

## **Identification and characterization of a bacterial periplasmic solute binding protein that binds L-amino acid amides.**

*Oliver B. Smith*<sup>1,2</sup>, *Rebecca L. Frkic*<sup>1,3</sup>, *Marina G. Rahman*<sup>2,4</sup>, *Colin J. Jackson*<sup>1,2,3,4‡</sup>, *Joe A. Kaczmarski*<sup>2,4‡</sup>

<sup>1</sup>Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia.

<sup>2</sup>ARC Centre of Excellence in Synthetic Biology, Australian National University, Canberra, ACT 2601, Australia.

<sup>3</sup>ARC Centre of Excellence for Innovations in Peptide & Protein Science, Australian National University, Canberra, ACT 2601, Australia.

<sup>4</sup>Research School of Biology, Australian National University, Canberra, ACT 2601, Australia.

‡Corresponding author(s): Joe Kaczmarski, [joe.kaczmarski@anu.edu.au](mailto:joe.kaczmarski@anu.edu.au); Colin Jackson, [colin.jackson@anu.edu.au](mailto:colin.jackson@anu.edu.au)

# ABSTRACT

Periplasmic solute-binding proteins (SBPs) are key ligand recognition components of bacterial ATP-binding cassette (ABC) transporters that allow bacteria to import nutrients and metabolic precursors from the environment. Periplasmic SBPs comprise a large and diverse family of proteins, of which only a small number have been empirically characterized. In this work, we identify a set of 610 unique uncharacterized proteins within the SBP\_bac\_5 family that are found in conserved operons comprising genes encoding (i) ABC transport systems and (ii) putative amidases from the FmdA\_AmdA family. From these uncharacterized SBP\_bac\_5 proteins, we characterize a representative periplasmic SBP from *Mesorhizobium* sp. A09 (*MeAmi*\_SBP) and show that *MeAmi*\_SBP binds L-amino acid amides but not the corresponding L-amino acids. An X-ray crystal structure of *MeAmi*\_SBP bound to L-serinamide highlights the residues that impart distinct specificity for L-amino acid amides and reveals a structural Ca<sup>2+</sup> binding site within one of the lobes of the protein. We show that the residues involved in ligand and Ca<sup>2+</sup> binding are conserved amongst the 610 SBPs from experimentally uncharacterized FmdA\_AmdA amidase-associated ABC transporter systems, suggesting these homologous systems are also likely to be involved in the sensing, uptake and metabolism of L-amino acid amides across many Gram-negative nitrogen-fixing soil bacteria. We propose that *MeAmi*\_SBP is involved in the uptake of such solutes to supplement pathways such as the citric acid cycle and the glutamine synthetase-glutamate synthase pathway. This work expands our currently limited understanding of microbial interactions with L-amino acid amides and bacterial nitrogen utilization.

# INTRODUCTION

Bacterial ATP-binding cassette (ABC) transporter systems enable bacteria to adapt to diverse environments by facilitating the selective uptake of key nutrients and metabolic precursors (1). While bacterial ABC transporter systems for sugars (2–5), oligopeptides (6,7) and amino acids (8–10) are well characterized, those specialized for the import and utilization of other important carbon and nitrogen sources, such as amides, remain poorly understood.

In Gram-negative bacteria, ABC transporter systems are typically encoded by a single operon that include genes for (i) a periplasmic solute binding protein (SBP) that initiates solute recognition and capture, (ii) transmembrane permease and ATP-binding domains for the transport of the solute into the cytoplasm and, in some cases, (iii) cytosolic enzyme(s) that catalyze the first step in solute metabolism (1,11). Because the periplasmic SBP components determine the ligand specificity of the transport system (12–15), identifying uncharacterized SBPs with novel ligand binding specificities can help expand our understanding of bacterial ABC transporters and their role in solute uptake and metabolism.

The SBP superfamily is a large group of proteins (> 3.4 million sequences in February 2024) (16,17) with diverse sequences and ligand-binding preferences (13,18). Although prokaryotic SBPs share a common core architecture comprising two primary  $\alpha/\beta$  domains linked by a flexible hinge region and a central ligand-binding cleft, their structures are diverse and cover 33 distinct Pfam structural families (19). Proteins of the “extracellular solute-binding protein, family 5” Pfam family (“SBP\_bac\_5”, PF00496) are amongst the most structurally divergent SBPs and contain three topological domains (13). Unlike other structural classes of SBPs, which show preferences for a particular ligand class (19), members of the SBP\_bac\_5 family bind a range of ligand types, including oligopeptides (20,21), glutathione (22,23) and metal ions (24,25). While a handful of the SBP\_bac\_5 proteins have been experimentally characterized, much of the sequence and functional diversity of the SBP\_bac\_5 family remains unexplored.

In this work, we use sequence similarity network (SSN) and genomic context analyses to identify a cluster of 610 unique uncharacterized proteins within the SBP\_bac\_5 family that are all found within ABC transporter operons that also contain an enzyme from the FmdA\_AmdA acetamidase/formamidase family. We experimentally characterize a representative of this group of proteins, *MeAmi\_SBP* from *Mesorhizobium* sp. AP09, and show that *MeAmi\_SBP* binds L-amino acid amides but not the corresponding L-amino acids. A crystal structure of

*MeAmi\_SBP* bound to L-serinamide highlights the key interactions and binding site properties that confer this novel, divergent, ligand specificity and reveals a  $\text{Ca}^{2+}$  binding site in one of the protein's lobes. Ligand-binding residues are conserved amongst the other proteins in this cluster, suggesting that this group of proteins has a role in the uptake and metabolism of L-amino acid amides. This work expands our limited understanding of microbial interactions with L-amino acid amides, especially in relation to bacterial nitrogen utilization.

## RESULTS

### Identification of a group of uncharacterized SBPs from amidase-containing ABC transporter operons in Gram-negative bacteria.

Sequence similarity networks (SSNs) are valuable tools for identifying and exploring uncharacterized regions of sequence-function space within protein families (26,27). We generated an SSN comprising ~400,000 protein sequences from the SBP\_bac\_5 family (**Figure 1A**). Mapping SWISS-PROT annotations and PDB structures onto this SSN show that the experimentally characterized members of the SBP\_bac\_5 family are concentrated within localized regions of the sequence space. Indeed, when the SSN was generated using an alignment score cut-off of 100 (corresponding to roughly 30% sequence identity) most of the distinct sequence clusters had no experimentally characterized members, highlighting that much of the functional diversity of this family remains unexplored.

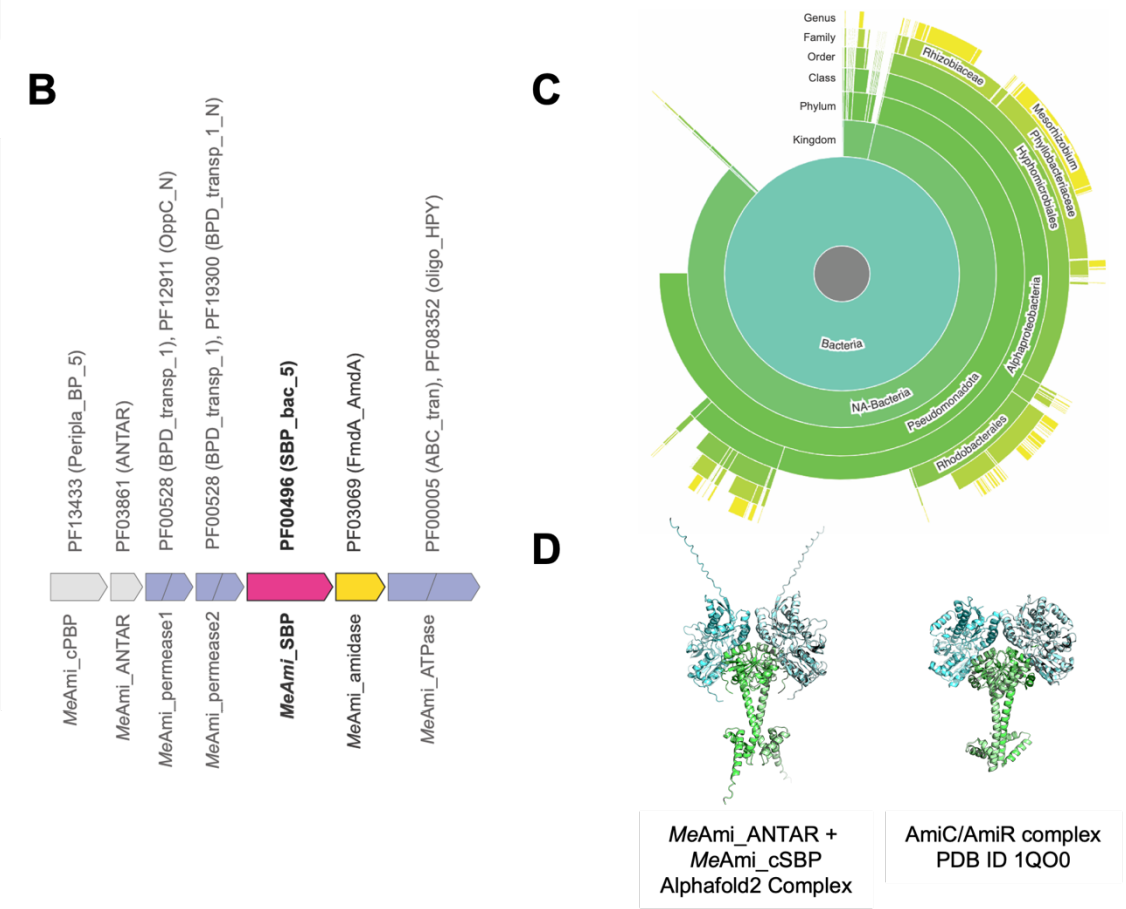
To identify novel and interesting ligand binding properties within the SBP\_bac\_5 family, we performed genomic context analyses of each SBP\_bac\_5 protein cluster to identify which SBP open reading frames were colocalized with ABC transporter cassettes and were also found near genes encoding enzymes that could suggest a biochemical requirement for processing novel ligands (28). One such group comprised 610 unique SBP\_bac\_5 proteins that shared a highly conserved ABC transporter operon structure containing a gene that encodes a protein from the acetamidase/formamidase family (PF03069, FmdA\_AmdA) (**Figure 1B**).

