UNIVERSITY *of* York

This is a repository copy of *The structural basis for high-affinity uptake of lignin-derived aromatic compounds by proteobacterial TRAP transporters*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/186677/

Version: Published Version

Article:

Bisson, Claudine, Salmon, Robert C, West, Laura et al. (4 more authors) (2021) The structural basis for high-affinity uptake of lignin-derived aromatic compounds by proteobacterial TRAP transporters. FEBS Journal. pp. 436-456. ISSN 1742-464X

https://doi.org/10.1111/febs.16156

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

The structural basis for high-affinity uptake of lignin-derived aromatic compounds by proteobacterial TRAP transporters

Claudine Bisson[1],*, Robert C. Salmon[1], Laura West[2], John B. Rafferty[1], Andrew Hitchcock[1] (iD), Gavin H. Thomas[2] and David J. Kelly[1] (iD)

1 Department of Molecular Biology and Biotechnology, The University of Sheffield, UK
2 Department of Biology, University of York, UK

Correspondence

D. J. Kelly, Department of Molecular Biology and Biotechnology, The University of Sheffield, Firth Court, Western Bank, Sheffield S10 2TN, UK
Tel: +44 (0) 114 222 4414
E-mail: d.kelly@sheffield.ac.uk

Present address

* Centre for Ultrastructural Imaging, King's College London, New Hunt's House, Guy's Campus, London, SE1 1UL, UK

Claudine Bisson and Robert C. Salmon contributed equally to this article

The organic polymer lignin is a component of plant cell walls, which like (hemi)-cellulose is highly abundant in nature and relatively resistant to degradation. However, extracellular enzymes released by natural microbial consortia can cleave the β-aryl ether linkages in lignin, releasing monoaromatic phenylpropanoids that can be further catabolised by diverse species of bacteria. Biodegradation of lignin is therefore important in global carbon cycling, and its natural abundance also makes it an attractive biotechnological feedstock for the industrial production of commodity chemicals. Whilst the pathways for degradation of lignin-derived aromatics have been extensively characterised, much less is understood about how they are recognised and taken up from the environment. The purple phototrophic bacterium *Rhodopseudomonas palustris* can grow on a range of phenylpropanoid monomers and is a model organism for studying their uptake and breakdown. *R. palustris* encodes a tripartite ATP-independent periplasmic (TRAP) transporter (TarPQM) linked to genes encoding phenylpropanoid-degrading enzymes. The periplasmic solute-binding protein component of this transporter, TarP, has previously been shown to bind aromatic substrates. Here, we determine the high-resolution crystal structure of TarP from *R. palustris* as well as the structures of homologous proteins from the salt marsh bacterium *Sagittula stellata* and the halophile *Chromohalobacter salexigens*, which also grow on lignin-derived aromatics. In combination with tryptophan fluorescence ligand-binding assays, our ligand-bound co-crystal structures reveal the molecular basis for high-affinity recognition of phenylpropanoids by these TRAP transporters, which have potential for improving uptake of these compounds for biotechnological transformations of lignin.

Introduction

High-affinity solute uptake is an essential requirement for the survival of microorganisms in low nutrient environments. There are three major families of high-affinity transport systems that rely on an extra-cytoplasmic solute-binding protein (SBP) to recognise and bind specific cargo and traffic it to hydrophobic subunits embedded in the cytoplasmic membrane [1]. The ATP-binding cassette (ABC) transporters

discovered almost 50 years ago are by far the best characterised of these systems, and as their name suggests drive transport using ATP hydrolysis [2]. Conversely, the tripartite tricarboxylate transporters and the tripartite ATP-independent periplasmic (TRAP) transporters are SBP-dependent secondary transporters that use either the proton motive force or sodium (Na^+) ions for solute uptake [3]. Since their discovery by Forward *et al.* in 1997 [4], the advent of mass genome sequencing has revealed that TRAP transporters are widely encoded in bacteria and archaea [5]. They are key to the survival of human pathogens such as *Vibrio cholerae* [6] and *Haemophillus influenzae* [7] as well as being involved in plant virulence [8] and uptake of carbon sources in many environmental bacteria [9–11].

Tripartite ATP-independent periplasmic transporters are typically comprised of three subunits, a periplasmic SBP (the P subunit) and two membrane protein components (the Q and M subunits) [4,12,13]. The larger M subunit is composed of 12 transmembrane helices (TMHs) and is thought to form the translocation pore, whilst the function of the smaller four TMH Q subunit, which has both the N and C termini positioned in the cytoplasm, has not yet been definitively established [5,14]. Although the structure of the transmembrane pore-region and the mechanism of solute transport have yet to be determined for TRAP transporters, the structure and ligand specificity of several SBPs have been determined, revealing an expanding list of biologically important solutes [3,15,16].

Lignin has a structural role in maintaining cell wall stability in plants, and after (hemi-) cellulose is the most abundant organic polymer on the planet [17,18]. Because of its environmental abundance, lignin is also of great biotechnological interest as a potential starting material for the derivation of commercially valuable products [19]. However, like (hemi)-cellulose, lignin is generally resistant to degradation [20]. In the environment, lignin polymers are degraded upon breakage of β-aryl ether linkages by extracellular laccases and peroxidases, which are released by a consortium of aerobic bacteria and fungi [21–23]. This in turn leads to the release of a range of aromatic compounds, mainly phenylpropanoids, such as the hydroxycinnamates (HCMs) coumarate and ferulate [24], which are more readily available to bacteria as a source of carbon. Many organisms from various ecological niches can further degrade these lignin-derived aromatic compounds; of these, aerobic microorganisms such as members of the *Roseobacter* clade use the well-characterised β-ketoadipate pathway to convert protocatechuate derived from phenolic compounds such as lignin monomers to β-ketoadipate and subsequently to the tricarboxylic acid cycle intermediates succinyl-CoA and acetyl-CoA [25].

The metabolically versatile photosynthetic purple bacterium *Rhodopseudomonas palustris* anaerobically degrades lignin-derived phenylpropanoid compounds via a 'non-β-oxidation' pathway, which has been described in detail elsewhere [26–28]. The gene cluster that encodes these enzymes also encodes a TRAP transporter (TarPQM) and an ABC transporter (CouPSTU) (Fig. 1A), the expression of which are under the control of the coumarate responsive transcriptional regulator CouR [28–30]. The SBP from the CouPSTU ABC transporter (CouP_Rhp) has been shown to bind a subset of phenylpropanoids, with the structure of CouP in complex with ferulate revealing the basis of molecular specificity in this family of transporters [29,31]. The SBP of TarPQM (TarP_Rhp) also binds several aromatic ligands, including coumarate, with sub-micromolar affinity [29,32]. However, neither the structure of TarP_Rhp nor that of any of its close homologues from other bacteria have been determined; thus, the molecular basis of aromatic ligand specificity in TRAP transporters remains unknown.

In this study, we use bioinformatics and phylogenetic analysis to identify putative lignin-derived phenylpropanoid TRAP transporters from a range of environmental bacteria known to be enriched for the use of TRAP transporters in their biology [33]. In addition to TarP_Rhp, two other TarP homologues were selected for detailed analysis, the proteins from the halophilic γ-proteobacterium *Chromohalobacter salexigens* (Csal_0280; TarP_Csal) and the α-proteobacterial Roseobacterium *Sagittula stellata* E-37 (SSE37_24379; TarP_Sse). Both of these aerobic marine-dwelling organisms are known to grow on a range of lignin-derived aromatic compounds [34–37]. Of these, TarP_Sse has been shown to be required for growth of *S. stellata* E-37 on ferulate and coumarate, confirming its involvement with phenylpropanoid metabolism [37]. In this study, tryptophan fluorescence ligand-binding assays show that TarP_Csal and TarP_Sse bind to a range of lignin-derived aromatic compounds with subtly different specificities to each other and to that of TarP_Rhp. High-resolution crystal structures of the three different TarP proteins in complex with a range of aromatic compounds reveal a highly similar mode of binding involving recognition of the carboxyl group of the ligand via an invariant arginine residue, which is conserved in all known TRAP transporter SBPs. Comparison of the three structures to each other and with other TRAP SBPs reveals that
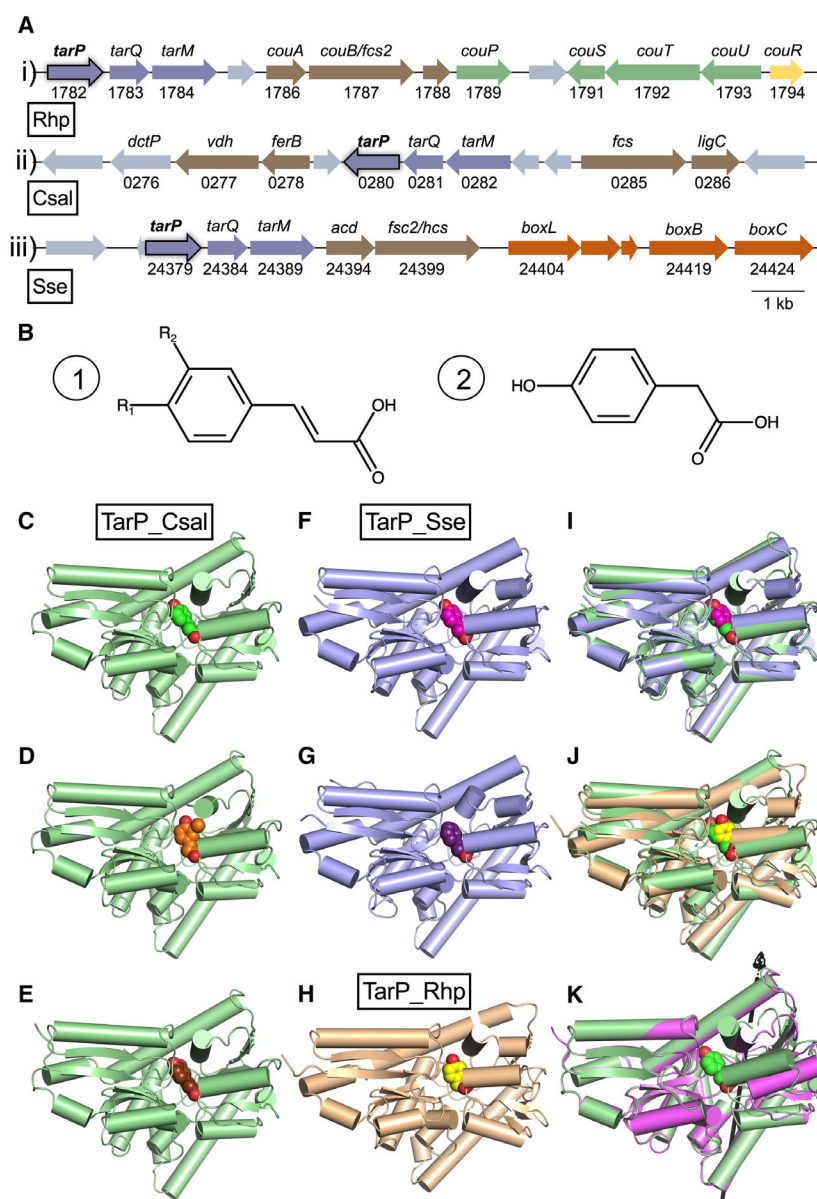
Fig. 1. The genomic context of *tarP* gene clusters, the molecular structure of HCM compounds used in this study and the conserved fold of TarP proteins. (A) The genetic arrangement of the *tarPQM* genes in the genomes of (i) *Rhodopseudomonas palustris* CGA009 (Rhp), (ii) *Chromohalobacter salexigens* DSM 3043 (Csal) and (iii) *Sagittula stellata* E-37 (Sse). Locus numbers and gene names are labelled below/above the corresponding gene arrows. The *tarPQM* genes are shown in mauve with the periplasmic binding protein subunit encoding *tarP* gene outlined in black. In panel (i), the *couPSTU* ABC transporter and *couR* transcriptional regulator genes are shown in green and yellow, respectively. In panel (iii), *box* genes encoding enzymes for benzoate degradation are shown in red. In all panels, genes encoding proteins that have known/hypothesised roles in HCM degradation are brown and genes encoding proteins with unknown roles in HCM degradation, genes not predicted to be involved in HCM degradation or hypothetical proteins of unknown function are grey. (B) (1) The general chemical structure of HCM compounds. R_1 and R_2 substituent group positions are at C4 and C3 in the phenyl ring, respectively. (2) The chemical structure of 4HPA. Images generated with Marvin JS (ChemAxon, Budapest, Hungary). (C–H) Overviews of the structures of TarP_Csal (green, C–E), TarP_Sse (blue, F, G) and TarP_Rhp (beige, H) in complex with ligands (coumarate, green/magenta; ferulate, orange; caffeate, brown; cinnamate, purple; and 4HPA, yellow). (I) A superposition of the TarP_Csal and TarP_Sse complexes with coumarate (coloured as in A and D). (J) A superposition of the TarP_Csal complex with coumarate and TarP_Rhp complex with 4HPA (coloured as in A and F). (K) An overlay of TarP_Csal in complex with coumarate and the apo structure of TarP_Csal (pink) superimposed on domain 1. The black arrow indicates the axis of rotation between the two domains. Helices are shown as cylinders, and ligands are shown as spheres to indicate the position of the binding pocket in each case. Images in C–K were generated using PYMOL.

