow respectively.

(DOCX)

**S2 Fig. Typical EMSA results for the interaction of SapA with nucleic acids on agarose gels.** A, DH5α *E. coli* gRNA incubated with increasing concentrations of wt SapA and mutants of SapA. Arrows on the left mark the position of rRNA bands and asterisks indicate starting well position B, DH5α *E. coli* gDNA incubated with increasing concentrations of BSA, wt SapA and SapA mutants.

(DOCX)

**S3 Fig. Alignment of amino acid sequences of NTHi SapA (UNIPROT = Q4QL73) with its closest structural homologs *HpHbpA* (UNIPROT = B8F653), *EcDppA***

**(UNIPROT = P23847) and *Pseudoaltermonas* sp. SM9913 DppA (UNIPROT = A7Y7W1).**

Non-conserved NTHi SapA residues discussed in the main text are highlighted in solid green. α-helices are displayed as squiggles respectively. β-strands are rendered as arrows, strict β-turns as TT letters. Figure generated with *ESPrpt 3.0* [67].

(DOCX)

**S1 Raw images.**

(PDF)

## Acknowledgments

The authors would like to acknowledge the staff at the DLS synchrotron for crystallographic data collection and Kevin Mason of Nationwide Children's Hospital, Columbus, Ohio, USA for the kind gift of NTHi genomic DNA and initial discussions. We would like to thank Derek Hood, Pratik Vikhe of MRC Harwell and Claire Strain-Damerell, Kyle Dent, Tracy Keates, Hannah Glover (DLS) for many insightful discussions and for critically reading the manuscript.

## Author Contributions

**Conceptualization:** Petra Lukacik, Panagis Filippakopoulos, Carol V. Robinson, Martin A. Walsh.

**Data curation:** C. David Owen.

**Formal analysis:** Petra Lukacik, C. David Owen, Gemma Harris, Jani Reddy Bolla, Sarah Picaud, Irfan Alibay, Joanne E. Nettleship, Louise E. Bird, Raymond J. Owens, Philip C. Biggin, Panagis Filippakopoulos, Carol V. Robinson, Martin A. Walsh.

**Funding acquisition:** Petra Lukacik, Panagis Filippakopoulos, Martin A. Walsh.

**Investigation:** Petra Lukacik, C. David Owen, Gemma Harris, Jani Reddy Bolla, Sarah Picaud, Irfan Alibay, Joanne E. Nettleship, Louise E. Bird, Raymond J. Owens, Philip C. Biggin, Panagis Filippakopoulos, Carol V. Robinson, Martin A. Walsh.

**Methodology:** Petra Lukacik.

**Writing – original draft:** Petra Lukacik.

**Writing – review & editing:** Petra Lukacik, C. David Owen, Gemma Harris, Jani Reddy Bolla, Sarah Picaud, Irfan Alibay, Joanne E. Nettleship, Louise E. Bird, Raymond J. Owens, Philip C. Biggin, Panagis Filippakopoulos, Carol V. Robinson, Martin A. Walsh.