The presence of a putative FmdA\_AmdA amidase family member in each of these operons suggested that these ABC transport systems may be involved in the import and processing of amine-containing compounds. While the FmdA\_AmdA family of proteins is primarily known for hydrolysis of acetate (29) and formate (30), some members of this family have activity on other amide-containing compounds, including lactams and aliphatic amides (31). Interestingly, the cluster of 610 amidase-associated operons that we identified is dominated by proteins from Gram-negative bacteria that are specialized in the uptake of nitrogen-containing compounds, such as *Rhizobiales*, *Mesorhizobium* and *Rhodobacteraceae* (32–34) (**Figure 1A, 1C**). Notably, no other clusters within the SBP\_bac\_5 family were associated with a FmdA\_AmdA protein and the structure of these amidase-associated operons was clearly distinct from previously well-studied amidase operons such as the *Pseudomonas aeruginosa* amidase operon (35,36), *Rhodococcus* sp. R312 amidase operon (37) and *Mycobacterium smegmatis* acetamidase

operon (38–40) (**Supplementary Figure 1**). This marked difference prompted further investigation into this distinct set of ABC transporter-associated proteins.

In addition to genes for the putative amidase and the core ABC transporter import proteins (i.e. the periplasmic SBP, membrane permease and ATP-binding domains), two additional genes were consistently present at the 5' end of each operon in this cluster. These genes were predicted to encode (i) a protein with a periplasmic binding protein fold (PF13433, Peripla\_BP\_5) that lacks a periplasmic signaling peptide and (ii) a protein belonging to the family of AmiR and NasR transcription antitermination regulators (ANTAR, PF03861) (**Figure 1B**). The presence of transcription antitermination regulators in all of these operons suggested that the expression of these ABC transporter and associated proteins may be regulated in a ligand-dependent ANTAR-mediated manner like the amide import operon of *P. aeruginosa* (41). This potential regulatory mechanism also prompted investigation into the SBP component of this conserved operon.

To establish the function and substrate specificity of this group of ABC transporters associated with potentially unique enzymatic processing and regulatory mechanisms, we focused on identifying the ligand specificity of the associated periplasmic SBPs. Rather than directly experimentally characterizing each member of this cluster (which share over 35% sequence identity), we used a profile hidden Markov model (43) to identify the most consensus-like, or representative, SBP sequence within this group of proteins. The selected representative protein, which we call *MeAmi*\_SBP, comes from the soil bacterium *Mesorhizobium* sp. AP09 and is annotated on UniProt as being a “peptide/nickel transport system substrate-binding protein/oligopeptide transport system substrate-binding protein” (UniProt ID: A0A395KPD0) (44). Importantly, *MeAmi*\_SBP has no closely related homologs that have been experimentally characterized, with the closest related protein with a confirmed function being the glutathione-binding protein, GsiB, from *Salmonella enterica* (BLAST E-value =  $9e^{-73}$ , 32% sequence identity) (45). Of note, the ANTAR protein in the *MeAmi* operon (*MeAmi*\_ANTAR) shares 32% sequence identity with the *P. aeruginosa* ANTAR protein, AmiR, and the cytosolic Peripla\_BP\_5 protein from the *MeAmi* operon (*MeAmi*\_cPBP) shares 29% sequence identity with the cytosolic ANTAR-associated ligand receptor from *P. aeruginosa*, AmiC. An Alphafold2 (42) model of these two proteins predict they will form a dimer of dimers similar to the AmiC/AmiR complex (**Figure 1D**).





**Figure 1. A group of 610 homologous SBP\_bac\_5 solute binding proteins from conserved amidase-containing ABC transporter operons.** **A.** A sequence similarity network (SSN) of the SBP\_bac\_5 Pfam using an alignment cut-off score of 100. Nodes representing proteins with published PDB structures are shown as green diamonds. Nodes representing proteins with reviewed SWISS-PROT annotations (but no PDB structures) are shown as cyan circles. The node representing *MeAmi\_SBP* is shown as a magenta square. A few key SWISS-PROT annotations are labelled. **(insert)** A close-up of the *MeAmi\_SBP*-containing cluster, using an alignment cut-off score of 250. Nodes are colored based on the Order of the host organism (except for *MeAmi\_SBP*, magenta square). **B.** The conserved genomic neighborhood/operon structure amongst *MeAmi\_SBP* and its homologs. Each open-reading frame is labelled above with the corresponding Pfam identifiers. Each gene is also labelled with the name we give the corresponding gene in the *MeAmi* operon from *Mesorhizobium* sp. AP09. **C.** The taxonomic distribution of proteins in the *MeAmi\_SBP* cluster generated using the EFI-EST Taxonomy tool (46). **D.** An AF2-predicted model of the *MeAmi\_ANTAR/MeAmi\_cPBP* complex (left) and the X-ray crystal structure of the *P. aeruginosa* AmiC/AmiR complex (right). In each structure, the respective PBP components are shown in light blue, and the respective ANTAR components are shown in green.

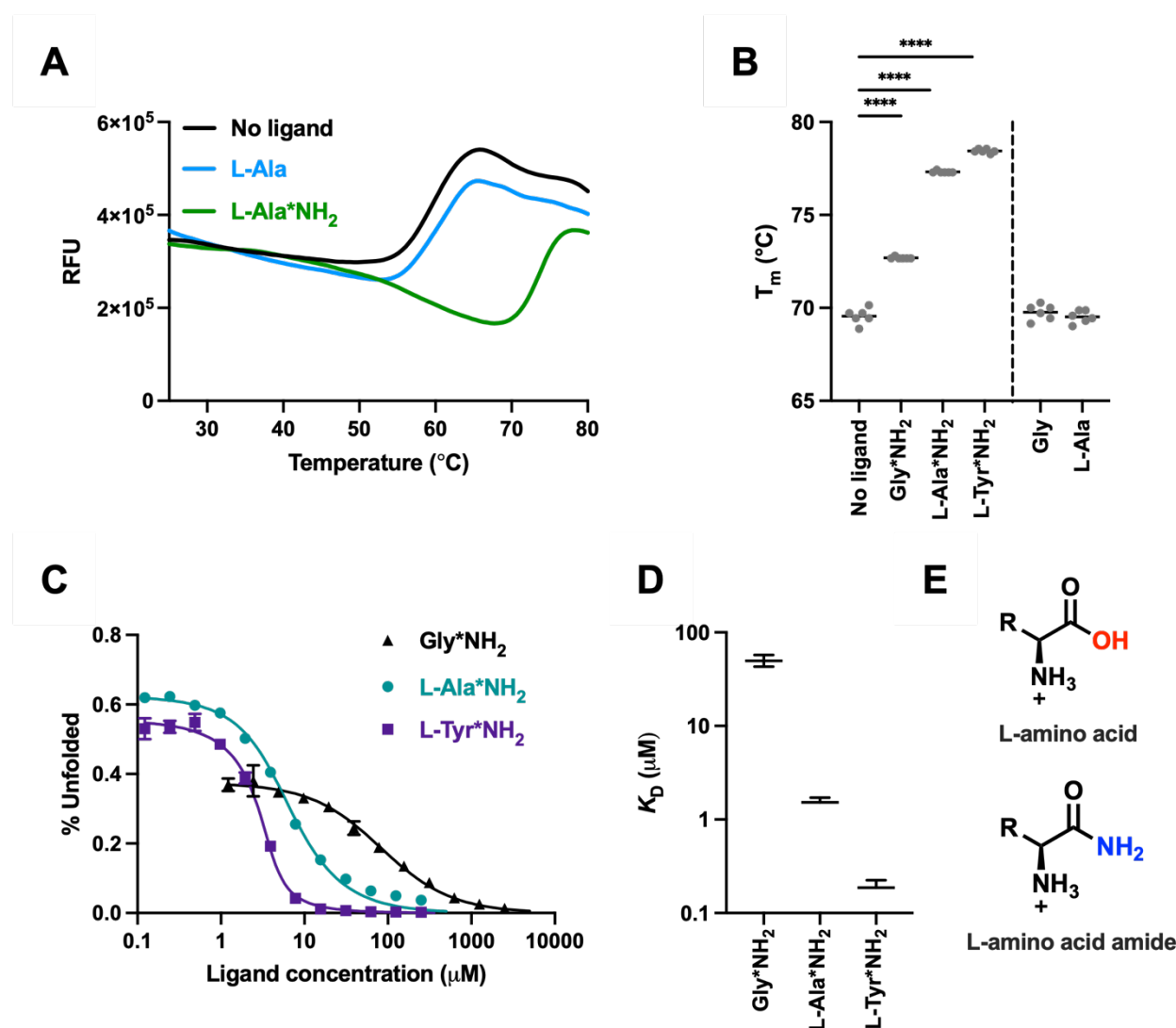
***MeAmi\_SBP* binds L-amino acid amides.** His<sub>6</sub>-tagged *MeAmi\_SBP* was expressed in the cytosol of *Escherichia coli* BL21 (DE3) cells and purified. *MeAmi\_SBP* is monomeric in solution (**Supplementary Figure 2**). To identify ligands of *MeAmi\_SBP*, we screened for binding to 510 biologically-relevant compounds from the Biolog Phenotype Microarray plates PM1-PM6 using a differential scanning fluorimetry (DSF)-based thermal shift assay. The PM1-PM6 screens are designed for use in bacterial phenotyping studies that assess the uptake and metabolism of common solutes (47), and therefore contain compounds that ABC transporter-associated SBPs, such as *MeAmi\_SBP*, are likely to bind (**Supplementary Tables 1-6**).