amino acid substitutions elsewhere in the binding pocket confer ligand specificity, whilst comparison of TarP_Rhp with CouP_Rhp reveals how the same ligand cargo is recognised differently in a secondary TRAP transporter and a primary ABC transporter from the same organism. These data expand the structural and biochemical characterisation of SBPs for the uptake of lignin-derived aromatics and the role of TRAP import systems in the utilisation of ecologically and biotechnologically important HCM compounds.

Results and Discussion

Enrichment of TRAP transporters with potential specificity for HCMs in marine proteobacteria

Although TRAP transporters are widespread across the bacterial domain, we have previously noted their enrichment in phylogenetically diverse marine bacteria, presumably to enable scavenging of a diverse range of organic acids present at low concentrations in the ocean/aquatic environments [33]. Remarkably, some of these organisms have over 20 genes encoding TRAP SBPs, compared to just one in the model organisms *Escherichia coli* and *Bacillus subtilis* [33]. To assess the potential diversity of the transporters within these organisms, we extracted the sequences for all full-length TRAP SBPs from a subset of marine bacteria enriched for TRAP transporters [33], namely *C. salexigens* DSM 3043, *Aurantimonas manganoxydans* (formerly *Aurantimonas* sp. str. SI85-9A1), *Ruegeria pomeroyi* DSS-3 (formerly *Silicibacter pomeroyi* DSS-3), *Jannaschia* sp. CCS1, *S. stellata* E-37 and *Labrenzia aggregata IAM 12614* (see Table S1 for a list of genomes and geneID abbreviations). The sequences were aligned and compared to the TRAP SBP subset from the known HCM degrader *R. palustris* (including TarP_Rhp), and genome context analysis was used to identify the linkage of potential catabolic genes. One striking clade, represented by at least one example from each genome (Fig. S1A,B), shares a close relationship with mandelate-binding SBPs and are linked to genes with likely functions in HCM degradation. On this selection basis, Csal_0280 from *C. salexigens* (TarP_Csal) and SSE37_24379 from *S. stellata* (TarP_Sse) in addition to TarP_Rhp (see Fig. 1A for gene contexts) were recombinantly expressed and purified for further biochemical characterisation.

Binding assays indicate related patterns of ligand specificity in the TarP SBP family

To investigate ligand specificity, we screened TarP_Sse and TarP_Csal proteins by tryptophan fluorescence spectroscopy in the presence of equimolar amounts of coumarate, caffeate, cinnamate or ferulate. The basic structure of these ligands is an alpha, beta unsaturated, phenylpropanoate. The phenyl ring has additional hydroxyl or methoxy substituent groups at the C3 and/or C4 positions (Fig. 1B). A change in fluorescence emission upon exposure to the ligand indicated a ligand-induced conformational change in the environment of intrinsic tryptophan residues. Ligands that resulted in large quenches (25–40%) were selected for ligand titrations, monitoring the peak fluorescence emission over time, allowing determination of dissociation constants (see Fig. S2 for representative titrations and Table 1 for K_d values). Based on the phylogenetic analysis, we expected that ligand specificity in TarP_Sse and TarP_Csal would be more similar to each other and perhaps subtly different to TarP_Rhp, which is less closely related (Fig. S1B). A sequence alignment for all three proteins is shown in Fig. S3. The binding assays show that: (a) TarP_Rhp has an order of magnitude tighter binding than TarP_Sse and TarP_Csal to coumarate, caffeate and ferulate; (b) TarP_Rhp and TarP_Sse bind cinnamate with lower affinity than coumarate; and (c) TarP_Csal does not appear to bind cinnamate or ferulate under these assay conditions but has similar binding affinities for coumarate and caffeate as TarP_Sse. With these results in mind, we set out to determine crystal structures of various protein/ligand complexes to explain these apparent differences in binding affinity.

Table 1. Hydroxycinnamate binding data for TarP SBPs. For each protein, the left column shows if a liganded structure (LS) has been obtained for the protein and ligand combination indicated (Y = yes, N = no). Note that an apo structure of TarP_Csal was also obtained and a structure of TarP_Rhp with 4HPA intrinsically bound. The right column gives the calculated mean K_d (nM) $\pm$ the standard deviation from three independent ligand titrations. n.d. specifies that no K_d could be determined because a fluorescence change was not detectable.

Ligand	TarP_Sse		TarP_Csal		TarP_Rhp[a]	
	LS	K_d (nM)	LS	K_d (nM)	LS	K_d (nM)
Coumarate	Y	120 ± 4	Y	158 ± 26	N	8 ± 5
Caffeate	N	148 ± 10	Y	225 ± 85	N	15 ± 3
Cinnamate	Y	247 ± 21	N	n.d.	N	50, 33[b]
Ferulate	N	486 ± 72	Y	n.d.	N	14 ± 7

[a]The TarP data have been published previously [29] and are included in this table for comparison.; [b]Individual K_d values from two independent titrations with cinnamate.

The structure of the TarP homologues reveals a conserved architecture of the binding pocket

Despite the low nanomolar binding affinity of TarP_Rhp for coumarate, caffeate and ferulate ([29]; Table 1), it was not possible to co-crystallise it with any of these ligands. Crystallisation of TarP_Rhp appeared to be dependent on the formation of a pyroglutamate at the N terminus of the protein (Fig. S4), which results in crystal formation taking many months (see Materials and methods for further details of crystallisation conditions). TarP_Rhp also co-purified and crystallised in complex with 4-hydroxyphenylacetate (4HPA) (Fig. 1B), which is a common by-product of aromatic amino acid fermentation in microorganisms [38] and can also be derived from the degradation of lignin [39]. TarP_Rhp binds 4HPA in solution with a K_d of ~ 0.4 nM (Fig. S2), which is ~ 20 times tighter than coumarate, providing an explanation as to why it remains intrinsically bound. Attempts to crystallise TarP_Rhp that had been gently denatured (to remove intrinsically bound ligands) and refolded were not successful. This may have been due to the presence of a di-sulphide bond between C117 and C235, which is visible in the final refined TarP_Rhp structure (Fig. S4). Despite not being able to determine structures of TarP_Rhp in complex with HCM ligands, the structure bound to 4HPA was determined to atomic resolution (1.1 Å), enabling us to carry out modelling studies with the protein coordinates. We also determined high-resolution co-crystal structures of both TarP_Sse and TarP_Csal in complex with coumarate, TarP_Sse with cinnamate and TarP_Csal with caffeate and ferulate; the latter structure is surprising given that fluorescence-based solution-binding assays did not detect binding of ferulate to TarP_Csal. The apo structure of TarP_Csal was also determined, enabling observation of the conformational change between an open (apo) and closed (ligand-bound) conformation of this subfamily of TRAP SBPs. Comparison of these structures (Fig. 1C–K), in the context of the binding assays, reveals a conserved mechanism of ligand binding and identifies differences within the binding pockets that facilitate ligand specificity.
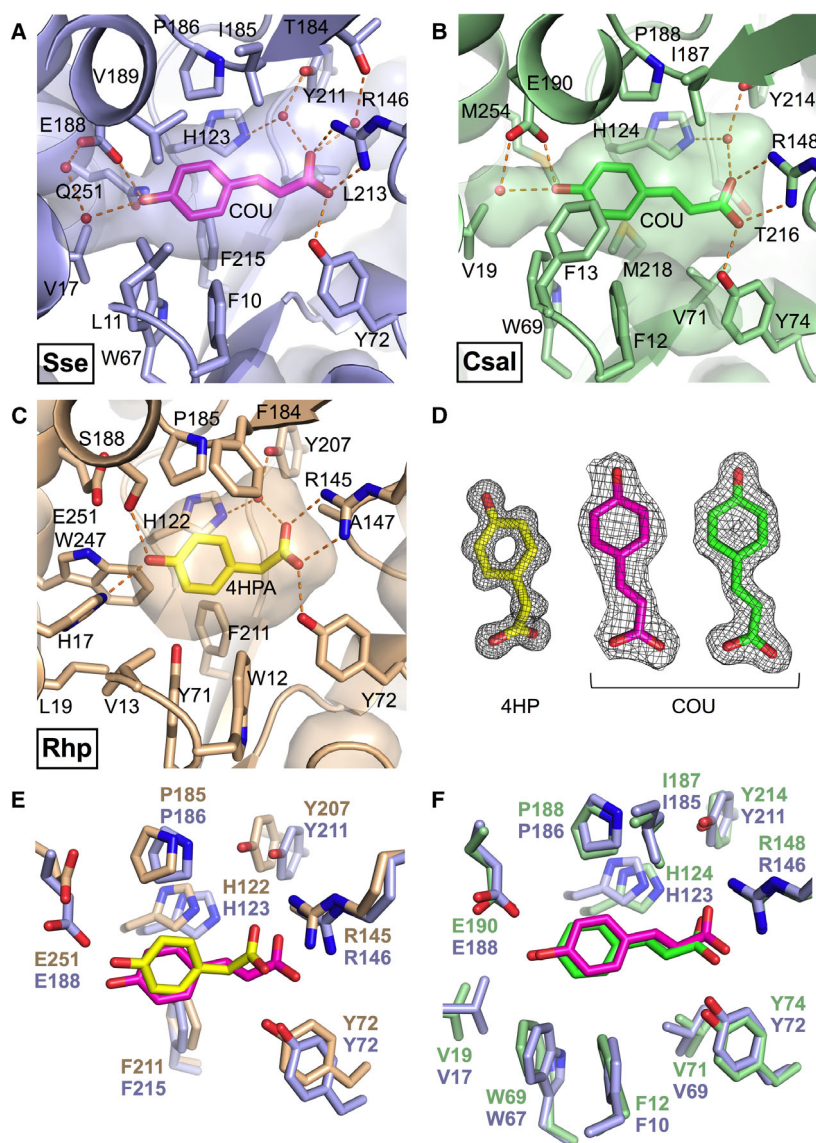
The arrangement of the binding pocket in the TarP family of SBPs