## References

1. WHO. Fact sheet—The top 10 causes of death: World Health Organisation. 2018. Available from: <http://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death>.
2. Murray CJL, Lopez AD, Mathers CD, Stein C. The Global Burden of Disease 2000 project: aims, methods and data sources. 57.
3. Eurostat. Respiratory diseases statistics—Statistics Explained. 2017. Available from: [http://ec.europa.eu/eurostat/statistics-explained/index.php/Respiratory\\_diseases\\_statistics](http://ec.europa.eu/eurostat/statistics-explained/index.php/Respiratory_diseases_statistics).
4. Ahearn CP, Gallo MC, Murphy TF. Insights on persistent airway infection by non-typeable *Haemophilus influenzae* in chronic obstructive pulmonary disease. *Pathog Dis*. 2017; 75(4). <https://doi.org/10.1093/femspd/ftx042> PMID: 28449098
5. Monasta L, Ronfani L, Marchetti F, Montico M, Vecchi Brumatti L, Bavcar A, et al. Burden of disease caused by otitis media: systematic review and global estimates. *PLoS One*. 2012; 7(4):e36226. <https://doi.org/10.1371/journal.pone.0036226> PMID: 22558393
6. Bardach A, Ciapponi A, Garcia-Marti S, Glujovsky D, Mazzoni A, Fayad A, et al. Epidemiology of acute otitis media in children of Latin America and the Caribbean: a systematic review and meta-analysis. *Int J Pediatr Otorhinolaryngol*. 2011; 75(9):1062–70. <https://doi.org/10.1016/j.ijporl.2011.05.014> PMID: 21665297
7. Hoban D, Felmingham D. The PROTEKT surveillance study: antimicrobial susceptibility of *Haemophilus influenzae* and *Moraxella catarrhalis* from community-acquired respiratory tract infections. *J Antimicrob Chemother*. 2002; 50 Suppl S1:49–59. <https://doi.org/10.1093/jac/45/10/1223> PMID: 12239228
8. Gilsdorf JR. What the pediatrician should know about non-typeable *Haemophilus influenzae*. *J Infect*. 2015; 71 Suppl 1:S10–4. <https://doi.org/10.1016/j.jinf.2015.04.014> PMID: 25917803
9. Swingler G, Fransman D, Hussey G. Conjugate vaccines for preventing *Haemophilus influenzae* type b infections. *Cochrane Database Syst Rev*. 2003;(4):CD001729. <https://doi.org/10.1002/14651858.CD001729> PMID: 14583937
10. Langereis JD, de Jonge MI. Invasive Disease Caused by Nontypeable *Haemophilus influenzae*. *Emerg Infect Dis*. 2015; 21(10):1711–8. <https://doi.org/10.3201/eid2110.150004> PMID: 26407156
11. Poolman JT, Bakaletz L, Cripps A, Denoel PA, Forsgren A, Kyd J, et al. Developing a nontypeable *Haemophilus influenzae* (NTHi) vaccine. *Vaccine*. 2000; 19 Suppl 1:S108–15. [https://doi.org/10.1016/S0264-410X\(00\)00288-7](https://doi.org/10.1016/S0264-410X(00)00288-7) PMID: 11163473
12. Behrouzi A, Vaziri F, Rahimi-Jamrani F, Afrough P, Rahbar M, Satarian F, et al. Vaccine Candidates against Nontypeable *Haemophilus influenzae*: a Review. *Iran Biomed J*. 2017; 21(2):69–76. <https://doi.org/10.18869/acadpub.ijb.21.2.69> PMID: 28088130