Despite *MeAmi\_SBP* being annotated as a peptide binding protein, no thermal shifts were observed for any of the ~100 di- and tripeptides present in the screen (**Supplementary Figure 3**). Similarly, there were no signs of ligand-induced stabilization by other known ligands of the SBP\_bac\_5 family, such as glutathione and octopine. A clear ligand-induced thermal shift ( $\Delta T_m$  14.1 °C compared to negative control) was observed in the presence of an L-amino acid amide, L-alaninamide, only (**Figure 2A**). L-alaninamide was the only amino acid amide in the initial screen; however, subsequent experiments also showed ligand-induced thermal shifts in the presence of 10 mM glycylamide and L-tyrosinamide, indicating *MeAmi\_SBP* binds L-amino acid amides with diverse side chain chemistry (**Figure 2B**). Isothermal analysis of ligand-titrated DSF experiments revealed that the affinity of *MeAmi\_SBP* for the three amino



acid amides spans two orders of magnitude: glycineamide ( $K_D$  of 49.7  $\mu\text{M}$ ), L-alanineamide ( $K_D$  of 1.53  $\mu\text{M}$ ) and L-tyrosineamide ( $K_D$  of 0.187  $\mu\text{M}$ ) (**Figure 2C, 2D**) (48).

Importantly, thermal shifts were not observed for any of the proteinogenic L-amino acids tested (**Figure 2B**), confirming selectivity for the L-amino acid amide moiety over the L-amino acid moiety (**Figure 2E**). Similarly, no thermal shifts were observed in the presence of non-amino acid amides such as acetamide and glucuronamide (**Supplementary Figure 3**). These data raised questions about (i) the structural basis of selectivity for L-amino acid amides over L-amino acids and other amidated compounds and (ii) the structural basis for large differences in ligand affinity that could be attributed solely to differences in side chain chemistry.



**Figure 2. *MeAmi\_SBP* from *Mesorhizobium* sp. A09 is a periplasmic solute binding protein specific for L-amino acid amides.** **A.** Representative DSF fluorescence traces of *MeAmi\_SBP* in the presence of L-alanine, L-alaninamide (L-Ala\* $\text{NH}_2$ ), and a no ligand control. RFU = relative fluorescent units. **B.** Melting temperatures ( $T_m$ ) of *MeAmi\_SBP* in the presence of 10 mM of various amino acid amides (L-X\* $\text{NH}_2$ ) and the corresponding amino acids as measured by DSF ( $n=6$ , horizontal line represents the mean). \*\*\*\* refers to  $p < 0.0001$  from unpaired parametric t-tests. **C.** Isothermal DSF curves from amino acid amide titrations into *MeAmi\_SBP*. Data is ( $n=3$ , error bars=SD). A one-site model (solid line) was fitted to the data using FoldAffinity (48). **D.** Binding affinities of *MeAmi\_SBP* for the L-amino acid amides. Data represent the mean, lower and upper limit of  $K_D$  values from the fit of the data in (C). **E.** General chemical structures of L-amino acids and L-amino acid amides. R= side chain (i.e. Gly: R=H, Ala: R= $\text{CH}_3$ , etc.).

**The structure of *MeAmi\_SBP* in complex with an amino acid amide.** We obtained a 1.55 Å crystal structure of *MeAmi\_SBP*, heterologously expressed and purified from *Escherichia coli*. As expected, *MeAmi\_SBP* adopts the SBP\_bac\_5 fold, with three subdomains (**Figure 3A**, **Supplementary Table 7**).

Clear electron density within the canonical SBP binding cleft indicated that *MeAmi\_SBP* had co-purified and crystallized bound to a ligand from the expression host. The high resolution of the data meant that we could confidently model L-serinamide within the binding site (**Figure 3B**). Analysis of the structure of *MeAmi\_SBP* in complex with L-serinamide revealed key interactions between the protein and the ligand. This includes (i) a salt-bridge between the N-terminal amine of L-serinamide and Asp413, (ii) hydrogen bonds between the L-serinamide carbonyl and both the NH group of the Trp34 sidechain and the amide NH of Ile411, and (iii) hydrogen bonds between the nitrogen of L-serinamide's carboxamide group and the backbone carbonyl groups of Tyr32 and Met408.

The structure of this complex helps to explain the selectivity of *MeAmi\_SBP* for amino acid amides over the corresponding L-amino acids: the hydrogen bonds observed between the carboxamide group of L-serinamide and *MeAmi\_SBP* would be disrupted if the carboxamide group was replaced with a carboxylic acid (as would be the case for the corresponding amino acids). Similarly, the key interaction between the N-terminal amino group of L-serinamide and Asp413 explains why *MeAmi\_SBP* does not bind amides that lack an N-terminal amino group, such as acetamide and glucuronamide.

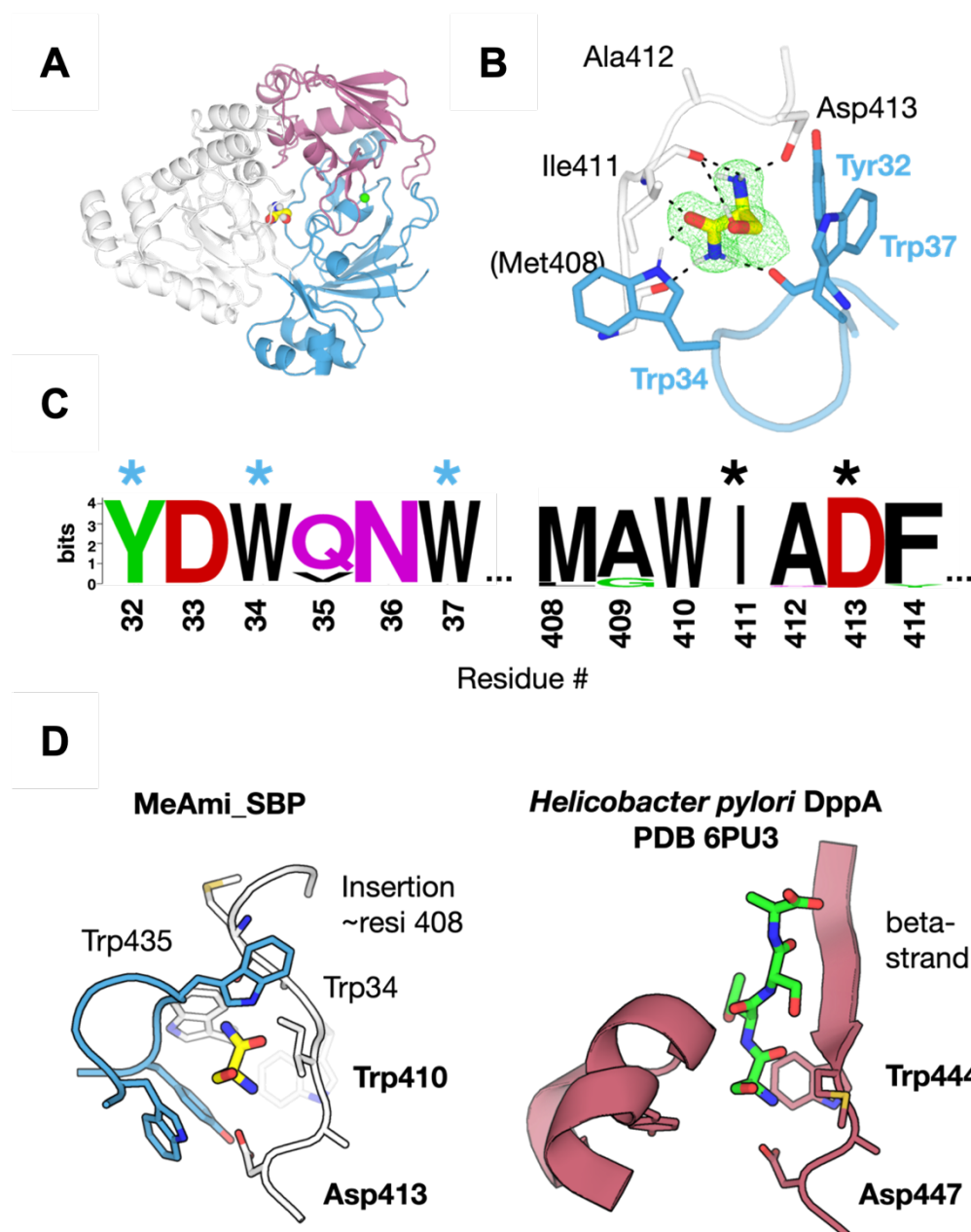
Given that our DSF experiments showed that *MeAmi\_SBP* binds glycineamide, L-alanineamide and L-tyrosineamide with various affinities, we used computational docking to predict and understand how these amino acid amides would bind *MeAmi\_SBP*. As expected, top-scoring poses show glycineamide, L-alanineamide and L-tyrosineamide occupying the same binding site in *MeAmi\_SBP* as L-serineamide, with the backbone motif held in place by the same interactions observed in the L-serineamide complex (**Supplementary Figure 4**). With the backbone of the L-amino acid amides held in this position, the side chains point away from the binding site where there is space within the binding cleft of *MeAmi\_SBP* to accommodate bulkier sidechains, such as that of L-tyrosineamide. While most of the interactions between the ligands and the protein are from the L-amino acid amide backbone, additional interactions with the ligand side chains likely contribute to the observed differences in binding affinity between glycineamide, L-alanineamide and L-tyrosineamide. For example, glycineamide, which had the lowest binding affinity in isothermal DSF experiments ( $K_D$  of 49.7  $\mu$ M) lacks a side chain and thus forms fewer stabilizing interactions. On the other hand, the side chain of L-tyrosineamide likely forms hydrophobic and pi-stacking interactions with Trp32, Trp410, Trp435 and Tyr32, which would contribute to the increased affinity ( $K_D$  of 0.187  $\mu$ M). The lack of water molecules in the side chain-binding site (**Figure 3B**) indicates side chain contributions to L-amino acid amide binding are primarily enthalpic, with no entropic contribution being made by the expulsion of water molecules from this binding site.