The overall fold of all three TarP family homologues is typical of DctP-type TRAP SBPs (InterPro: IPR038404), consisting of two α/β domains connected by hinge regions with a ligand-binding site buried at the domain interface. The organisation of the secondary structure elements in each of the three proteins is conserved, and the closed, ligand-bound conformation is highly similar (RMSD Cα: 1.2–1.8 Å) (Fig. 1C–K). Comparison of the TarP_Csal apo structure and the complex with coumarate shows that domain 1 (residues N-1–124, 150–155, 213–258, 298–324-C) and domain 2 (residues 129–148, 157–210, 262–293) are connected by a number of flexible regions (residues 125–128, 149–156, 211–212, 259–261, 294–297) (Fig. S5), enabling the protein to open and close by a 13° rotation of one domain against the other around an axis that lies between the two domains (Video S1 and S2, Fig. 1K). Notwithstanding that we may have crystallised a partially open conformation, this degree of domain rotation is much smaller than in the archetypal TRAP SBP, SiaP, where the domain rotation is ~ 28° between the open and closed forms (comparisons made between the apo structure, PDB ID: 2CEY and ligand-bound structure, PDB ID: 2XXK) [15]. This may be explained by both the larger size of the ligand (sialic acid) and the requirement to capture an extensive network of intrinsically bound water molecules along with the ligand in SiaP [40], which may require a more open conformation. In all three TarP proteins in this study, the binding pocket is largely lined with hydrophobic residues that generate packing interactions with the ligand, as well as conserved hydrogen bonding groups, which interact with the phenyl ring substituents at one end of the ligand and a carboxyl group at the other, conferring ligand specificity. The details of these interactions are shown in Figs 2A–C and 3A–C with the ligand fit to the density shown in Figs 2D and 3D. The features of the structures are further analysed in the following sections.

A conserved glutamate controls binding at the phenyl ring end of the ligand

In TarP_Sse and TarP_Csal, a sequentially and structurally conserved glutamate from domain two (E188/E190, respectively) is responsible for forming a hydrogen bond to one or both hydroxyl groups on the phenyl rings of coumarate, caffeate and ferulate, but not cinnamate, which lacks ring substituents (Figs 2A,B and 3A–C). With a coumarate ligand, there is an additional set of water-mediated hydrogen bonds between the glutamate and the phenyl ring -OH of the coumarate (involving two buried waters in TarP_Sse and one buried water with TarP_Csal) (Fig. 2A,B). In TarP_Rhp, this glutamate is not sequentially conserved (aligns to G187 in the TarP_Rhp sequence), and in the

Fig. 2. Coumarate binding in TarP SBPs and the comparison with 4HPA binding in Tarp_Rhp. (A–C) A detailed view of the residues surrounding the binding pocket and interactions made between the protein and ligand in the TarP_Rhp/4HPA complex (beige/yellow, A), the TarP_Sse/coumarate complex (blue/magenta, B) and the TarP_Csal/coumarate complex (green/green, C). Hydrogen bonds are drawn as orange dashed lines, and the inside surface of the binding pocket is shown as a partially transparent surface. Buried water molecules are shown as small red spheres. (D) Ligand density (black mesh, contoured at 1.3 Å) for each of the structures described in A–C. (E) A comparison of the positions of conserved residues surrounding the ligand-binding site in TarP_Rhp/4HPA (beige/yellow) and TarP_Sse/coumarate (blue/magenta). (F) A comparison of the positions of conserved residues surrounding the ligand-binding site in TarP_Sse/coumarate (blue/magenta) and TarP_Csal/coumarate (green/green). In (E) and (F), only conserved sidechains are shown and residue numbers are coloured according to each structure. All images in this figure were generated using PYMOL.

structure, a serine (S188) satisfies the demand for a hydrogen bond to the hydroxyl group of the 4HPA in the corresponding region of the binding pocket, along with an additional interaction to H17 (Fig. 2C). There is, however, a glutamate positioned nearby (E251) in TarP_Rhp, that is spatially equivalent to E188/E190 in the other two SBPs (Fig. 2E,F). Although we could not obtain co-crystal structures of TarP_Rhp with any tight-binding HCM ligands, based on their increased backbone length compared to 4HPA (Fig. 1B), and the spatial conservation of E251, it seems likely that this glutamate may facilitate binding of HCMs in TarP_Rhp. This suggests that for this subfamily of SBPs, a glutamate acting as a hydrogen bond acceptor (given the pH of the crystallisation experiments) is

important for binding the hydroxyl group(s) on the phenyl ring of the ligand. We note that the full hydrogen bonding potential of the hydroxyl group is not utilised by any of the proteins for which we have determined structures.

Understanding differences in cinnamate-binding affinity

Our assays show that although TarP_Sse and TarP_Csal both bind coumarate with similar affinity, TarP_Sse has 2-fold weaker binding for cinnamate and TarP_Csal shows no binding to cinnamate (Table 1). TarP_Rhp also has much lower binding affinity for cinnamate than coumarate (~ 5-fold weaker, see Table 1 and ref. [29]).
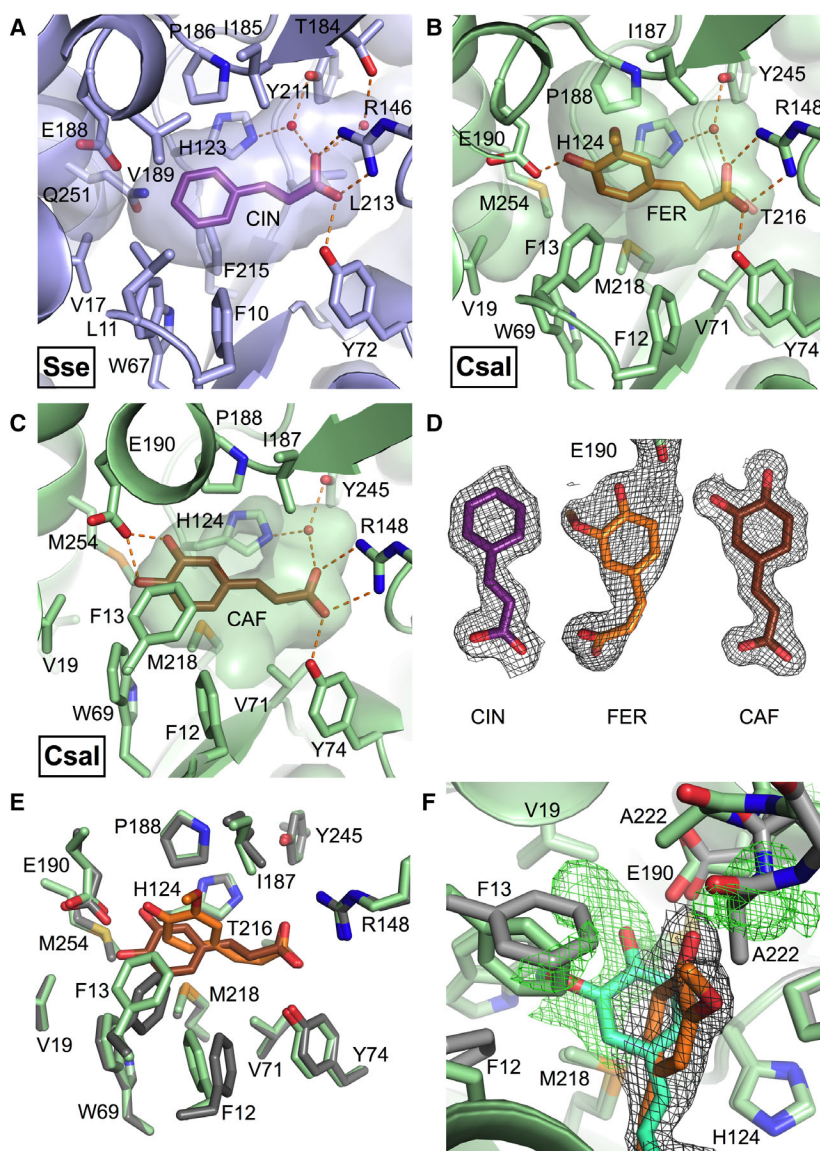
Fig. 3. Binding of cinnamate, caffeate and ferulate in TarP SBPs. (A–C) A detailed view of the residues surrounding the binding pocket and interactions made between the protein and ligand in: (A) the TarP_Sse/cinnamate complex (blue/purple), (B) the TarP_Csal/ferulate complex (green/orange) and (C) the TarP_Csal/caffeate complex (green/brown). Hydrogen bonds are drawn as orange dashed lines, and the inside surface of the binding pocket is shown as a partially transparent surface. Waters are shown as small red spheres. (D) Ligand density (black mesh, contoured at 1.3 σ) for each of the structures described in A–C. (E) A superposition of the sidechains surrounding the ligand-binding site in the TarP_Csal/ferulate (green/orange) and TarP_Csal/caffeate (grey/brown) structures. (F) After refinement of the TarP_Csal/ferulate structure, the difference map (green mesh, contoured at 3 σ) indicated that there are two binding modes of ferulate in the structure (the major binding mode is shown in orange, surrounded by the experimental map, contoured at 3 σ). The minor binding mode of ferulate (aqua) binds in a similar position to caffeate, cinnamate and coumarate from the other structures. This location is associated with a different position of F12, F13 and A222 (shown in grey from TarP_Csal/cafferate structure), evidence for which can also be seen in the difference map. All images in this figure were generated using PYMOL.

Together this suggests that the absence of the phenyl ring hydroxyl group in cinnamate has a detrimental effect on binding, which is not surprising given the clear demand for hydrogen bond formation between the ligand and the protein in this region of the binding pocket.

Comparison of the TarP_Sse/cinnamate and TarP_Sse/coumarate complexes shows that despite the loss of the hydrogen bond to E188, there is no dramatic change in the position of the ligand, the surrounding residues (sidechain movements < 0.4 Å) or the adjacent water structure. The volume of the binding pocket is approximately equivalent in both structures, with a slight increase in volume in the cinnamate complex compared to the coumarate complex (36 Å³ vs 34.9 Å³, respectively) (Table 2). This

perhaps reflects the loss of the hydrogen bond in the cinnamate complex, which may generate a slightly more open binding pocket. These subtle changes are consistent with a 2-fold change in binding affinity, with K_d values of 120 ± 4 nM and 247 ± 21 nM for TarP_Sse with coumarate and cinnamate, respectively (Table 1).

In TarP_Csal, the structural basis for the total lack of cinnamate binding is less clear because the hydrogen bonding groups around the ligand-binding pocket are conserved between TarP_Csal and TarP_Sse. However, the residues that provide packing interactions between the ligand and the walls of the binding pocket are different. A direct comparison between coumarate binding in TarP_Csal and TarP_Sse demonstrates that

Table 2. Solvent accessible volume of the binding pockets in each of the TarP structures (Å³). CouP in complex with ferulate and 4-hydroxy-phenyl pyruvate (PDB ID: 4JB0) is included for comparison. Ligand codes: 4HPA, 4-hydroxyphenylacetate; COU, coumarate; CIN, cinnamate; CAF, caffeate; FER, ferulate; ENO, 4-hydroxyphenylpyruvate and GOL, glycerol. Calculation was made with CASTp v3.0 using a 1.4 Å probe radius [62].