13. Mason KM, Munson RS Jr., Bakaletz LO. A mutation in the sap operon attenuates survival of nontypeable *Haemophilus influenzae* in a chinchilla model of otitis media. *Infect Immun*. 2005; 73(1):599–608. <https://doi.org/10.1128/IAI.73.1.599-608.2005> PMID: 15618200
14. Mason KM, Bruggeman ME, Munson RS, Bakaletz LO. The non-typeable *Haemophilus influenzae* Sap transporter provides a mechanism of antimicrobial peptide resistance and SapD-dependent potassium acquisition. *Mol Microbiol*. 2006; 62(5):1357–72. <https://doi.org/10.1111/j.1365-2958.2006.05460.x> PMID: 17064364
15. Vogel AR, Szelestey BR, Raffel FK, Sharpe SW, Gearing RL, Justice SS, et al. SapF-mediated heme-iron utilization enhances persistence and coordinates biofilm architecture of *Haemophilus*. *Front Cell Infect Microbiol*. 2012; 2:42. <https://doi.org/10.3389/fcimb.2012.00042> PMID: 22919633
16. Shelton CL, Raffel FK, Beatty WL, Johnson SM, Mason KM. Sap transporter mediated import and subsequent degradation of antimicrobial peptides in *Haemophilus*. *PLoS Pathog*. 2011; 7(11):e1002360. <https://doi.org/10.1371/journal.ppat.1002360> PMID: 22072973
17. Mason KM, Raffel FK, Ray WC, Bakaletz LO. Heme utilization by nontypeable *Haemophilus influenzae* is essential and dependent on Sap transporter function. *J Bacteriol*. 2011; 193(10):2527–35. <https://doi.org/10.1128/JB.01313-10> PMID: 21441512
18. Rodriguez-Arce I, Al-Jubair T, Euba B, Fernandez-Calvet A, Gil-Campillo C, Marti S, et al. Moonlighting of *Haemophilus influenzae* heme acquisition systems contributes to the host airway-pathogen interplay in a coordinated manner. *Virulence*. 2019; 10(1):315–33. <https://doi.org/10.1080/21505594.2019.1596506> PMID: 30973092
19. Tanaka KJ, Pinkett HW. Oligopeptide-binding protein from nontypeable *Haemophilus influenzae* has ligand-specific sites to accommodate peptides and heme in the binding pocket. *J Biol Chem*. 2019; 294(3):1070–82. <https://doi.org/10.1074/jbc.RA118.004479> PMID: 30455346
20. Raffel FK, Szelestey BR, Beatty WL, Mason KM. The *Haemophilus influenzae* Sap transporter mediates bacterium-epithelial cell homeostasis. *Infect Immun*. 2013; 81(1):43–54. <https://doi.org/10.1128/IAI.00942-12> PMID: 23071138
21. Hsu CR, Chang IW, Hsieh PF, Lin TL, Liu PY, Huang CH, et al. A Novel Role for the *Klebsiella pneumoniae* Sap (Sensitivity to Antimicrobial Peptides) Transporter in Intestinal Cell Interactions, Innate Immune Responses, Liver Abscess, and Virulence. *J Infect Dis*. 2019; 219(8):1294–306. <https://doi.org/10.1093/infdis/jiy615> PMID: 30476200
22. Mount KL, Townsend CA, Rinker SD, Gu X, Fortney KR, Zwickl BW, et al. *Haemophilus ducreyi* SapA contributes to cathelicidin resistance and virulence in humans. *Infect Immun*. 2010; 78(3):1176–84. <https://doi.org/10.1128/IAI.01014-09> PMID: 20086092
23. Rinker SD, Gu X, Fortney KR, Zwickl BW, Katz BP, Janowicz DM, et al. Permeases of the sap transporter are required for cathelicidin resistance and virulence of *Haemophilus ducreyi* in humans. *J Infect Dis*. 2012; 206(9):1407–14. <https://doi.org/10.1093/infdis/jis525> PMID: 22930807
24. Groisman EA, Parra-Lopez C, Salcedo M, Lipps CJ, Heffron F. Resistance to host antimicrobial peptides is necessary for *Salmonella* virulence. *Proc Natl Acad Sci U S A*. 1992; 89(24):11939–43. <https://doi.org/10.1073/pnas.89.24.11939> PMID: 1465423
25. Parra-Lopez C, Baer MT, Groisman EA. Molecular genetic analysis of a locus required for resistance to antimicrobial peptides in *Salmonella typhimurium*. *EMBO J*. 1993; 12(11):4053–62. PMID: 8223423
26. Parra-Lopez C, Lin R, Aspedon A, Groisman EA. A *Salmonella* protein that is required for resistance to antimicrobial peptides and transport of potassium. *EMBO J*. 1994; 13(17):3964–72. PMID: 8076592
27. Rodas PI, Contreras I, Mora GC. *Salmonella enterica* serovar Typhi has a 4.1 kb genetic island inserted within the sapABCDF operon that causes loss of resistance to the antimicrobial peptide protamine. *J Antimicrob Chemother*. 2010; 65(8):1624–30. <https://doi.org/10.1093/jac/dkq197> PMID: 20551214
28. Lopez-Solanilla E, Garcia-Olmedo F, Rodriguez-Palenzuela P. Inactivation of the sapA to sapF locus of *Erwinia chrysanthemi* reveals common features in plant and animal bacterial pathogenesis. *Plant Cell*. 1998; 10(6):917–24. PMID: 9634580
29. Bertsson RP, Smits SH, Schmitt L, Slotboom DJ, Poolman B. A structural classification of substrate-binding proteins. *FEBS Lett*. 2010; 584(12):2606–17. <https://doi.org/10.1016/j.febslet.2010.04.043> PMID: 20412802
30. Fukami-Kobayashi K, Tateno Y, Nishikawa K. Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *J Mol Biol*. 1999; 286(1):279–90. <https://doi.org/10.1006/jmbi.1998.2454> PMID: 9931266
31. Scheepers GH, Lycklama ANJA, Poolman B. An updated structural classification of substrate-binding proteins. *FEBS Lett*. 2016; 590(23):4393–401. <https://doi.org/10.1002/1873-3468.12445> PMID: 27714801