Sequence analysis of the related SBP proteins from the same cluster as *MeAmi\_SBP* reveals that all residues involved in binding L-amino acid amide backbone in *MeAmi\_SBP* are highly conserved (**Figure 3C**), suggesting that other proteins within this cluster likely have a similar ligand binding profile to *MeAmi\_SBP*. This, together with the conserved operon structure amongst all 610 proteins in this group of proteins, suggests that this distinct group of amidase-associated ABC transporters share a common substrate and physiological function.

We also compared the structure of *MeAmi\_SBP* with published structures of other members of the SBP\_bac\_5 family, including the most structurally related proteins: this included the periplasmic dipeptide-binding proteins DppA from *Helicobacter pylori* (PDB 6PU3, sequence ID = 28.5%) (49) (**Figure 3D**), DppA from *E. coli* K-12 (PDB 1DPP, sequence ID= 30.7%) (50), the dipeptide transport protein from *Yersinia pestis* (PDB 5F1Q, sequence ID = 30 %) (51), and the glutathione-binding protein A (GbpA) from *Haemophilus parasuis* SH0165 (PDB 3M8U, sequence ID = 30%) (23) (**Supplementary Figure 5**). The binding sites of these proteins and

their mode of binding to the ligands share similar characteristics in that they all have conserved residues at positions equivalent to Trp410 and Asp413, which interact directly with the shared NH<sub>2</sub> group of their ligands (**Supplementary Figure 5**). However, notable differences in the *MeAmi\_SBP* structure include an insertion near residue Met408 that appears to disrupt the beta strand observed in other published SBP\_bac\_5 structures and facilitate binding to L-amino acid amides only. The insertion positions the backbone carbonyl of Met408 so that it can interact with the NH<sub>2</sub> group of L-serinamide and, along with several additional bulky residues in *MeAmi\_SBP*, such as Trp435, Tyr32 and Trp34, also appears to decrease the overall size of the binding pocket compared to the other ligand-bound SBP\_bac\_5 proteins. This may prevent the binding of larger ligands such as dipeptides. The binding site structure thus justifies the divergence of the ligand binding partners of *MeAmi\_SBP* compared to otherwise highly structural homologous proteins.

It is important to note that SBPs are dynamic proteins that often sample both closed states (in which the subdomains are close together) and open states (in which the subdomains are further apart) in solution, and that the structure of *MeAmi\_SBP* bound to L-serinamide likely represents the closed conformation of the protein. Indeed, an AF2-predicted model of apo-*MeAmi\_SBP* shows a more open conformation than what is observed in the crystal structure of the *MeAmi\_SBP* in complex with L-serinamide (**Supplementary Figure 6**) and may represent a physiologically-relevant open, unliganded state. Considering that the L-serinamide and the other L-amino acid amide ligands form bridging interactions with residues from the two core domains of *MeAmi\_SBP*, it is likely that ligand binding would stabilize the closed conformation of the protein that is observed in the crystal structure.



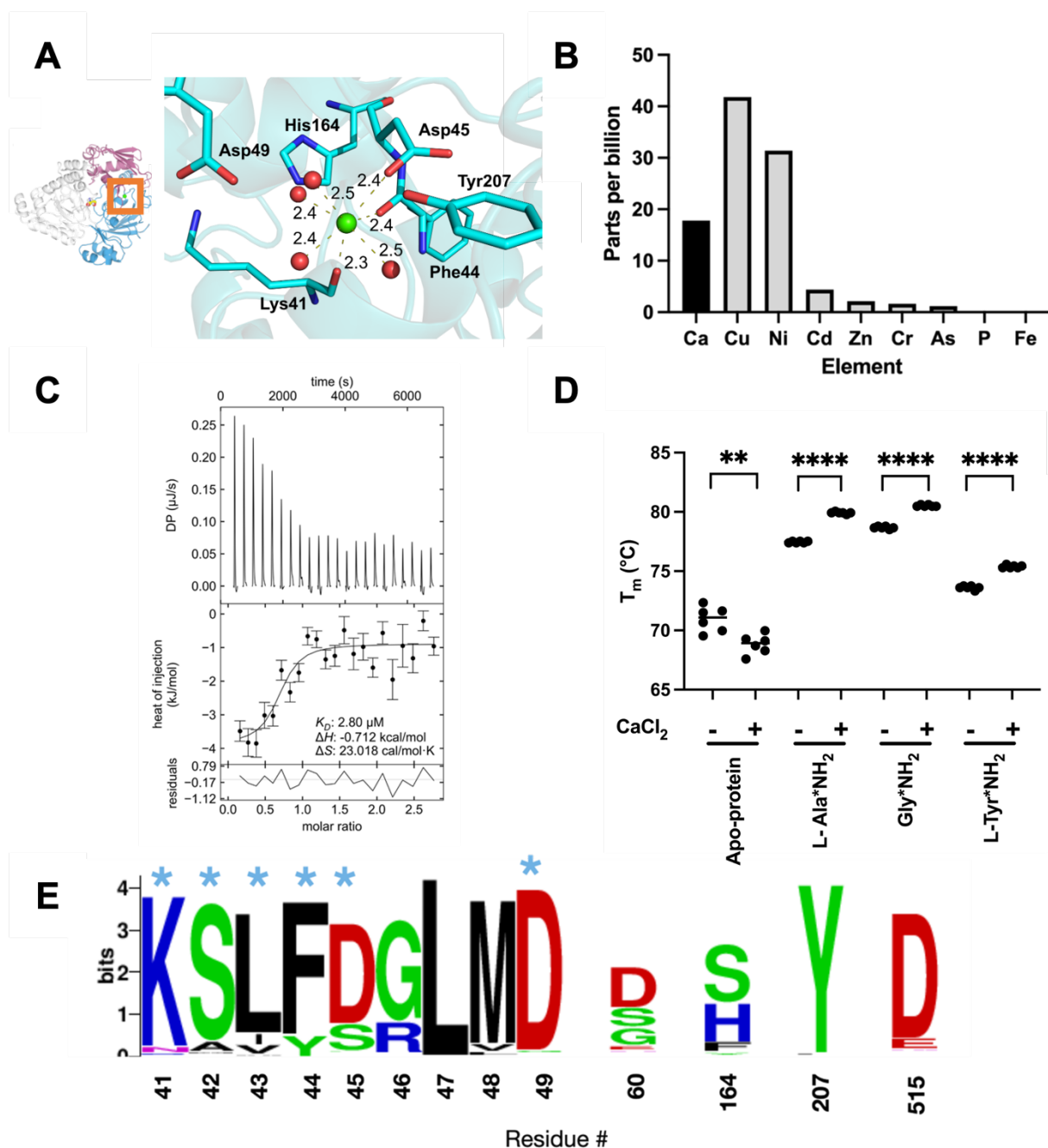
**Figure 3. Structural analysis of MeAmi\_SBP supports specific binding to amino acid amides. A.** X-ray crystal structure of MeAmi\_SBP showing the three subdomains of the protein (white, pink and blue), the bound L-serinamide molecule (spheres) within the central binding site, and the bound  $\text{Ca}^{2+}$  ion (green sphere). **B.** The amino acid amide binding site of MeAmi\_SBP showing key interactions. Hydrogens were added in ChimeraX to emphasize the H-bonding network. A Polder omit map around the ligand is shown as a green mesh (Contoured at 3 sigma, 3 Å carve around ligand). **C.** Sequence WebLogo (52) showing sequence conservation amongst MeAmi\_SBP homologs at residue positions important for binding amino acid amides in MeAmi\_SBP. Residues whose sidechains interact with the L-serinamide ligand in MeAmi\_SBP are highlighted with an asterisk, with blue asterisks referring to residues on the small core lobe of MeAmi\_SBP, and black asterisks referring to residues on the large core lobe of MeAmi\_SBP. **D.** Comparison of the binding cavities of MeAmi\_SBP and the dipeptide binding protein, DppA, from *Helicobacter pylori* (PDB 6PU3), highlighting similarities (e.g. positioning of

Asp413/447) and differences (e.g. disruption of beta-strand in *MeAmi\_SBP*) that contribute to differing ligand specificities.

***MeAmi\_SBP* has a  $\text{Ca}^{2+}$  binding site.** The crystal structure of *MeAmi\_SBP* also revealed clear electron density near residues Asp45, Tyr207 and Asp49 that was consistent with a metal ion binding site (**Figure 4A, Supplementary Figure 7**). Coordination geometry (pentagonal bipyramidal), environmental B-factors and bond lengths were most consistent with this being a  $\text{Ca}^{2+}$  ion (**Supplementary Figure 8**).  $\text{Ca}^{2+}$  binding to His<sub>6</sub>-tagged *MeAmi\_SBP* was subsequently confirmed using inductively coupled plasma mass spectrometry (ICP-MS) and isothermal titration calorimetry (ITC). Specifically, calcium was detected by ICP-MS at 17.8 parts per billion in a purified sample of 10  $\mu\text{M}$  *MeAmi\_SBP* (**Figure 4B**). While high concentrations of nickel and zinc were also detected in this sample of His<sub>6</sub>-tagged *MeAmi\_SBP*, the presence of nickel and zinc could be attributed to these metal ions binding the His<sub>6</sub>-tag rather than *MeAmi\_SBP*: removal of the His<sub>6</sub>-tag abolished nickel, zinc and copper binding (**Supplementary Figure 9**). Further, ITC titrations showed that  $\text{Ca}^{2+}$  binds His<sub>6</sub>-cleaved *MeAmi\_SBP* with high affinity ( $K_D$  2.80  $\mu\text{M}$ ) (**Figure 4C**), which aligns with the metal coordination properties of another calcium-coordinating SBP in SBP\_bac\_5 (53) and similar (low  $\mu\text{M}$   $K_D$ ) metal-binding affinity identified for the primary metal cofactor of different class members in the broader superfamily (54,55). While  $\text{Ca}^{2+}$  increased ligand-induced thermal shifts for L-alaninamide, glycineamide and L-tyrosinamide, it slightly decreased the thermostability of apo-*MeAmi\_SBP* (**Figure 4D**). This may be attributed to altered structural dynamics that decrease thermostability, which is consistent with the increase in *MeAmi\_SBP* entropy conferred by  $\text{Ca}^{2+}$  binding observed using ITC. These data suggest a role for  $\text{Ca}^{2+}$  as a structural metal ion in ligand-bound *MeAmi\_SBP*.