	TarP_Rhp	TarP_Sse	TarP_Csal	CouP
COU	–	34.9	36.8	–
CIN	–	36.0	–	–
CAF	–	–	39.2	–
FER	–	–	50.4	–
FER/ ENO	–	–	–	30.1
4HPA	39.6	–	–	–
GOL	–	–	–	27.8

although the volume of the pocket is similar (Table 2), the ligand position is slightly different (Fig. 4A,B). This is in part due to Q251 in TarP_Sse, the Nε atom of which is positioned 3.2 Å away from the oxygen atoms of the phenyl ring -OH group of the coumarate, forming a steric interaction, and 3 Å away from the nearest oxygen atom of E188, forming a long hydrogen bond (Fig. 2A). This forces the phenyl ring to sit higher in the pocket than it does in TarP_Csal, where this glutamine is replaced with a methionine (M254), which cannot form the same set of interactions with the ligand in this region of the binding pocket and

therefore occupies a subtly different space within the structure (compare Fig. 4A,B). This amino acid difference results in a ~ 1 Å change in the position of the phenyl ring -OH group, causing the hydrogen bonding network between the ligand -OH and E190/E188 to differ between the two proteins (Fig. 4A,B). In TarP_Csal, E190 forms two hydrogen bonds, one to the hydroxyl group of the phenyl ring (2.6 Å) and second to a water molecule (HOH1) (2.6 Å) that is also hydrogen-bonded to the phenyl ring -OH (2.9 Å), forming a ring of five bonded atoms (Fig. 4A). In TarP_Sse, the displacement of the ligand -OH group by Q251 draws E188 into a slightly different position outside of hydrogen bonding distance of HOH1. To fulfil the hydrogen bonding network, a second, solvent-exposed water molecule (HOH2) forms a bridging interaction from E188 to the phenyl ring -OH via HOH1 (distances: HOH2 → HOH1 2.9 Å and HOH1 → phenyl -OH 3.0 Å), forming a ring of 6 bonded atoms. HOH1 and HOH2 are held in these positions by mainchain interactions with the amino groups of N16 and V17, respectively (Fig. 4B). In the TarP_Sse/cinnamate complex, the position of HOH1 and HOH2 are conserved, satisfying the hydrogen bonding network around E188 despite the loss of the ligand -OH group (Fig. 4C). Our observed changes in the water networks involved in substrate binding by TarP_Csal and TarP_Sse, in which ligand binding depends on a more limited set of hydrogen bonding
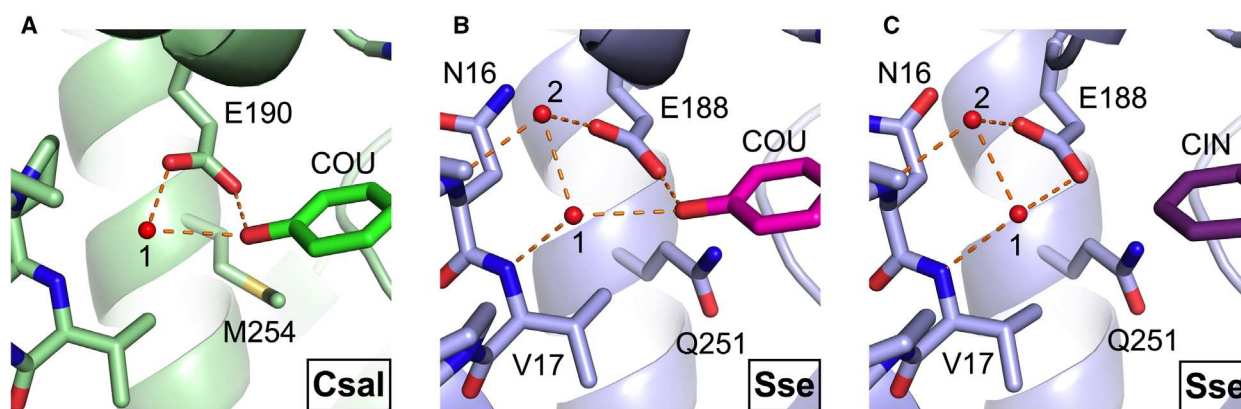

Fig. 4. Sequence-specific differences in water structure prohibit cinnamate binding in TarP_Csal. (A) In TarP_Csal (pale green), the coumarate (COU, green) phenyl ring -OH group forms hydrogen bonds (orange dashes) to E190 and a water molecule (HOH1, numbered 1), which is also hydrogen-bonded to E190, forming a ring of four bonded atoms. (B) In TarP_Sse (pale blue) the coumarate (magenta) is in a slightly different position due to the position of the Q251 sidechain; this changes the water structure around the phenyl ring -OH of the ligand. HOH1 is outside of hydrogen bonding distance to E188, thus a second water (HOH2, numbered 2) forms bridging interactions with the ligand, forming a ring of five bonded atoms. The two water positions are stabilised by hydrogen bonds to mainchain amino groups (N16 and V17). (C) In the TarP_Sse (pale blue) structure with cinnamate (CIN, purple) the water structure is maintained despite the loss of the phenyl ring -OH group on the cinnamate, with HOH1 and HOH2 providing stabilising interactions with E188. All images in this figure were generated using PYMOL.

interactions for TarP_Csal, offer a possible explanation for the lack of cinnamate binding by TarP_Csal. The absence of the hydroxyl group in cinnamate could be responsible for disrupting the water network with a concomitant decrease in binding affinity for the ligand.

In SiaP, achieving high-affinity ligand binding relies on a strict network of 14 buried water molecules within the binding pocket and disrupting this network by point mutations has been shown to severely reduce the binding affinity of sialic acid [40]. Our study shows that water molecules are also important for modulating ligand-binding affinity in the TarP SBPs, with 1–4 buried water molecules playing a key role in the hydrogen bonding networks surrounding the ligands in all of the proteins investigated here.

A conserved Y/R pair and a buried water molecule are responsible for binding the carboxyl end of the ligand

In all three TarP proteins, a strictly conserved Y/R pair (TarP_Rhp: Y72/R145, TarP_Sse: Y72/R146 and TarP_Csal: Y74/R148) forms a core component of a highly networked salt bridge between the arginine and the carboxyl group of the ligand (Figs 2A–C, 3A–C and 5A–D). The constellation of residues surrounding
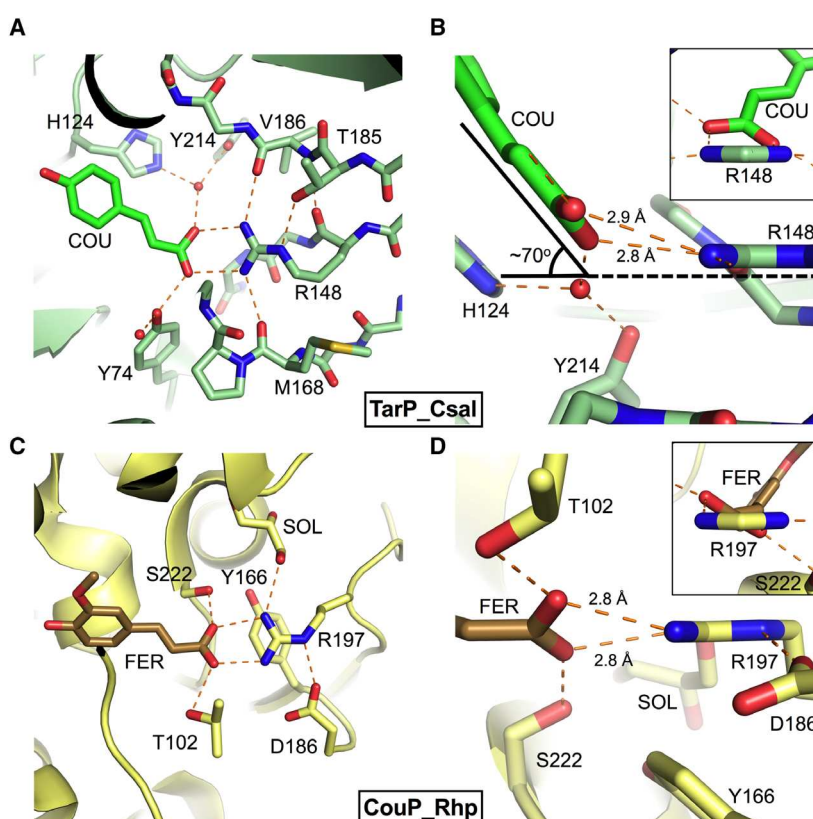

Fig. 5. The critical arginine forms a highly networked salt bridge with the ligand. (A) The highly networked salt bridge between R148 and the carboxyl group of the ligand in the TarP_Csal/coumarate (COU) complex (pale green/green) (TarP_Csal is used here as an example, the other TarP proteins have a similar arrangement around their equivalent invariant arginine). Hydrogen bonds are shown as orange dashed lines and water molecules are small red spheres. Only important side chains are shown. (B) The carbon backbone of the coumarate, including the carboxyl group, sits ∼ 70° outside of the plane of the guanidinium group of R148 (side view). The inset shows the twist of the carboxyl group relative to the plane of the guanidinium group (head-on view), which causes one of the pair of hydrogen bonds to be slightly longer than the other (2.9 Å vs 2.8 Å). (C) In the CouP_Rhp complex with ferulate (FER) (yellow/gold) (PDB ID: 4JB0) the arginine that forms a salt bridge with the ligand originates from a loop region and does not form the same kind of highly interconnected network of interactions as that seen in TarP_Csal. A molecule of glycerol (SOL) forms one of the hydrogen bonds with R197, but no other water molecules were modelled within hydrogen bonding distance. The mainchain of the helix bearing T102 has been hidden for clarity. (D) In CouP_Rhp, the carbon backbone of the ligand lies approximately in-plane with the R197 (side view); however, in the head-on view (inset) it is clear that the carboxyl group of the ligand is twisted ∼ 45° with respect to the guanidinium group of R197. Despite this, the pair of hydrogen bonds remain equidistant (both 2.8 Å). All images in this figure were generated using PYMOL.

the carboxyl group of the ligand is completed by a second conserved tyrosine and an additional histidine residue, which interact with the carboxyl via a spatially conserved buried water molecule (TarP_Rhp: H122/Y207, TarP_Sse: H123/Y211, TarP_Csal: H124/Y214). The number of buried water molecules in this region appears to be species-specific. TarP_Rhp and TarP_Sse both have one, whilst TarP_Csal has an additional buried water molecule 3.6 Å away from the first, which forms an extra hydrogen bond to the ligand carboxyl. Its position is stabilised by hydrogen bonds to T184 (which is a valine in TarP_Sse) and the mainchain (Figs 2B and 3A). As with the interactions around the phenyl ring end of the ligand and the example of SiaP, this suggests that the specific location of buried water molecules is important for ligand recognition in TRAP SBPs.