32. Mao B, Pear MR, McCammon JA, Quioco FA. Hinge-bending in L-arabinose-binding protein. The "Venus's-flytrap" model. *J Biol Chem.* 1982; 257(3):1131–3. PMID: [7035444](#)
33. Berntsson RP, Doeven MK, Fusetti F, Duurkens RH, Sengupta D, Marrink SJ, et al. The structural basis for peptide selection by the transport receptor OppA. *EMBO J.* 2009; 28(9):1332–40. <https://doi.org/10.1038/emboj.2009.65> PMID: [19300437](#)
34. Liptfert J, Skinner GM, Keegstra JM, Hensgens T, Jager T, Dulin D, et al. Double-stranded RNA under force and torque: similarities to and striking differences from double-stranded DNA. *Proc Natl Acad Sci U S A.* 2014; 111(43):15408–13. <https://doi.org/10.1073/pnas.1407197111> PMID: [25313077](#)
35. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A.* 2001; 98(18):10037–41. <https://doi.org/10.1073/pnas.181342398> PMID: [11517324](#)
36. Tam R, Saier MH Jr. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol Rev.* 1993; 57(2):320–46. <https://doi.org/10.1128/mr.57.2.320-346.1993> PMID: [8336670](#)
37. Dunten P, Mowbray SL. Modeling of the structure of the *Haemophilus influenzae* heme-binding protein suggests a mode of heme interaction. *Protein Sci.* 1995; 4(11):2335–40. <https://doi.org/10.1002/pro.5560041111> PMID: [8563630](#)
38. Yang X, Li H, Huang Y, Liu S. The dataset for protein-RNA binding affinity. *Protein Sci.* 2013; 22(12):1808–11. <https://doi.org/10.1002/pro.2383> PMID: [24127340](#)
39. Ben-Hur A. BD. Sequence Motifs: Highly Predictive Features of Protein Function. In: Guyon I. NM, Gunn S., Zadeh L.A., editor. *Feature Extraction Studies in Fuzziness and Soft Computing*, vol 207. 207. Berlin, Heidelberg: Springer; 2006.
40. Ghosal A, Upadhyaya BB, Fritz JV, Heintz-Buschart A, Desai MS, Yusuf D, et al. The extracellular RNA complement of *Escherichia coli*. *Microbiologyopen.* 2015. <https://doi.org/10.1002/mbo3.235> PMID: [25611733](#)
41. Beckloff N, Laube D, Castro T, Furgang D, Park S, Perlin D, et al. Activity of an antimicrobial peptide mimetic against planktonic and biofilm cultures of oral pathogens. *Antimicrob Agents Chemother.* 2007; 51(11):4125–32. <https://doi.org/10.1128/AAC.00208-07> PMID: [17785509](#)
42. Juneau RA, Pang B, Weimer KE, Armbruster CE, Swords WE. Nontypeable *Haemophilus influenzae* initiates formation of neutrophil extracellular traps. *Infect Immun.* 2011; 79(1):431–8. <https://doi.org/10.1128/IAI.00660-10> PMID: [20956567](#)
43. Remijsen Q, Kuijpers TW, Wirawan E, Lippens S, Vandenabeele P, Vanden Berghe T. Dying for a cause: NETosis, mechanisms behind an antimicrobial cell death modality. *Cell Death Differ.* 2011; 18(4):581–8. <https://doi.org/10.1038/cdd.2011.1> PMID: [21293492](#)
44. Weiner DJ, Bucki R, Janmey PA. The antimicrobial activity of the cathelicidin LL37 is inhibited by F-actin bundles and restored by gelsolin. *Am J Respir Cell Mol Biol.* 2003; 28(6):738–45. <https://doi.org/10.1165/rcmb.2002-0191OC> PMID: [12600826](#)
45. Jones EA, McGillivray G, Bakaletz LO. Extracellular DNA within a nontypeable *Haemophilus influenzae*-induced biofilm binds human beta defensin-3 and reduces its antimicrobial activity. *J Innate Immun.* 2013; 5(1):24–38. <https://doi.org/10.1159/000339961> PMID: [22922323](#)
46. Sugiyama Y, Nakamura A, Matsumoto M, Kanbe A, Sakanaka M, Higashi K, et al. A Novel Putrescine Exporter SapBCDF of *Escherichia coli*. *J Biol Chem.* 2016; 291(51):26343–51. <https://doi.org/10.1074/jbc.M116.762450> PMID: [27803167](#)
47. Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, et al. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* 2007; 35(6):e45. <https://doi.org/10.1093/nar/gkm047> PMID: [17317681](#)
48. Winter G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography.* 2010; 43:186–90.
49. Leslie AGW. The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr.* 2006; 62:48–57. <https://doi.org/10.1107/S0907444905039107> PMID: [16369093](#)
50. Scaling Evans P. and assessment of data quality. *Acta Crystallogr D Biol Crystallogr.* 2006; 62:72–82. <https://doi.org/10.1107/S0907444905036693> PMID: [16369096](#)
51. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr.* 1994; 50(Pt 5):760–3. <https://doi.org/10.1107/S0907444994003112> PMID: [15299374](#)
52. Xds Kabsch W. *Acta Crystallogr D Biol Crystallogr.* 2010; 66(Pt 2):125–32. <https://doi.org/10.1107/S0907444909047337> PMID: [20124692](#)
53. Long F, Vagin AA, Young P, Murshudov GN. BALBES: a molecular-replacement pipeline. *Acta Crystallogr D Biol Crystallogr.* 2008; 64(Pt 1):125–32. <https://doi.org/10.1107/S0907444907050172> PMID: [18094476](#)

54. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr*. 2007; 40(Pt 4):658–74. <https://doi.org/10.1107/S0021889807021206> PMID: 19461840
55. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 2010; 66(Pt 2):213–21. <https://doi.org/10.1107/S0907444909052925> PMID: 20124702
56. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004; 60(Pt 12 Pt 1):2126–32. <https://doi.org/10.1107/S0907444904019158> PMID: 15572765
57. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr*. 2011; 67(Pt 4):355–67. <https://doi.org/10.1107/S0907444911001314> PMID: 21460454
58. Blanc E, Roversi P, Vornrhein C, Flensburg C, Lea SM, Bricogne G. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr D Biol Crystallogr*. 2004; 60(Pt 12 Pt 1):2210–21. <https://doi.org/10.1107/S0907444904016427> PMID: 15572774
59. van Beusekom B, Touw WG, Tatineni M, Somani S, Rajagopal G, Luo J, et al. Homology-based hydrogen bond information improves crystallographic structures in the PDB. *Protein Sci*. 2018; 27(3):798–808. <https://doi.org/10.1002/pro.3353> PMID: 29168245
60. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2010; 66(Pt 1):12–21. <https://doi.org/10.1107/S0907444909042073> PMID: 20057044
61. Nicholls RA, Fischer M, McNicholas S, Murshudov GN. Conformation-independent structural comparison of macromolecules with ProSMART. *Acta Crystallogr D Biol Crystallogr*. 2014; 70(Pt 9):2487–99. <https://doi.org/10.1107/S1399004714016241> PMID: 25195761
62. Kleywegt GJ, Jones TA. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr D Biol Crystallogr*. 1994; 50(Pt 2):178–85. <https://doi.org/10.1107/S0907444993011333> PMID: 15299456
63. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci*. 2018; 27(1):135–45. <https://doi.org/10.1002/pro.3290> PMID: 28884485
64. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*. 2007; 23(1):127–8. <https://doi.org/10.1093/bioinformatics/btl529> PMID: 17050570
65. Davis BJ. Disc Electrophoresis. II. Method and Application to Human Serum Proteins. *Ann N Y Acad Sci*. 1964; 121:404–27. <https://doi.org/10.1111/j.1749-6632.1964.tb14213.x> PMID: 14240539
66. Bollag DM, Rozycki MD, Edelstein SJ. *Protein Methods*: Wiley and Sons Inc.; 1996.
67. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 2014; 42(Web Server issue):W320–4. <https://doi.org/10.1093/nar/gku316> PMID: 24753421