In *MeAmi\_SBP*,  $\text{Ca}^{2+}$  is coordinated by the side chain of Asp45, backbone carbonyl of Lys41, backbone carbonyl of Phe44, and four crystallographic waters that are held in place by the backbones and side chains of the surrounding residues (Ser42, Tyr207, Asp515, Asp49, Asp60 and His164). These residues are highly conserved across the other 609 SBPs found in the cluster with *MeAmi\_SBP* (**Figure 4E**), suggesting that homologous SBPs probably bind  $\text{Ca}^{2+}$  at this position as well.





**Figure 4. A conserved Ca<sup>2+</sup> binding site in MeAmi\_SBP.** **A.** Ca<sup>2+</sup>-binding site coordination geometry and distances (in Å) in the MeAmi\_SBP crystal structure. **B.** ICP-MS analysis of the inorganic elemental content of MeAmi\_SBP directly after SEC purification. **C.** ITC thermogram of His<sub>6</sub>-cleaved MeAmi\_SBP titrated with CaCl<sub>2</sub> after removal of copurified Ca<sup>2+</sup> ions by dialysis using EDTA. **D.** Protein melting temperatures measured using DSF for MeAmi\_SBP in the presence of several amino acid amides  $\pm$  1 mM CaCl<sub>2</sub>. \*\*\*\* refers to  $p < 0.0001$  and \*\* refers to  $p < 0.01$  from unpaired parametric t-tests. **E.** Sequence WebLogo (52) showing conservation of residues within the Ca<sup>2+</sup> coordination site in all homologs found in the same sequence cluster as MeAmi\_SBP. Residues whose sidechains directly interact with the Ca<sup>2+</sup> or the coordinating water molecules in MeAmi\_SBP are highlighted with an asterisk.



## DISCUSSION

The SBP\_bac\_5 family comprises periplasmic SBPs that are associated with ABC transport systems and that bind a range of substrates, including peptides, heme, nickel, opines, and other molecules (20–25). Our sequence similarity network analysis of this family highlighted that most of the sequence and function space of this family remains uncharacterized (**Figure 1**). In this work, we focused on a set of 610 uncharacterized proteins within the SBP\_bac\_5 family that are all found in conserved operons comprising genes encoding ABC transport systems and a putative amidase from the FmdA\_AmdA family. Considering amides represent a valuable source of carbon and nitrogen for bacteria, and our understanding of import systems specific for such compounds is poor, we aimed to determine the ligand specificity of these homologous SBPs to provide insight into these amidase-associated uptake systems.

We found that although *MeAmi*\_SBP, the most consensus-like representative from this subgroup, is annotated as a peptide-binding protein, it showed no ligand-induced thermal shifts in the presence of any of the di- or tri- peptides screened (**Figure 2**). We experimentally identified that *MeAmi*\_SBP instead binds the L-amino acid amides L-serinamide, glycineamide, L-tyrosinamide and L-alaninamide with high (nM- $\mu$ M  $K_D$ ) affinity (**Figure 2**). The structure of *MeAmi*\_SBP highlights the structural determinants for specificity for the amidated amino acids over the canonical L-amino acids and other amide-containing molecules. Distinct specificity for amino acid amides is conferred through a combination of (i) a salt bridge between the N-terminal amine and Asp413, (ii) hydrogen bonds between the C-terminal carbonyl and Trp34 and Ile411 and (iii) hydrogen bonds between the carboxamide group, Tyr32 and Met408 (**Figure 3**). This three-component interaction determines ligand specificity, and ligand affinity is further tuned across two orders of magnitude based on side chain interaction compatibility with a hydrophobic binding site pocket, best highlighted by the differences in affinity between glycineamide and L-tyrosinamide. *MeAmi*\_SBP likely diverges from its computational annotation as a peptide-binding protein due to a loop insertion in the binding site that reduces the volume of the pocket and occludes larger molecules that bind to homologs such as *H. pylori* DppA (49) (**Figure 3**). However, it is important to note that *MeAmi*\_SBP may also bind to other compounds not tested in this study that contain L-amino acid amide-like motifs. Most significantly, this work presents the first structural and experimental evidence for periplasmic SBP family proteins

binding to L-amino acid amides. The conservation of binding site residues across the *MeAmi\_SBP* cluster in the SSN indicates homologous binding properties exist in other bacteria.

The structure of *MeAmi\_SBP* also revealed a  $\text{Ca}^{2+}$  binding site within its main small lobe, which was subsequently confirmed by both ITC and ICP-MS (**Figure 4**). While the structural role of calcium is evident, its physiological significance in these proteins remains uncertain. Most significantly, this work presents the first evidence for *SBP\_bac\_5* proteins implicated in calcium binding through multiple experimental methods, although calcium binding has been identified in other classes of the periplasmic SBP superfamily previously (56–58). The evidence for calcium binding having a structural role in *MeAmi\_SBP* is compelling because the ion identified in the crystal structure was adventitiously bound during expression and proven to interact with *MeAmi\_SBP* using ICP-MS and ITC. The conservation of coordination site residues across the *MeAmi\_SBP* cluster in the SSN indicates calcium binding is a likely conserved property in the homologous proteins from other bacterial species.

The prevalence of Gram-negative nitrogen-fixing soil bacteria within this group, some of which are known to thrive on short chain amines (59), suggests an evolutionary advantage for soil bacteria capable of catalyzing amidase reactions on L-amino acid amides. This would metabolize them to the respective L-amino acid and ammonia, facilitating access to amino acid utilization pathways such as the citric acid cycle (60) and ammonia assimilation pathways such as the glutamine synthetase-glutamate synthase pathway (61), respectively. Characterizing the substrate specificity of the amidases found within these operons would help provide further insight into amide metabolism in these bacteria. This might also identify enzymes capable of producing enantiomerically pure L-amino acids from racemic amino acid amides.

The conserved operon structure and the presence of ANTAR protein and an additional cytosolic PBP in the operon also hints that the expression of these genes may be regulated via an ANTAR-dependent mechanism, like that of the amidase operon in *Pseudomonas aeruginosa* (35,36,41). Sequence and structural analyses of *MeAmi\_ANTAR* and *MeAmi\_cPBP* shows similarity to the *P. aeruginosa* AmiC/AmiR complex, but further research is needed to identify whether it responds to L-amino acid amides and to confirm the ANTAR target sequences, as canonical dual hexaloop motif sequences (62) could not be identified in this work. Future studies should determine if the system is induced by L-amino acid amides or other amide-

containing compounds. This will expand our understanding of ANTAR regulated ABC transporter operons.

Overall, these findings contribute to our understanding of ABC transporter-associated SBPs binding to amides and their regulation in bacteria. The most compelling aspect of this work is that we provide the first experimental evidence for a cluster of periplasmic SBPs that bind L-amino acid amides. This evidence, in conjunction with the genetic elements co-located with *MeAmi\_SBP* and sequence homology to co-clustered proteins from the source SSN, expands our understanding of microbial interactions with L-amino acid amides and general bacterial nitrogen utilization.

## METHODS

Key chemicals, reagents, recipes, and software are listed in Supplementary Information (**Supplementary Table 8-10**).

### Sequence Similarity Network and Bioinformatics

A sequence similarity network (SSN) of Pfam PF00496 (“Bacterial extracellular solute-binding proteins, family 5 Middle”) was calculated using EFI-EST (28) with the UniRef50 database. Default input settings were used. After the initial calculation, networks were calculated at a range of alignment score thresholds, with no sequence length restrictions. SSNs were visualized using the YFiles Organic Layout in Cytoscape 3.9.1 (63). The stringency of the SSN was increased until all unique SWISS-PROT descriptions were sorted into separate clusters of the SSN; this occurred at an alignment score of 100. The resulting isofunctional SSN was submitted to the EFI-GNT server (46) for analysis of per-cluster genomic context. *MeAmi\_SBP* was selected as a representative of the target cluster using HMMER (43) and a hidden Markov model calculated from the MAFFT (64) alignment of all sequences in the cluster of interest.

### Protein expression and purification.

SignalP 6.0 (65) was used to identify the signal peptide cleavage site between residues 25 and 26 of *MeAmi\_SBP* sequence (probability=0.97) (**Supplementary Figure 10**). The signaling peptide (MKKFLASTVAASALALMLGMTSARA) was removed from the sequence encoding *MeAmi\_SBP* and a DNA sequence encoding an N-terminal His<sub>6</sub> tag and TEV protease cleavage

site (GSSHHHHHHSSGENLYFQG) was added (**Supplementary Table 11**). This was codon-optimized for expression in *E. coli*, synthesized and cloned between NdeI and XhoI sites in pET-29b(+) vectors by Twist Bioscience (**Supplementary Table 12**). The plasmid was transformed into chemically-competent *E. coli* BL21(DE3) (NEB) using the recommended heat-shock protocol and plated onto LB Miller agar supplemented with 50 mg mL<sup>-1</sup> kanamycin before being incubated overnight at 37 °C.