The arginine from the Y/R pair is sequentially and spatially conserved in all TRAP family proteins [16,41,42]. Positioned at the end of β6 within domain 2, it is anchored by several hydrogen bonds to the mainchain of two adjacent β-strands (β7 and β8); carbonyl and amide of V186 (TarP_CsaI), T184 (TarP_Sse) and T183 (TarP_Rhp) and the carbonyl of M168 (TarP_CsaI), M166 (TarP_Sse) and A165 (TarP_Rhp). In TarP_Csal and TarP_Sse, an additional hydrogen bond between the Nε atom of the arginine and a threonine sidechain (T183 and T185, respectively) further enforces this position. This results in the arginine essentially substituting for the C-terminal residue of β6 (G149, G148 and A147 in TarP_Csal, TarP_Sse and TarP_Rhp, respectively) by fulfilling the hydrogen bonds on the edge of the surrounding strands. This generates a split β-sheet, of which the arginine is a key structural component. The surrounding secondary structure restricts the rotational freedom of the arginine sidechain, forcing it to adopt an unusual rotamer in the TarP_Rhp and TarP_Sse structures. In the SBP from the HCM ABC transporter, CouP_Rhp (PDB ID: 4JB0) [29], an equivalent arginine (R197), is responsible for forming the same kind of interaction with the carboxyl group of similar ligands (Fig. 5C). However, the fold of CouP_Rhp is quite different to TarP_Rhp and the structural context around the arginine is also distinct. In the CouP_Rhp complex with ferulate, R197 packs against an underlying tyrosine (Y166), and apart from a single hydrogen bond to D168, the only other hydrogen bond, other than with the ligand, is to a molecule of glycerol that sits close by in the crystal structure. As this arginine is solvent-exposed and on a loop region, it seems feasible that it functions as a flexible cap to the binding pocket. This contrasts with the situation in the TarP

SBPs, where the arginine has an integral role in stabilising secondary structure elements within domain 2, is buried and is highly conformationally constrained.

In all our TarP structures, the arginine forms an end-on, bidentate twin-nitrogen, twin-oxygen interaction with the carboxyl group of the ligand. An in-plane interaction is favoured for intramolecular salt bridges within proteins, as this optimises the overlap of the guanidinium hydrogen atoms with the syn lone pairs of the carboxyl oxygen atoms. However, in many high-resolution crystal structures deviations up to 50° out of plane are common [43,44]. In all of the TarP_Csal and TarP_Sse complexes, the carbon backbone of the ligand (including the carbon of the carboxyl group) lies ∼ 70° outside of the plane of the guanidinium group (Fig. 5B), which is unusual when compared to these types of interactions within proteins in general and also, more specifically, within TRAP proteins [41]. The only structure in this study that is in-plane is the TarP_Rhp/4HPA complex, which is probably due to the smaller size and greater conformational freedom of the ligand, compared to the HCMs.

In CouP_Rhp, the carbon backbone of the ligand is more in-plane with the guanidinium group of the arginine, but here the two groups are twisted 45° in relation to each other (Fig. 5D). A suboptimal arrangement of the ligand carboxyl in relation to the critical arginine may indicate a shared mechanism of solute release, ensuring that binding of the ligand to the SBP is not favoured over release and subsequent transport. This may be even more important in the TRAP transporters where transport is independent of ATP hydrolysis. In SiaP, the carboxyl group of the sialic acid is in-plane with the critical arginine, but SiaP has a double arginine motif providing end-on and side-on interactions. The much bigger size of sialic acid compared to HCMs requires a bigger binding pocket and the ligand occupies a different region of the binding pocket to the HCMs. Moreover, binding of sialic acid relies on a network of several water molecules. Thus, even though the overall fold is the same in both of these TRAP SBPs, the interactions between the ligand and the protein are quite different, suggesting that SiaP may require a different mechanism of solute release.

Finally, we note that the electron density around the coumarate in both the TarP_Sse and TarP_Csal structures shows that the carbon backbone of the ligand is visibly bent, suggesting that there is an element of strain in the conformation of the ligand, which is reflected in the refinement, with plane distortion centred around the C3 atom of the ligand. Atom C3 is −0.28 and 0.15 Å away from a least squared plane

defined by the 6 neighbouring atoms of the ligand in TarP_Sse and TarP_Csal, respectively (Fig. S6). This could also indicate one mechanism of solute release in TRAP transporters.

Domain closure engages the critical arginine with the carboxyl group of the ligand

A comparison of the open apo and closed coumarate-bound structures of TarP_Csal shows that the region surrounding R148 is the main moving component during domain closure. In the open structure, the electron density around R148 is much weaker than the map of the surrounding protein structure; therefore, we modelled the sidechain guided by the strongest map peaks. This places the Nε and NH2 of R148 within hydrogen bonding distance of the carbonyl of M168 (2.8 and 2.7 Å, respectively). The mainchain amide and carbonyl of R148 also form long hydrogen bonds with the mainchain carbonyl of G184 (3.0 Å) and the mainchain amide of V186 (3.1 Å), respectively (Fig. 6). The R148 sidechain is essentially pinned to β7 (via M168), sitting much further away from β8 (~ 8 Å) than it does in the closed complex. The refined sidechain atoms of R148 had 2-fold higher B factors than the surrounding protein, suggesting that it is flexible. Consequently, the water structure surrounding R148 is disordered, such that many of the exposed hydrogen bonding groups along the edges of β7 and β8 do not have corresponding solvent molecules. This contrasts with the many ordered solvent molecules in and around the binding pocket in the apo structure. All but four of these water molecules are expelled to bulk solvent during domain closure of TarP_Csal (Fig. 3B), providing an entropic driver for ligand binding. Domain closure forces the sidechain of R148 closer to β8, which requires a change in the R148 rotamer, flipping the direction of the Nε by 180° so that the hydrogen atom points towards β8 and picks up an interaction with T185. Of the four hydrogen bonds formed by R148 in the open structure, two are retained during the conformational change to the closed complex (R148 NH2-M168 carbonyl and R148 carbonyl-V186 mainchain amide), one is lost and one is replaced, whilst four others are made. As well as the expulsion of bulk solvent, this net gain in hydrogen bonding groups, forming part of an extended network of interaction around the carboxyl group of the ligand, represents a key enthalpic driver for binding affinity in TarP SBPs. Video S1 and S2 illustrate the major changes in domain movement and bonding described in this section.
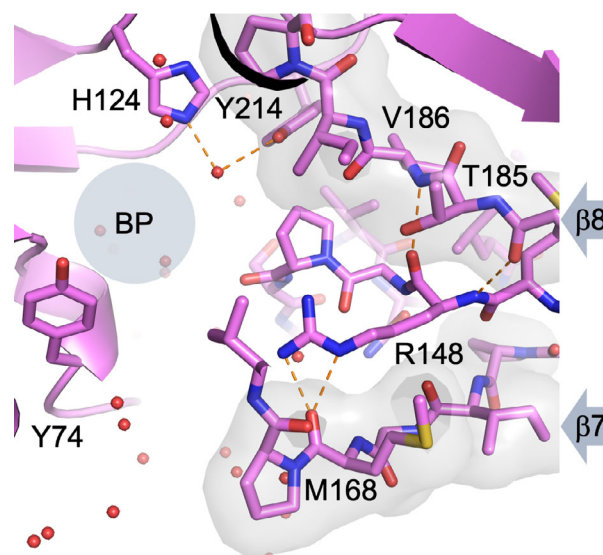

Fig. 6. Ligand binding is driven by the conformational change in the region surrounding the critical arginine (see also Video S1 and S2). In the open conformation of apo-TarP_Csal (pink), the R148 sidechain is pinned against β7 via two hydrogen bonds (orange dashed lines) to the mainchain carbonyl of M168, whilst its mainchain amide and carbonyl form hydrogen bonds to mainchain groups on β8 (T185 carbonyl and V186 amide). The direction of β7 and β8 are indicated by grey arrows and the strands are drawn with a transparent molecular surface (grey) to show their proximity to R148. The approximate location of the binding pocket is indicated by the grey circle (labelled BP) and the residues that mediate binding of the carboxyl group of the ligand in the closed complex (Y74, H124, Y214) are shown for context. All the water molecules in this region of the structure are shown as small red spheres. Interestingly, the buried water molecule that forms part of the network of interactions around the ligand in the closed structure (hydrogen-bonded between H124 and Y214) is prebound in the open structure. All images in this figure were generated using PYMOL.

Ferulate has multiple modes of binding in TarP_Csal

In our binding assays, we could not detect any binding of TarP_Csal to ferulate (Table 1); however, with the high concentrations of ligand in the crystallisation experiment, we were able to obtain a co-crystal structure of this complex. There is evidence from the electron density map that ferulate has two modes of binding in the structure, with the conformation of the binding pocket adjusting to accommodate either (Fig. 3E,F). The minor mode of binding (not modelled in the deposited structure) is very similar to that of caffeate, including evidence from the difference map that the surrounding protein also has a second conformation that closely matches its conformation in the

caffeate complex. The major binding mode, representing ~ 80% of the protein complexes in the crystal (based on the relative size of the map peaks), has the carboxyl group in approximately the same position as caffeate, but the phenyl ring of the ligand is rotated by ~ 160° and is packed against the opposite face of the binding pocket, under A222. This position requires the helix bearing A222 to move by ~ 2.5 Å (Cα-Cα) compared to its location in the complex with caffeate. On the other side of the binding pocket, F12 and F13 adopt different rotamers to fill the space that is created by the repositioning of the ligand within the binding pocket of the protein. Analysis of the volume of the binding pocket in the TarP_Csal/ferulate complex shows that it expands as a result of these conformational changes (50.4 Å^3 vs 39.2 Å^3 with caffeate) (Table 2). Presumably, the lack of binding of ferulate to TarP_Csal in our assays is due to ineffective closure of the protein around the ligand, which is in part due to the arrangement of hydrophobic groups around the binding pocket in TarP_Csal, which are incompatible with the position of the C3-methoxy group of the ligand. TarP_Sse has weak, but detectable binding for ferulate (K_d 486 ± 72 nm) and a comparison of the TarP_Csal and TarP_Sse structures shows that F13 is replaced with a leucine (L11) in TarP_Sse (Fig. 2B,C), generating space adjacent to where the ferulate C3-methoxy group would sit in the optimal binding mode. Thus, it seems feasible that TarP_Sse could achieve a fully closed complex, explaining why binding of ferulate could be detected in this species variant.

Modelling coumarate binding in TarP_Rhp

Although HCM ligands bind tightly to TarP_Rhp ([29]; Table 1), we could not obtain crystals of these complexes (see Materials and methods section for further detail). We therefore used the structures of our protein/ligand complexes from the other two species variants to model coumarate binding in TarP_Rhp (K_d 8 ± 5 nm). Using the protein structure from the TarP_Rhp/4HPA complex and without moving any of the sidechains, we manually docked a molecule of coumarate, using the conformation of the ligand from the TarP_Sse/coumarate complex. By optimising the hydrogen bond distances between the ligand and the protein at both ends of the binding pocket, we were able to obtain a reasonable fit (Fig. 7A). The only problem area was a clash between the phenyl ring of the coumarate and the indole ring of W247. With subtle changes in sidechain positions, this clash could be alleviated, so to remove bias from our interpretation, we performed computational modelling with flexible fitting. Water molecules were removed from the model, and rotational freedom was allowed for all of the sidechains surrounding the binding pocket and for the conformationally free bonds of the ligand. The top result places coumarate in the expected position, with

Fig. 7. Modelling coumarate binding in TarP_Rhp. (A) Coumarate (cyan) can be modelled into the TarP_Rhp protein structure (beige) by manual fitting without moving any of the sidechains or water molecules. Reasonable hydrogen bonding distances could be maintained, with the only clash arising between the phenyl ring of the coumarate and the indole ring of W251. (B) Computational modelling was carried out on TarP_Rhp (grey) with flexible fitting. The top result placed coumarate (blue) in the expected position, with the S192 side chain rotated by 120° to generate the space enabling E255, W251 and H21 to hydrogen bond with the phenyl ring -OH group. The buried water molecule that sits between H126 and Y211 (small black sphere) could be directly modelled back into the structure whilst maintaining reasonable hydrogen bonding distances. Hydrogen bonds are drawn as orange dashed lines and distances are shown in Å. The molecular surface of coumarate is drawn with a grey transparent surface in both panels. All images in this figure were generated using PYMOL.
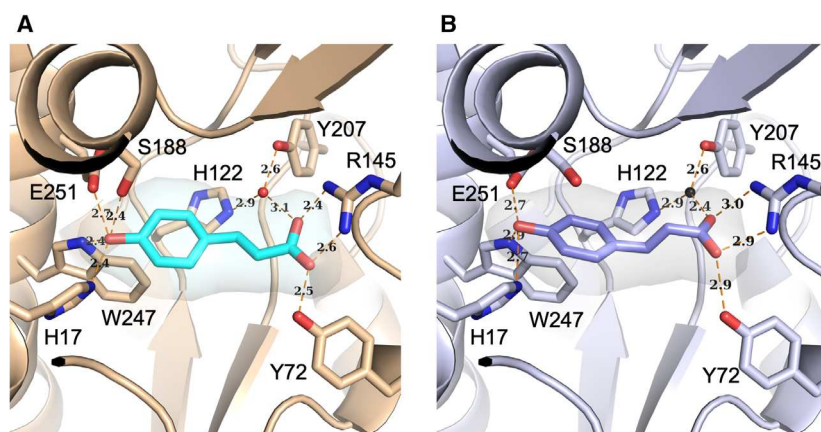
a slight conformational change in the ligand backbone and a subtle rotation of the pose of the ligand in the binding pocket, compared to the manually docked model. The only major difference in sidechain rotamer is S188, which rotates by 120°, creating space for E251, W247 and H17 to form hydrogen bonds with the phenyl ring -OH of the ligand (Fig. 7B). This result agrees with our initial observation regarding the demand for a glutamate hydrogen bond acceptor to interact with the phenyl -OH group of the HCM ligand in this subfamily of TRAP SBPs. Very little movement in any of the other residues is required to bind coumarate, and we could directly model back in the buried water molecule that sits between Y207 and H122 from the TarP_Rhp/4HPA structure whilst maintaining reasonable hydrogen bonding distances. We therefore conclude that the structure of TarP_Rhp is compatible with binding coumarate and other HCMs, notwithstanding our inability to generate crystals of these complexes.

Comparison of the binding pockets of TarP family SBPs with those of other TRAP SBPs

To put our findings into a wider context, we drew upon a high-throughput study of ligand specificity in TRAP SBPs that determined structures of a range of TRAP SBPs in complex with their endogenously bound ligands [16]. Some of the ligands that co-crystallised in these structures have been confirmed to be the natural cargo of the SBP in question, whilst others have yet to be fully characterised. Three proteins from the Vetting *et al.* study [16] co-crystallise with aromatic/HCM-like molecules. These are SBPs from *Bordetella bronchiseptica* strain RB50 bound to mandelate (PDB ID: 4P56, Uniprot: Q7WJQ1), *Ru. pomeroyi* strain DSS-3 bound to a number of hydroxybenzoate derivatives (PDB ID: 4PAI, 4PAF and 4PBH Uniprot: Q5LSJ5) and *Polaromonas* sp. Strain JS666 (PDB ID: 4MNC Uniprot: Q122C7) bound to benzoyl formate compounds. All three species (or their close relatives) have aromatic degradation pathways [45,46].

A comparison of these three structures to our TarP proteins (Fig. S7) reveals some conserved structural motifs and some marked differences. The overall folds are clearly related to the TarP SBPs (all atom RMSDs; PDB ID: 4P56, 1.6 Å, PDB ID: 4PAI, 6.0 Å and PDB ID: 4MNC, 6.5 Å, compared to TarP_Csal). Of the three examples, the hydroxybenzoate-bound SBP from *Ru. pomeroyi* has the lowest sequence conservation around the binding pocket. Like SiaP, it has a double arginine motif at the carboxyl end of the ligand and

the rotamer of the critical arginine is different to the TarP SBP family. The hydrogen bonding groups around the hydroxy-benzene are similar to those in the TarP SBPs, but there are many unfulfilled hydrogen bonds in this region. The binding pocket is also very spacious (73.8 Å^3) given the small size of the ligand and contains a number of water molecules that are connected to external bulk solvent, suggesting that the pocket is not fully closed. Together, this implies that in the crystal structure the ligand may be bound suboptimally and therefore may not be the true cargo of this SBP.

Like the SBP from *Ru. pomeroyi*, the benzoyl formate-bound SBP from *Polaromonas* sp. Strain JS666 differs significantly from the TarP SBP-subfamily, both in terms of the relative organisation of the secondary structure elements within its fold and the sequence conservation around the end of the binding pocket where the benzene ring is bound. There are no conserved acidic hydrogen bonding groups around the benzene ring of the ligand and this region of the binding pocket is largely hydrophobic, dominated by a tyrosine residue (Y44) that blocks the end of the pocket. Three water molecules accompany the ligand in the pocket, making mainchain interactions with the protein and packing against the benzene ring of the ligand. The critical arginine is held in a highly networked salt bridge with the additional threonine on the adjacent beta strand, as in TarP.

Finally, the mandelate-bound SBP from *B. bronchiseptica* RB50 is the most structurally similar to the TarP SBPs. The position of the arginine and its rotamer are exactly the same as in the TarP SBPs (Fig. 2A–C,E–F), including its involvement in a highly networked salt bridge. There is also high sequence conservation around the rest of the binding pocket, including the strictly conserved buried water that sits between the Y/H pair (Y234/H146) to one side of the ligand carboxyl group. The other waters in the pocket are not conserved, and they are linked to bulk solvent via a channel adjacent to the ligand. This is due to subtle packing differences between the loop at the N-terminal end of β6 and the neighbouring secondary structure in the mandelate-binding protein that creates more space, whereas in TarP_Csal, for example, the sidechains of D86, P150 and T216 fill this channel. The glutamate that binds the -OH group of the ligand phenyl ring in the TarP SBPs in spatially conserved in the *B. bronchiseptica* RB50 protein (E212), but because it does not engage with the mandelate it occupies a different position; therefore, there are a number of unsatisfied hydrogen bonds to the surrounding residues. A tryptophan residue (W274) dominates the

space in this region of the binding pocket applying a selection for the smaller mandelate ligand, over, for example, larger HCMs. However, with a subtle rotation of the sidechain, the binding pocket would closely resemble that of a TarP SBP. In support of this, orthologues of the *B. bronchiseptica* RB50 mandelate-binding SBP group very closely with the TarP SBPs as shown in the phylogenetic tree in Fig. S1B.

Conclusion

We have demonstrated the basis of ligand specificity in a group of TRAP transporter SBPs that bind and transport HCM ligands formed as part of lignin degradation in marine environments. Like other known TRAP transporter family members, the proteins in our study rely on a critical interaction between the carboxyl group of the ligand and a conserved arginine residue on the protein that is conformationally restricted within the structure. This differs from SBPs from ABC transporters, where an arginine with a similar role forms a flexible latch to close the binding pocket. Along with variations in the surrounding network of hydrogen bonding groups and buried water molecules, this interaction confers ligand specificity to this family of SBPs. We have further shown that the position of spatially conserved water molecules within the binding pocket can subtly modulate ligand-binding specificity, even between very structurally similar ligands. This supports evidence from studies with SiaP that show the position of water molecules is important and represents a shared mechanism for ligand specificity. In addition, analysis of our protein/ligand complexes and comparison with other TRAP transporter SBPs suggests that there is a level of strain on the bound conformation of the ligand (Fig. S6), related both to the structurally constrained binding pocket and relative orientation of the critical arginine to the carboxyl group of the ligand. This may represent a mechanism for ligand release in these ATP-independent transport systems, but further work is required to test this directly. Overall, a deeper understanding of how proteobacteria use TRAP transporters to acquire HCMs may help in exploiting such transporters in the future production of commercially valuable chemicals from lignin-derived aromatic feedstocks.

Materials and methods

Phylogenetic analysis and identification of TarP homologues

Amino acid sequences for 119 TRAP transporter SBPs were extracted from a subset of marine bacteria, namely *C. salexigens* DSM 3043, *A. manganoxydans* (formerly *Aurantimonas* sp. str. SI85-9A1), *Ru. pomeroyi* DSS-3 (formerly *Silicibacter pomeroyi* DSS-3), *Jannaschia* sp. CCS1, *S. stellata* E-37 and *Labrenzia aggregate* IAM 12614 (see Table S1 for a list of genomes and geneID abbreviations). The sequences were aligned in JALVIEW [47] using a MUSCLE alignment [48] (Fig. S1). Phylogenetic trees were also produced in JALVIEW.

Construction of overexpression plasmids

The coding sequences of *csal_0280* (*C. salexigens*) and *sse37_24379* (*S. stellata*) were retrieved from the NCBI database and codon-optimised gene constructs (JCAT web tool [49]), minus their predicted N-terminal leader sequence (SignalP [50]), were designed. The synthetic genes were synthesised (GBlocks; Integrated DNA Technologies, Leuven, Belgium) and assembled into pET21b(+) vector via Gibson assembly [51], resulting in the pET*csal_0280* and pET*sse37_24379* overexpression vectors. The gene construct for TarP_Rbp production was created by amplifying the nucleotide sequence encoding the *rpa1782* (*tarP*) gene minus the predicted N-terminal signal sequence (SignalP; 50) (951 bp) from *R. palustris* (strain CGA009) genomic DNA using primers pET*TarP*_F (5′-AATGTA<u>CCATGG</u>ATCAGGACAAAACTGTCAACTGG-3′; *Nco*I site underlined) and pET*TarP*_R (5′-AACTAT<u>CTCGAG</u>CAGCCCCGCGTCGTACTT-3′; *Xho*I site underlined). The amplified *rpa1782* gene was cloned into the *Nco*I and *Xho*I sites of pET22b(+), so it was in frame with sequence encoding the PelB leader peptide leader for periplasmic folding of recombinant C-terminally hexa-histidine tagged protein.