A single colony was picked and used to inoculate a seed culture in 10 mL LB media supplemented with 50 mg mL<sup>-1</sup> kanamycin. The seed culture was incubated at 37°C for 8-12 hours with shaking at 200 rpm and then diluted 1:100 to inoculate a 1 L culture of simple autoinduction media (5 g L<sup>-1</sup> Yeast Extract, 20 g L<sup>-1</sup> Tryptone, 5 g L<sup>-1</sup> NaCl, 6 g L<sup>-1</sup> Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O, 3 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.19 g L<sup>-1</sup> MgCl<sub>2</sub>, 6 g L<sup>-1</sup> glycerol, 0.5 g L<sup>-1</sup> glucose, 2 g L<sup>-1</sup> D-lactose) supplemented with 50 mg mL<sup>-1</sup> kanamycin in Thomson Ultra Yield Flasks (66). This was incubated at 30°C for 24 hours with shaking at 250 rpm. Cell pellets were harvested by centrifugation at 4730 rcf for 15 min at 20°C. Cell pellets were stored at -20 °C until purification.

MeAmi\_SBP was purified using immobilized metal affinity chromatography (IMAC) followed by size exclusion chromatography (SEC). Cell pellets were resuspended in 25 mL equilibration buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4). Cells were lysed by sonication (2 x 5 min at 50% power), followed by addition of 0.5 µL Turbonuclease from *Serratia marcescens* (Sigma Aldrich) and centrifugation at 17400 rcf for 75 min at 4°C to remove cell debris. The supernatant was filtered through a 0.45 µm syringe filter prior to loading onto a 5 mL HisTrap HP (Cytiva) column pre-equilibrated with equilibration buffer at 2.5 mL min<sup>-1</sup>. Following loading, the column was washed with 10 column volumes of equilibration buffer, followed by elution of the target protein with five columns of IMAC elution buffer (500 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4). Eluate fractions were pooled and loaded onto a size exclusion column (HiLoad Superdex 26/600 200 prep grade, GE Healthcare) pre-equilibrated with Tris buffer (50 mM Tris, 300 mM NaCl, pH 7.4) or HEPES buffer (50 mM HEPES, 300 mM NaCl, pH 7.4), depending on downstream use. SDS-PAGE (SurePAGE™, Bis-Tris, 10×8, 4-12%, GenScript) was used to assess protein purity, following which protein concentration was determined using a NanoDrop One<sup>c</sup> Microvolume UV-Vis Spectrophotometer (ThermoFisher Scientific) using the predicted molar extinction coefficient from ProtParam (91,790 M<sup>-1</sup> cm<sup>-1</sup>).

## Differential scanning fluorimetry (DSF)

For screening, Biolog Microbial Phenotyping Microarray Plates PM1-PM6 were resuspended in 125  $\mu$ L MilliQ water and 5  $\mu$ L of each well was aliquoted into MicroAmp Optical 96-Well Reaction Plates (ThermoFisher Scientific) to prepare screening plates. DSF experiments were performed in Tris buffer (50 mM Tris, 300 mM NaCl, pH 7.4) using a final protein concentration of 0.5 mg mL<sup>-1</sup>. A 1.33  $\times$  master mix was prepared containing *MeAmi*\_SBP, Protein Thermal Shift buffer, Protein Thermal Shift dye and Tris buffer according to the Protein Thermal Shift™ Dye Kit (ThermoFisher Scientific) protocols. 15  $\mu$ L of master mix was aliquoted into each well of the prepared screening plates to a final volume of 20  $\mu$ L. Prepared plates were centrifuged at 2250 rcf for 1 min and sealed using MicroAmp Optical Adhesive Film (ThermoFisher Scientific). Melt experiments were performed using a QuantStudio Real-Time PCR thermocycler (ThermoFisher Scientific). Samples were heated from 25 °C to 95 °C at 0.05 °C s<sup>-1</sup>. Protein Thermal Shift Software (ThermoFisher Scientific) was used for data analysis. The midpoint of the unfolding transition was obtained from the maximum of the first derivative of fluorescence.

For testing of additional ligands, each ligand was prepared at 100 mM in Tris buffer and 2  $\mu$ L was aliquoted into MicroAmp Optical 96-Well Reaction Plates (ThermoFisher Scientific). Plates were then prepared using 18  $\mu$ L of a 1.11  $\times$  master mix according to the Protein Thermal Shift™ Dye Kit (ThermoFisher Scientific).

For isothermal analysis, 2  $\mu$ L of serially diluted ligands were aliquoted into MicroAmp Optical 96-Well Reaction Plates (ThermoFisher Scientific). Plates were then prepared using 18  $\mu$ L of a 1.11  $\times$  master mix according to the Protein Thermal Shift™ Dye Kit (ThermoFisher Scientific) recommendations. Protein concentration was 8.4  $\mu$ M in each well. Data analysis was performed using FoldAffinity (48). For each isothermal analysis, the signal D window range was 40-85 °C. A global model was used to fit fluorescence with a  $\Delta$ CP of 0 kcal K<sup>-1</sup> mol<sup>-1</sup>, and a one-site model was used for the isothermal fit. Results were exported and plotted in GraphPad Prism.

## TEV Protease (TEVp)-mediated cleavage of N-terminal His<sub>6</sub> tag

The His<sub>7</sub>-TEV (L56V, S135G, S219V) variant of the Tobacco Etch Virus protease (TEVp) was expressed and purified as described elsewhere (67). TEVp-cleavage of IMAC-purified His<sub>6</sub>-*MeAmi*\_SBP was performed overnight in TEVp-cleavage buffer (10 mM Tris, 50 mM NaCl, 0.5

mM EDTA, 1 mM DTT, pH 8.0) using a protease to substrate mass ratio of 1:10. The following day, the cleaved substrate was recovered from the reaction by using a 5 mL HisTrap HP (Cytiva) column pre-equilibrated with equilibration buffer to remove uncleaved substrate, the cleaved His<sub>6</sub>-tag, and the His<sub>7</sub>-tagged protease, followed by size exclusion chromatography as described previously.

### **Isothermal titration calorimetry (ITC)**

His<sub>6</sub>-cleaved *MeAmi*\_SBP was dialyzed extensively prior to ITC to remove Ca<sup>2+</sup> ions and adventitiously bound ligand obtained via copurification: protein was dialyzed for 24 hours against 5 L of 50 mM HEPES, 300 mM NaCl, 10 mM EDTA, pH 7.4, followed by four passes against 5 L of 5 mM HEPES, 300 mM NaCl, pH 7.4, followed by one pass against 5 L of ITC buffer (50 mM HEPES, 300 mM NaCl, pH 7.4, 0.02% TWEEN20). Protein was concentrated to 10 mL using an Amicon Ultra-15 10K MWCO Centrifugal Filter Unit (Merck Millipore) and loaded into SnakeSkin Dialysis Tubing, 10K MWCO, 35 mm dry I.D. (ThermoFisher Scientific).

All titrations were performed in a TA Instruments Benchtop Nano ITC (small volume). All stocks and dilutions used for ITC were prepared using exactly matched buffer from the final round of dialysis. Proteins were concentrated using an Amicon Ultra-15 10K MWCO Centrifugal Filter Unit (Merck Millipore). Protein and ligand samples were degassed prior to each ITC experiment. All titrations were performed as titrations of ligand into protein at 25°C with a stirring rate of 350 rpm. Initial and final baselines were generated over 120 s. The first injection of each titration was a 1 µL dummy injection, followed by 22 injections of 2 µL each, with an injection interval of 300 s. The background heat was estimated as the average heat associated with each injection in a control titration of ligand into buffer and subtracted from each titration. Results were analyzed in NITPIC (68) for baseline detection, blank subtraction, and integration of baseline-subtracted power. SEDPHAT (69) was used to fit integrated heats to the single binding site model ( $A + B \longleftrightarrow AB$ , heteroassociation) by global fitting with the Simplex algorithm. Thermograms were plotted in GUSSI (70).

### **AlphaFold2 Structural Models**

An AlphaFold2 model of *MeAmi*\_SBP was generated using the “AlphaFold2\_mmseqs2” ColabFold Notebook (71), which uses AlphaFold2 (42) and MMSeqs2 (72). The sequence corresponding to the TEVp-cleaved *MeAmi*\_SBP protein was used. Default settings were used.



Structural figures were prepared using PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

## Protein Crystallization and Data Collection

The His<sub>6</sub>-cleaved *MeAmi* was purified through size exclusion chromatography, as described above, using 4× crystallization buffer (40mM Tris, 150mM NaCl, pH 7.4). The purified protein sample was concentrated to 40 mg mL<sup>-1</sup> using Amicon Ultra-15 3K MWCO Centrifugal Filter Unit (Merck Millipore) and subsequently diluted 1:4 with MilliQ water (yielding 10 mg mL<sup>-1</sup> of protein in 1x crystal buffer (10mM Tris, 37.5 mM NaCl, pH 7.4). Sparse-matrix crystallography trials with ShotGun SG1 (Molecular Dimensions) and Index (Hampton Research) screens were performed at 18°C using sitting-drops of 400 nl protein mixed with 400 nl reservoir solution (prepared using a Formulatrix NT8).

A single large, cubic crystal formed after 3 months in well H5 of the ShotGun SG1 screen (30% w/v PEG 4000). The crystal was mounted onto a loop and set on goniometer without cryoprotecting or flash cooling. Diffraction data were collected in-house using an Agilent SuperNova X-ray source and processed using CrysAlisPro (Oxford Diffraction /Agilent Technologies UK Ltd, Yarnton, England) in space group P 2<sub>1</sub> 2<sub>1</sub> 2<sub>1</sub>. Data were truncated using Aimless (73) until statistics were of suitable quality in the outer shell. The structures were solved by molecular replacement (Phaser MR, CCP4 (74)) using the two domains of the AF2-predicted model of *MeAmi*\_SBP (domain 1= residues 1-265, 488-523; domain 2= residues 266-488) as the search models. AF2 model pLDDT values were converted to B-factors in Phenix prior to molecular replacement. Refinement was carried out in Phenix.refine (75). Multiple rounds of refinement and model building in Coot (76) were carried out until R-factors converged and the structure was modeled as best as possible to the electron density. MolProbity (77) was used for structure validation. In the final structure, residue number 1 corresponds to the first residue (glycine) in the TEVp-cleaved *MeAmi*\_SBP sequence. The metal binding site was analyzed using the CheckMyMetal (CMM): Metal Binding Site Validation Server (78). Polder omit maps were generated using phenix.polder (79) with default settings.