Production and purification of recombinant protein

Escherichia coli BL21 (DE3) transformed with either pET*csal_0280* or pET*sse37_24379* was grown at 37 °C to an optical density at 600 nm (OD_{600}) of 0.6 in LB medium containing carbenicillin (50 µg·mL^{-1}) (Melford Laboratories, Ipswich, UK). Overexpression of genes encoding recombinant TarP_Sse or TarP_Csal was induced by the addition of 0.4 mM isopropyl-b-D-thiogalactopyranoside (IPTG), and cells were incubated at 25 °C at 250 r.p.m. for a further 5 or 3 h, respectively. TarP_Rhp was produced using the same method, but cells were grown postinduction for 2 h at 37 °C. Selenomethionine (SeMet) protein was produced using the same method as for the native proteins, but IPTG-induced overexpression was carried out in M9 minimal media supplemented with 40 mg·L^{-1} L-selenomethionine [52]. Cells were harvested by centrifugation (10 000 *g*, 10 min, 4 °C) and resuspended in 20 mM sodium phosphate buffer pH 7.4, 500 mM sodium chloride and 20 mM imidazole, before lysis by sonication (MSE Soniprep, MSE, London, UK; 4 × 20 s bursts). Cell debris was removed by centrifugation (15 000 *g*,

25 min, 4 °C), and the cell-free extract was applied to a Hi-trap HP Nickel affinity column (GE Healthcare, Amersham, UK). Native and SeMet proteins were eluted over a 20–500 mM imidazole gradient in 20 mM sodium phosphate buffer pH 7.4 containing 500 mM sodium chloride. To remove any endogenously bound ligands that may have co-purified with TarP_Sse and TarP_Csal, purified protein was urea-treated and dialysed, as described in Salmon *et al.* [29]. TarP_Rhp was not urea-treated due to the presence of a di-sulphide bond in the structure, which caused issues with cor-rect re-folding. Proteins were concentrated and buffer exchanged into 50 mM Tris/HCl pH 7.2, 100 mM NaCl prior to structural studies.

Tryptophan fluorescence spectroscopy

Changes in the UV fluorescence of intrinsic tryptophan residues in recombinant TarP_Sse and TarP_Csal were measured on a Cary eclipse fluorimeter (Agilent Ltd, Stock-port, UK) in 10 mM Tris/HCl pH 7.4 at 30 °C in a 3 mL stirred quartz cuvette. For emission scan experiments, sam-ples were excited at 280 nm (5 nm slit width) and emission was recorded at 300–400 nm (20 nm slit width). Ligand titrations were performed with 0.2 μM recombinant protein (unfolded, dialysed and refolded for TarP_Sse and TarP_C-sal) in 10 mM Tris/HCl pH 7.4 buffer at 30 °C with excita-tion at 280 nm and emission at 340 nm using 5 nm excitation and 20 nm emission slit widths, respectively.

Protein crystallisation

SeMet TarP_Sse was concentrated to 10 mg·mL^{-1} prior to the addition of 6 mM coumarate. Automated crystallisation screens were carried out with a Hydra II crystallisation robot using commercial screens (Nextal, Molecular Dimen-sions, Sheffield, UK) (290 K). This identified several high-salt conditions that were subsequently optimised using hanging-drop vapour diffusion with a 1 : 1 ratio of protein to mother liquor. This resulted in the production of large cuboidal crystals that grew over a few days in conditions containing 0.2 M NaCl, 0.1 M sodium cacodylate pH 6.25 and 1.8 M ammonium sulphate. Native TarP_Sse was crys-tallised in complex with cinnamate (6 mM) from 1.6 M tri-sodium citrate, yielding crystals in the same crystal form as those grown from ammonium sulphate. Crystallisations of TarP_Sse with coumarate (6 mM) in conditions containing 0.2 M CaCl$_2$, 0.1 M Tris/HCl pH 7.5 and 22.5% (w/v) PEG 6000 produced a second crystal form, which diffracted to higher resolution.

SeMet TarP_Csal was concentrated to 11 mg·mL^{-1} prior to the addition of 6 mM caffeate. Automated crystallisation screens were carried out in the same way as for TarP_Sse, which identified several PEG 6000 based conditions that were subsequently optimised. This identified crystallisation conditions containing 0.2 M MgCl$_2$, 0.1 M MES pH 6.0 and

20% (w/v) PEG 6000. Native apo-protein crystals and native co-crystals with coumarate were both produced in conditions containing 0.2 M MgCl$_2$, 0.1 M HEPES pH 7.0 and 25% (w/v) PEG 6000. The TarP_Csal complex with ferulate crystallised in 0.2 M MgCl$_2$, 0.1 M Tris pH 8.5 and 20% (w/v) PEG 8000.

Native TarP_Rhp (7 mg·mL^{-1}) was screened in the pres-ence of 6 mM coumarate by sitting drop vapour diffusion. Thin, plate-like crystals grew from a drop containing 0.1 M Tris/HCl pH 8 and 20% (w/v) PEG 6000. However, unlike the other two TRAP proteins, where crystals grew in days, crystals of TarP_Rhp grew over a number of months and were not readily reproducible in either commercial screens or by manual optimisations.

Data collection, structural determination and analysis

All crystals were cryoprotected in their mother liquor plus 25% (w/v) ethylene glycol (PEG-based conditions) or glyc-erol (ammonium sulphate conditions) and then subse-quently mounted on a liquid nitrogen cold stream (100 K) prior to data collection. All datasets were collected on the MX beamlines at the Diamond Light Source (Table 3). For TarP_Sse, Seleno-MAD data were collected from a single crystal of the protein in complex with coumarate at two wavelengths (12 663 and 12 659 eV). Data were processed using Xia2 [53], which determined that the crystal belonged to the spacegroup P2$_1$ with cell dimensions of $a \approx 83$ Å $b \approx 88$ Å $c \approx 97$ Å and angles of $\alpha = \gamma = 90°$ $\beta \approx 92.3°$. SHELXCDE [54] was used to determine a selenium substruc-ture, from which preliminary phases were determined and an initial model was built. Forty-one selenium sites (~ 10 per subunit in the AU) were found for TarP_Sse, from which ~ 70% of the four subunits within the asymmetric unit were built automatically by SHELXE. Model building was completed using PHENIX PHASE and BUILD [55] before ligand density was interpreted in COOT [56] using ligand coordinates generated in JLIGAND [57]. The model coordi-nates were refined in REFMAC5 [58]. The TarP_Sse/cinna-mate complex belonged to the same spacegroup as the SeMet structure and was determined to 1.9 Å resolution by molecular replacement with PHASER [59] within CCP4I [60] using the protein coordinates of the SeMet/coumarate com-plex as a search model.

For TarP_Csal, Seleno-MAD data from a single crystal grown in the presence of caffeate were collected at three wavelengths (12 748, 12 672 and 12 700 eV). The data were processed by Xia2 in spacegroup P2$_1$2$_1$2 with cell dimensions of $a = 81.91$ Å $b = 119.59$ Å $c = 61.96$ Å and angles of $\alpha = \beta = \gamma = 90°$. The structure was determined to 1.67 Å resolution by Seleno-MAD using SHELXCDE [54], which auto-built a preliminary poly-Ala backbone from 18 selenium sites contained within two molecules in the asymmetric unit. Model building, refinement and interpretation of the ligand

Table 3. X-ray crystallography data collection and refinement statistics.

	TarP_Rhp + 4HPA (PDB ID: 7NQG)	TarP_Sse + coumarate (SeMet) PEAK (PDB ID: 7NR2)	TarP_Sse + coumarate (SeMet) INFLECTION	TarP_Sse + cinnamate (PDB ID: 7NRA)	TarP_Csal + caffeate (SeMet) PEAK (PDB ID: 7NRR)	TarP_Csal + caffeate (SeMet) INFLECTION	TarP_Csal + caffeate (SeMet) HREM	TarP_Csal + coumarate (PDB ID: 7NSW)	TarP-Csal + ferulate (PDB ID: 7NTD)	TarP_Csal Apo (PDB ID: 7NTE)
Data collection										
Beamline	DLS, i03	DLS, i02	DLS, i02	DLS, i03	DLS, i04	DLS, i04	DLS, i04	DLS, i04-1	DLS, i04	DLS, i03
Wavelength (Å)	0.97625	0.97918	0.97938	0.9500	0.97258	0.97841	0.97625	0.91731	0.97949	0.9500
Resolution range (Å)[c]	50.50–1.1 (1.13–1.1)	50.50–2.13 (2.19–2.13)	50.62–2.21 (2.27–2.21)	64.02–1.91 (1.96–1.91)	67.47–1.67 (1.71–1.67)	67.48–1.72 (1.76–1.72)	67.58–1.7 (1.74–1.7)	34.25–1.67 (1.71–1.67)	48.89–1.75 (1.80–1.75)	53.43–1.60 (1.64–1.60)
Space group	$P2_12_12_1$	$P2_1$	$P2_1$	$P2_1$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P1$	$P2_1$
Unit cell (a, b, c, α, β, γ)	38.67, 50.5, 142.56, 90, 90, 90	82.84, 88.08, 96.58, 90, 92.28, 90	83.0, 88.32, 96.78, 90, 92.23, 90	64.12, 70.2, 85.01, 90, 93.12, 90	81.73, 119.54, 61.94, 90, 90, 90	81.73, 119.59, 61.96, 90, 90, 90	81.91, 119.59, 61.96, 90, 90, 90	82.17, 119.94, 62.04, 90, 90, 90	49.86, 52.28, 68.53, 101.68, 98.59, 104.9	63.77, 63.62, 73.70, 90, 103.39, 90
Total reflections[c]	719 653 (23 572)	262 774 (18 734)	236 061 (12 881)	183 739 (14 119)	751 869 (19 828)	722 153 (23 772)	750 666 (22 368)	505 837 (31 641)	123 467 (9132)	261 785 (18 665)
Unique reflections[c]	113 835 (7978)	77 509 (5746)	69 735 (5149)	57 762 (4328)	70 139 (4567)	64 890 (4569)	67 210 (4644)	71 530 (5218)	62 516 (4572)	74 222 (5395)
Multiplicity[c]	6.3 (3.0)	3.4 (3.3)	3.4 (3.5)	3.2 (3.3)	10.7 (4.3)	11.1 (5.2)	11.2 (4.8)	7.1 (6.1)	2.0 (2.0)	3.5 (3.5)
Completeness (%)[c]	99.6 (96.1)	99.7 (99.8)	99.6 (99.8)	98.6 (99.8)	98.6 (88.8)	99.4 (96.2)	99.2 (93.9)	99.6 (99.4)	96.8 (95.4)	98.2 (96.8)
Mean $I/\sigma(I)$[c]	17 (2.2)	12.5 (2.3)	11.7 (2.0)	9.9 (2.1)	18.4 (2.0)	17.3 (2.2)	20.2 (2.1)	9.8 (1.4)	6.2 (1.2)	13.0 (2.1)
Wilson B factor	7.1	28.1	31.1	15.7	16.6	18.0	15.5	21.0	23.4	15.9
R_{merge}[a]	0.050 (0.436)	0.056 (0.432)	0.061 (0.528)	0.092 (0.559)	0.079 (0.568)	0.085 (0.584)	0.077 (0.599)	0.114 (1.145)	0.069 (0.54)	0.051 (0.530)
R_{pim}[b]	0.023 (0.328)	0.062 (0.343)	0.058 (0.406)	0.079 (0.410)	0.029 (0.346)	0.033 (0.314)	0.028 (0.334)	0.049 (0.560)	0.069 (0.54)	0.038 (0.379)
Anomalous slope		1.54	1.372		1.385	1.466	1.427			
Anomalous multiplicity[c]		1.6 (1.7)	1.6 (1.7)		5.5 (2.3)	5.8 (2.7)	5.8 (2.5)			
Anomalous completeness (%)[c]		91.2 (88.8)	90.6 (93)		98 (84.8)	99.2 (94)	98.7 (89.6)			
Refinement										
R_{factor}	0.118	0.18		0.20	0.206			0.197	0.191	0.176
R_{free}	0.138	0.236		0.261	0.265			0.251	0.252	0.217
RMSD (bonds)	0.0129	0.0103		0.0073	0.0072			0.0103	0.0071	0.011
RMSD (angles)	1.87	1.79		1.50	1.50			1.64	1.46	1.44
No. of non-H atoms										
Protein	2533	9918		5015	5102			5105	5006	5088
Ligands	32	95		22	29			27	38	3
Water	373	425		277	245			236	472	274
Average B factors										

Table 3. (Continued).