## FoldSeek

Similar structures were searched for using FoldSeek (80). All available databases were searched with Foldseek in 3Di/AA mode.



## Computational Docking of Ligands into *MeAmi\_SBP*

Schrödinger (Schrödinger Release 2022-1: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2021.) was used for computational docking. Briefly, the crystal structure of *MeAmi\_SBP* was prepared using default settings in the protein preparation wizard (pH 7.4, including a minimization step) and the ligands were prepared using the LigPrep wizard and default settings (retaining stereochemistry). The ligands were then docked into the *MeAmi\_SBP* using the induced fit protocol with default settings. The OPLS4 Force Field (81) was used.

## Inductively Coupled Plasma Mass Spectrometry (ICP-MS)

ICP-MS was performed by technical staff at the Australian National University Research School of Earth Sciences ICP-MS Research Facility. 5 mL of 10  $\mu$ M *MeAmi\_SBP* was prepared in MS buffer (50 mM HEPES, 300 mM NaCl, pH 7.4). Both sample and suspension blank were diluted 1:10 in triplicate using 2% HNO<sub>3</sub> to acidify samples and precipitate protein component. One of each triplicate was spiked with 100 ppb of Agilent Intelliquant 68 multi-element standard No.1 for ICP (IQ1), 5% HNO<sub>3</sub>, 100  $\mu$ g mL<sup>-1</sup> and 100 ppb of Agilent Intelliquant 12 multi-element standard No.2 for ICP, 5% HNO<sub>3</sub>, 100  $\mu$ g mL<sup>-1</sup> (IQ2). Solutions were first run on an Agilent 5110 OES to determine elemental composition of solutions. Samples were then run on a ThermoFisher iCapRQ Q-ICP-MS, excluding Na because this was very high in the suspension liquid. Final concentrations were processed by suppressing the 100 ppb addition of IQ1 and IQ2 and subtracting components of the blank sample.

## DATA AVAILABILITY

The 1.5 Å P<sub>21</sub> × P<sub>21</sub> × P<sub>21</sub> crystal structure of the *MeAmi\_SBP*–L-serinamide complex was submitted to the Protein Data Bank (PDB ID 8UPI). Raw data and associated files are available at doi: 10.5281/zenodo.10070929. Any other data is available upon request.

## AUTHOR CONTRIBUTIONS

OBS performed protein expression, purification, and experiments. RLF performed crystallography experiments, collected X-ray diffraction data, and processed the data. MGR performed protein expression, purification and initial DSF experiments. OBS, JAK and CJJ wrote the manuscript. CJJ supervised the project and provided support throughout. All authors edited and contributed to the final manuscript. All authors have read and approved the final version of the manuscript.

## ACKNOWLEDGEMENTS

This research was conducted by the Australian Research Council Centre of Excellence in Synthetic Biology (project number CE200100029) and funded by the Australian Government. We thank the Australian Research Council Centre of Excellence in Synthetic Biology for seed funding that supported this project. We thank Brett Knowles and Robin Grunn at the Australian National University Research School of Earth Sciences ICP-MS Research Facility for performing the ICP-MS experiments and for helpful discussions. We thank Dr Michael Gardiner for assistance with X-ray crystallography.

## REFERENCES

1. Davidson AL, Dassa E, Orelle C, Chen J. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol Mol Biol Rev.* 2008 Jun;72(2):317–64.
2. Boos W, Shuman H. Maltose/Maltodextrin System of *Escherichia coli*: Transport, Metabolism, and Regulation. *Microbiol Mol Biol Rev.* 1998 Mar;62(1):204–29.
3. Ejby M, Fredslund F, Andersen JM, Vujičić Žagar A, Henriksen JR, Andersen TL, et al. An ATP Binding Cassette Transporter Mediates the Uptake of  $\alpha$ -(1,6)-Linked Dietary Oligosaccharides in *Bifidobacterium* and Correlates with Competitive Growth on These Substrates. *J Biol Chem.* 2016 Sep 16;291(38):20220–31.
4. Webb AJ, Homer KA, Hosie AHF. Two Closely Related ABC Transporters in *Streptococcus mutans* Are Involved in Disaccharide and/or Oligosaccharide Uptake. *J Bacteriol.* 2008 Jan;190(1):168–78.
5. Lansky S, Salama R, Shulami S, Lavid N, Sen S, Schapiro I, et al. Carbohydrate-Binding Capability and Functional Conformational Changes of AbnE, an Arabino-oligosaccharide Binding Protein. *J Mol Biol.* 2020 Mar 27;432(7):2099–120.

6. Monnet V. Bacterial oligopeptide-binding proteins. *Cell Mol Life Sci CMLS*. 2003 Oct;60(10):2100–14.
7. Bareia T, Pollak S, Guler P, Puyesky S, Eldar A. Major distinctions between the two oligopeptide permease systems of *Bacillus subtilis* with respect to signaling, development and evolutionary divergence. *Microbiology*. 2023 Sep 27;169(9):001382.
8. Hosie a H, Poole PS. Bacterial ABC transporters of amino acids. *Res Microbiol*. 2001;152(3–4):259–70.
9. Nishijyo T, Park SM, Lu CD, Itoh Y, Abdelal AT. Molecular Characterization and Regulation of an Operon Encoding a System for Transport of Arginine and Ornithine and the ArgR Regulatory Protein in *Pseudomonas aeruginosa*. *J Bacteriol*. 1998 Nov;180(21):5559–66.
10. Matsubara K, Ohnishi K, Kiritani K. Nucleotide sequences and characterization of liv genes encoding components of the high-affinity branched-chain amino acid transport system in *Salmonella typhimurium*. *J Biochem (Tokyo)*. 1992 Jul;112(1):93–101.
11. Tomii K, Kanehisa M. A comparative analysis of ABC transporters in complete microbial genomes. *Genome Res*. 1998 Oct;8(10):1048–59.
12. De Boer M, Gouridis G, Vietrov R, Begg SL, Schuurman-Wolters GK, Husada F, et al. Conformational and dynamic plasticity in substrate-binding proteins underlies selective transport in ABC importers. *eLife*. 2019 Mar;8:1–28.
13. Scheepers GH, Lycklama a Nijeholt JA, Poolman B. An updated structural classification of substrate-binding proteins. *FEBS Lett*. 2016;590(23):4393–401.
14. Doeven MK, Abele R, Tampé R, Poolman B. The Binding Specificity of OppA Determines the Selectivity of the Oligopeptide ATP-binding Cassette Transporter\*. *J Biol Chem*. 2004 Jul 30;279(31):32301–7.
15. Yu J, Ge J, Heuveling J, Schneider E, Yang M. Structural basis for substrate specificity of an amino acid ABC transporter. *Proc Natl Acad Sci*. 2015 Apr 21;112(16):5243–8.
16. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. *Nucleic Acids Res*. 2021 Jan 8;49(D1):D412–9.
17. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, et al. InterPro in 2022. *Nucleic Acids Res*. 2023 Jan 6;51(D1):D418–27.
18. 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 Feb;286(1):279–90.
19. Cerna-Vargas JP, Sánchez-Romera B, Matilla MA, Ortega Á, Krell T. Sensing preferences for prokaryotic solute binding protein families. *Microb Biotechnol*. 2023 Aug 7;16(9):1823–33.
20. Berntsson RPA, 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 May 1;28(9):1332–40.
21. Olson ER, Dunyak DS, Jurss LM, Poorman RA. Identification and characterization of dppA, an *Escherichia coli* gene encoding a periplasmic dipeptide transport protein. *J Bacteriol*. 1991 Jan;173(1):234–44.

22. Wang Z, Xia X, Zhang M, Fang J, Li Y, Zhang M. Purification and Characterization of Glutathione Binding Protein GsiB from *Escherichia coli*. *BioMed Res Int*. 2018 Nov 1;2018:3429569.
23. Vergauwen B, Elegheert J, Dansercoer A, Devreese B, Savvides SN. Glutathione import in *Haemophilus influenzae* Rd is primed by the periplasmic heme-binding protein HbpA. *Proc Natl Acad Sci*. 2010;107(30):13270–5.
24. Navarro C, Wu LF, Mandrand-Berthelot MA. The nik operon of *Escherichia coli* encodes a periplasmic binding-protein-dependent transport system for nickel. *Mol Microbiol*. 1993 Sep;9(6):1181–91.
25. Song L, Zhang Y, Chen W, Gu T, Zhang SY, Ji Q. Mechanistic insights into staphylopin-mediated metal acquisition. *Proc Natl Acad Sci U S A*. 2018 Apr 10;115(15):3942–7.
26. Atkinson HJ, Morris JH, Ferrin TE, Babbitt PC. Using Sequence Similarity Networks for Visualization of Relationships Across Diverse Protein Superfamilies. *PLOS ONE*. 2009 Feb 3;4(2):e4345.
27. Copp JN, Akiva E, Babbitt PC, Tokuriki N. Revealing Unexplored Sequence-Function Space Using Sequence Similarity Networks. *Biochemistry*. 2018 Aug 7;57(31):4651–62.
28. Zallot R, Oberg N, Gerlt JA. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry*. 2019 Oct 15;58(41):4169–82.
29. Draper P. The Aliphatic Acylamide Amidohydrolase of *Mycobacterium smegmatis*: Its Inducible Nature and Relation to Acyl-Transfer to Hydroxylamine. *Microbiology*. 1967;46(1):111–23.
30. Wyborn NR, Mills J, Williams SG, Jones CW. Molecular characterisation of formamidase from *Methylophilus methylotrophus*. *Eur J Biochem*. 1996 Sep 1;240(2):314–22.
31. Gonsalvez IS, Isupov MN, Littlechild JA. Crystallization and preliminary X-ray analysis of a gamma-lactamase. *Acta Crystallogr D Biol Crystallogr*. 2001 Feb;57(Pt 2):284–6.
32. Garrido-Oter R, Nakano RT, Dombrowski N, Ma KW, McHardy AC, Schulze-Lefert P. Modular Traits of the Rhizobiales Root Microbiota and Their Evolutionary Relationship with Symbiotic Rhizobia. *Cell Host Microbe*. 2018 Jul 11;24(1):155-167.e5.
33. Demtröder L, Pfänder Y, Masepohl B. *Rhodobacter capsulatus* AnfA is essential for production of Fe-nitrogenase proteins but dispensable for cofactor biosynthesis and electron supply. *MicrobiologyOpen*. 2020 Mar 23;9(6):e1033.
34. Tatsukami Y, Nambu M, Morisaka H, Kuroda K, Ueda M. Disclosure of the differences of *Mesorhizobium loti* under the free-living and symbiotic conditions by comparative proteome analysis without bacteroid isolation. *BMC Microbiol*. 2013 Jul 31;13(1):180.
35. Wilson SA, Drew RE. Transcriptional analysis of the amidase operon from *Pseudomonas aeruginosa*. *J Bacteriol*. 1995 Jun;177(11):3052–7.
36. Norman RA, Richard A. Norman, Poh CL, Pearl LH, O'Hara BP, Drew R. Steric hindrance regulation of the *Pseudomonas aeruginosa* amidase operon. *J Biol Chem*. 2000 Sep 29;275(39):30660–7.
37. Chebrou H, Bigey F, Arnaud A, Galzy P. Amide metabolism: a putative ABC transporter in *Rhodococcus* sp. R312. *Gene*. 1996 Dec 5;182(1–2):215–8.

38. Parish T, Mahenthiralingam E, Draper P, Davis EO, Colston EO. Regulation of the inducible acetamidase gene of *Mycobacterium smegmatis*. *Microbiol Read Engl*. 1997 Jul;143 ( Pt 7):2267–76.
39. Arunkumar Venkatesan, Venkatesan A, Kannan Palaniyandi, Palaniyandi K, Sujatha Narayanan, Narayanan S. Molecular characterization of AmiC, a positive regulator in acetamidase operon of *Mycobacterium smegmatis*. *Cell Stress Chaperones*. 2018 Jul 1;23(4):539–50.
40. Sundararaman B, Palaniyandi K, Venkatesan A, Narayanan S. Expression, purification and functional characterization of AmiA of acetamidase operon of *Mycobacterium smegmatis*. *Microbiol Res*. 2014 Nov 1;169(11):873–80.
41. Wilson SA, Wachira SJ, Norman RA, Pearl LH, Drew RE. Transcription antitermination regulation of the *Pseudomonas aeruginosa* amidase operon. *EMBO J*. 1996 Nov 1;15(21):5907–16.
42. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021 Aug;596(7873):583–9.
43. Finn RD, Finn R, Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011;39(Web Server Issue):W29–37.
44. Consortium TU. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2020 Nov;49(D1):D480–9.
45. Wang Z, Zhang M, Shi X, Xiang Q. Purification and Characterization of an ATPase GsiA from *Salmonella enterica*. *BioMed Res Int*. 2017;2017:3076091.
46. Oberg N, Zallot R, Gerlt JA. EFI-EST, EFI-GNT, and EFI-CGFP: Enzyme Function Initiative (EFI) Web Resource for Genomic Enzymology Tools. *J Mol Biol*. 2023 Jul 15;435(14):168018.
47. Bochner BR. Global phenotypic characterization of bacteria. *Fems Microbiol Rev*. 2009 Jan;33(1):191–205.
48. Niebling S, Burastero O, Bürgi J, Günther C, Defelipe LA, Sander S, et al. FoldAffinity: binding affinities from nDSF experiments. *Sci Rep*. 2021 May 5;11(1):9572.
49. Rahman MM, Machuca MA, Khan MF, Barlow CK, Schittenhelm RB, Roujeinikova A. Molecular Basis of Unexpected Specificity of ABC Transporter-Associated Substrate-Binding Protein DppA from *Helicobacter pylori*. *J Bacteriol*. 2019 Oct 15;201(20):e00400-19.
50. Dunten P, Mowbray SL. Crystal structure of the dipeptide binding protein from *Escherichia coli* involved in active transport and chemotaxis. *Protein Sci Publ Protein Soc*. 1995 Nov;4(11):2327–34.
51. Kim Y, Zhou M, Shatsman S, Anderson W, Joachimiak A, Center for Structural Genomics of Infectious Diseases (CSGID). Crystal Structure of Periplasmic Dipeptide Transport Protein from *Yersinia pestis*. PDB ID 5F1Q. *Protein Data Bank*. 2015 Dec 23;
52. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004 Jun;14(6):1188–90.
53. Munshi P, Stanley CB, Ghimire-Rijal S, Lu X, Myles DA, Cuneo MJ. Molecular details of ligand selectivity determinants in a promiscuous  $\beta$ -glucan periplasmic binding protein. *BMC Struct Biol*. 2013 Oct 4;13(1):18.

54. Vyas MN, Jacobson BL, Quijcho FA. The calcium-binding site in the galactose chemoreceptor protein. Crystallographic and metal-binding studies. *J Biol Chem*. 1989 Dec 5;264(34):20817–21.
55. Ford BA, Michie KA, Paulsen IT, Mabbutt BC, Shah BS. Novel functional insights into a modified sugar-binding protein from *Synechococcus* MITS9220. *Sci Rep*. 2022;12(1):4805.
56. Vyas NK, Vyas MN, Quijcho FA. A novel calcium binding site in the galactose-binding protein of bacterial transport and chemotaxis. *Nature*. 1987 Jun 18;327(6123):635–8.
57. Culurgioni S, Harris G, Singh AK, King SJ, Walsh MA. Structural Basis for Regulation and Specificity of Fructooligosaccharide Import in *Streptococcus pneumoniae*. *Struct Lond Engl* 1993. 2017 Jan 3;25(1):79–93.
58. Lansky S, Salama R, Shulami S, Lavid N, Sen S, Schapiro I, et al. Carbohydrate-Binding Capability and Functional Conformational Changes of AbnE, an Arabino-oligosaccharide Binding Protein. *J Mol Biol*. 2020 Mar 27;432(7):2099–120.
59. Mosley OE, Gios E, Close M, Weaver L, Daughney C, Handley KM. Nitrogen cycling and microbial cooperation in the terrestrial subsurface. *ISME J*. 2022 Nov 1;16(11):2561–73.
60. Walshaw DL, Wilkinson A, Mundy M, Smith M, Poole PS. Regulation of the TCA cycle and the general amino acid permease by overflow metabolism in *Rhizobium leguminosarum*. Vol. 143, *Microbiology*. Microbiology Society; 1997. p. 2209–21.
61. Kondorosi A, Svab Z, Kiss GB, Dixon RA. Ammonia assimilation and nitrogen fixation in *Rhizobium meliloti*. *Mol Gen Genet MGG*. 1977 Jan 1;151(2):221–6.
62. Walshe JL, Siddiquee R, Patel K, Ataide SF. Structural characterization of the ANTAR antiterminator domain bound to RNA. *Nucleic Acids Res*. 2022 Mar 21;50(5):2889–904.
63. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res*. 2003 Nov;13(11):2498–504.
64. Katoh K, Misawa K, Kuma K ichi, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002 Jul 15;30(14):3059–66.
65. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol*. 2022 Jul;40(7):1023–5.
66. Studier FW. Protein production by auto-induction in high-density shaking cultures. *Protein Expr Purif*. 2005 May 1;41(1):207–34.
67. Raran-Kurussi S, Cherry S, Cherry S, Cherry S, Zhang D, Waugh DS. Removal of Affinity Tags with TEV Protease. *Methods Mol Biol*. 2017 Jan 1;1586:221–30.
68. Scheuermann TH, Brautigam CA. High-precision, automated integration of multiple isothermal titration calorimetric thermograms: new features of NITPIC. *Methods San Diego Calif*. 2015 Apr 1;76:87–98.
69. Zhao H, Piszczek G, Schuck P. SEDPHAT--a platform for global ITC analysis and global multi-method analysis of molecular interactions. *Methods San Diego Calif*. 2015 Apr;76:137–48.



70. Brautigam CA, Zhao H, Vargas C, Keller S, Schuck P. Integration and global analysis of isothermal titration calorimetry data for studying macromolecular interactions. *Nat Protoc*. 2016 Apr;11(5):882–94.
71. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods*. 2022 Jun;19(6):679–82.
72. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol*. 2017 Nov;35(11):1026–8.
73. Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr*. 2013 Jul;69(7):1204–14.
74. McCoy AJ. Phaser crystallographic software. *J Appl Crystallogr* [Internet]. 2007;40. Available from: <https://doi.org/10.1107/S0021889807021206>
75. Zwart PH, Afonine PV, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, et al. Automated structure solution with the PHENIX suite. *Methods Mol Biol Clifton NJ*. 2008;426:419–35.
76. Emsley P, Cowtan K, IUCr. Coot : model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004 Dec;60(12):2126–32.
77. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2010 Jan 1;66(Pt 1):12–21.
78. Zheng H, Cooper DR, Porebski PJ, Shabalin IG, Handing KB, Minor W. CheckMyMetal: a macromolecular metal-binding validation tool. *Acta Crystallogr Sect Struct Biol*. 2017 Mar 1;73(Pt 3):223–33.
79. Liebschner D, Afonine PV, Moriarty NW, Poon BK, Sobolev OV, Terwilliger TC, et al. Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr Sect Struct Biol*. 2017 Feb 1;73(Pt 2):148–57.
80. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, et al. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol*. 2023 May 8;1–4.
81. Lu C, Wu C, Ghoreishi D, Wei Chen, Chen W, Wang L, et al. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J Chem Theory Comput*. 2021 Jun 7;17(7):4291–300.