	TarP_Rhp + 4HPA (PDB ID: 7NQG)	TarP_Sse + coumarate (SeMet) PEAK (PDB ID: 7NR2)	TarP_Sse + coumarate (SeMet) INFLECTION	TarP_Sse + cinamate (PDB ID: 7NRA)	TarP_Csal + caffeate (SeMet) PEAK (PDB ID: 7NRR)	TarP_Csal + caffeate (SeMet) INFLECTION	TarP_Csal + caffeate (SeMet) HREM	TarP_Csal + coumarate (PDB ID: 7NSW)	TarP-Csal + ferulate (PDB ID: 7NTD)	TarP_Csal Apo (PDB ID: 7NTE)
Mainchain	9.3	35.9		30.9	21.2			23.0	30.3	19.0
Side chains	12.4	40.2		36.6	30.2			31.6	40.9	27.8
Ligands	12.4	41.6		30.9	26.8			23.7	62.4	23.1
Water	25.5	36.5		27.7	25.1			27.0	42.5	26.9
Ramachandran favoured (%)	99.68	98.15		97.24	98.16			98.31	98.59	98.62
Molprobity score	0.9	1.67		1.37	1.16			1.06	1.2	1.0

$^a R_{merge} = \Sigma_{hkl}\Sigma_i |I_i - I_m|/\Sigma_{hkl}\Sigma_i I_i$; $^b R_{pim} = \Sigma_{hkl} \sqrt{1/n - 1}\Sigma_{i=1}|I_i - I_m|/\Sigma_{hkl}\Sigma_i I_i$, where I_i and I_m are the observed intensity and mean intensity of related reflections, respectively.; ^{c}Values in parenthesis are for data in the high-resolution shell.

density was carried out for TarP_Sse. Crystals of TarP_Csal in complex with coumarate grew in the same crystal form, resulting in a 1.67 Å structure determined by molecular replacement. The complex with ferulate crystallised in space-group P1 with cell dimensions of $a = 49.86$, $b = 52.28$, $c = 68.53$, $\alpha = 101.68°$, $\beta = 98.59°$ and $\gamma = 104.9°$. A 1.75 Å structure was determined by molecular replacement. The apo structure of TarP_Csal was crystallised in a second crystal form, resulting in a 1.6 Å resolution structure. The crystal belonged to the spacegroup P2$_1$, with cell dimensions of $a = 63.77$ Å $b = 63.62$ Å $c = 73.70$ Å and angles of $\alpha = \gamma = 90°$ and $\beta = 103.39°$. The structure was determined by molecular replacement with PHASER [59] using two search models generated from each domain of the closed monomer of TarP_Csal.

For TarP_Rhp, high-resolution (1.1 Å) native data were processed using Xia2 [53], revealing that the crystal belonged to the space group P2$_1$2$_1$2$_1$ with cell dimensions of $a = 38.67$ Å $b = 50.5$ Å $c = 142.56$ Å and angles of $\alpha = \beta = \gamma = 90°$. The structure was determined ab initio using Arcimboldo [61], which auto-built the poly-Ala backbone (305/336 amino acids) of a monomer of TarP_Rhp that was contained within the asymmetric unit. Even though the protein was co-crystallised in the presence of 5 mM coumarate, the ligand density in the binding pocket corresponded to a single molecule of 4HPA, which had presumably co-purified with the protein. The resulting structure was built and refined in the same way as the TarP_Sse and TarP_Csal structures. The electron density showed that after cleavage of the PelB leader peptide, two residues had been lost from the N terminus of the protein, resulting in an N-terminal glutamine, which had subsequently spontaneously cyclised to form a pyroglutamate. The position of the pyroglutamate in the crystal lattice was such that it packs within a depression on a neighbouring molecule, generating several crystal contacts. It is therefore likely that crystallisation is dependent on the formation of the N-terminal pyroglutamate, which can form spontaneously in solution over a matter of weeks, potentially explaining why crystallisation took many months. Intriguingly, the cyclised glutamine is predicted to be the N-terminal residue in the mature protein, indicating that the formation of an N-terminal pyroglutamate may be important for stability of this protein. In all structures, the predicted N-terminal glutamine residue in the mature protein is numbered as residue 1.

Binding pocket volumes were calculated with CASTp using a 1.4 Å probe radius [62]. Structure validation was carried out with MOLPROBITY [63], and COOT [56]. An analysis of the flexible regions and motion between the open and closed structures of TarP_Csal was carried out with DYN-DOM [64].

Computational modelling

Computational modelling was carried out with FlexAID (NRGSuite v2.48I plugin for PYMOL v2.1.0 [65];

The PyMOL Molecular Graphics System, Version 2.0, Schrödinger, LLC.) using desolvated protein coordinates from the TarP_Rhp/4HPA structure and a molecule of coumarate from the Trap_Sse/coumarate structure. Side-chain flexibility was enabled for 15 sidechains around the binding pocket (R145, Y72, F211, Y67, W12, H17, W247, E251, S188, H122, F184, L191, V13, F250 and F192), which was defined by a sphere with radius 8 Å centred on the binding position of 4HPA. Flexibility was also enabled for the two bonds in coumarate that have rotational freedom and distance constraints of 2.7 Å were applied between the carboxyl group oxygen atoms of coumarate and the NH1 and NH2 nitrogen atoms of R145.

Accession numbers and data availability

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

Author contributions

CB involved in investigation, formal analysis, methodology, writing – original draft, validation, visualisation. RCS involved in investigation, formal analysis, writing – original draft, validation, visualisation. LW involved in investigation. JBR involved in investigation, resources, writing – review and editing. AH involved in formal analysis, writing – original draft, visualisation. GHT involved in conceptualisation, resources, supervision, writing – original draft, visualisation. DJK involved in conceptualisation, funding acquisition, project administration, resources, supervision, writing – original draft.

Peer Review

The peer review history for this article is available at https://publons.com/publon/10.1111/febs.16156.

References

1 Scheepers GH, Lycklama A, Nijeholt JA & Poolman B (2016) An updated structural classification of substrate-binding proteins. *FEBS Lett* **590**, 4393–4401.

2 Locher KP (2016) Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat Struct Mol Biol* **23**, 487–493.

3 Rosa LT, Bianconi ME, Thomas GH & Kelly DJ (2018) Tripartite ATP-independent periplasmic (TRAP) transporters and tripartite tricarboxylate transporters (TTT): from uptake to pathogenicity. *Front Cell Infect Microbiol* **8**, 33.

4 Forward JA, Behrendt MC, Wyborn NR, Cross R & Kelly DJ (1997) TRAP transporters: a new family of periplasmic solute transport systems encoded by the *dctPQM* genes of *Rhodobacter capsulatus* and by homologs in diverse gram-negative bacteria. *J Bacteriol* **179**, 5482–5493.

5 Mulligan C, Fischer M & Thomas GH (2011) Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS Microbiol Rev* **35**, 68–86.

6 Chowdhury N, Norris J, McAlister E, Lau SYK, Thomas GH & Boyd EF (2012) The VC1777-VC1779 proteins are members of a sialic acid-specific subfamily of TRAP transporters (SiaPQM) and constitute the sole route of sialic acid uptake in the human pathogen *Vibrio cholerae*. *Microbiology* **158**, 2158–2167.

7 Severi E, Randle G, Kivlin P, Whitfield K, Young R, Moxon R, Kelly DJ, Hood D & Thomas GH (2005) Sialic acid transport in *Haemophilus influenzae* is essential for lipopolysaccharide sialylation and serum resistance and is dependent on a novel tripartite ATP-independent periplasmic transporter. *Mol Microbiol* **58**, 1173–1185.

8 Zhao J & Binns AN (2016) Involvement of *Agrobacterium tumefaciens* galacturonate tripartite ATP-independent periplasmic (TRAP) transporter GaaPQM in virulence gene expression. *Appl Environ Microbiol* **82**, 1136–1146.

9 Takahashi H, Inagaki E, Kuroishi C & Tahirov TH (2004) Structure of the *Thermus thermophilus* putative periplasmic glutamate/glutamine-binding protein. *Acta Crystallogr D* **60**, 1846–1854.

10 Gonin S, Arnoux P, Pierru B, Lavergne J, Alonso B, Sabaty M & Pignol D (2007) Crystal structures of an extracytoplasmic solute receptor from a TRAP transporter in its open and closed forms reveal a helix-swapped dimer requiring a cation for alpha-keto acid binding. *BMC Struct Biol* **7**, 11.

11 Lecher J, Pittelkow M, Zobel S, Bursy J, Bönig T, Smits SH, Schmitt L & Bremer E (2009) The crystal structure of UehA in complex with ectoine—a comparison with other TRAP-T binding proteins. *J Mol Biol* **389**, 58–73.

12 Mulligan C, Geertsma ER, Severi E, Kelly DJ, Poolman B & Thomas GH (2009) The substrate-binding protein imposes directionality on an electrochemical sodium gradient-driven TRAP transporter. *Proc Natl Acad Sci USA* **106**, 1778–1783.

13 Mulligan C, Leech AP, Kelly DJ & Thomas GH (2012) The membrane proteins SiaQ and SiaM form an essential stoichiometric complex in the sialic acid tripartite ATP-independent periplasmic (TRAP) transporter SiaPQM (VC1777-1779) from *Vibrio cholerae*. *J Biol Chem* **287**, 3598–3608.

14 Wyborn NR, Alderson J, Andrews SC & Kelly DJ (2001) Topological analysis of DctQ, the small integral membrane protein of the C4-dicarboxylate TRAP transporter of *Rhodobacter capsulatus*. *FEMS Microbiol Lett* **194**, 13–17.

15 Müller A, Severi E, Mulligan C, Watts AG, Kelly DJ, Wilson KS, Wilkinson AJ & Thomas GH (2006) Conservation of structure and mechanism in primary and secondary transporters exemplified by SiaP, a sialic acid binding virulence factor from *Haemophilus influenzae*. *J Biol Chem* **281**, 22212–22222.

16 Vetting MW, Al-Obaidi N, Zhao S, San Francisco B, Kim J, Wichelecki DJ, Bouvier JT, Solbiati JO, Vu H, Zhang X *et al.* (2015) Experimental strategies for functional annotation and metabolism discovery: targeted screening of solute binding proteins and unbiased panning of metabolomes. *Biochemistry* **54**, 909–931.

17 Boerjan W, Ralph J & Baucher M (2003) Lignin biosynthesis. *Annu Rev Plant Biol* **54**, 519–546.

18 Gargulak JD, Lebo SE Jr & McNally TJ (2015) Lignin. In Encyclopedia of Chemical Technology (Kirk-Othmer, ed), John Wiley & Son, New Jersey, NJ.

19 Gall DL, Ralph J, Donohue TJ & Noguera DR (2017) Biochemical transformation of lignin for deriving valued commodities from lignocellulose. *Curr Opin Biotechnol* **45**, 120–126.

20 Davis K & Moon TS (2020) Tailoring microbes to upgrade lignin. *Curr Opin Chem Biol* **7**, 23–29.

21 Heinzkill M, Bech L, Halkier T, Schneider P & Anke T (1998) Characterization of laccases and peroxidases from wood-rotting fungi (family *Coprinaceae*). *Appl Environ Microbiol* **64**, 1601–1606.

22 McLeod MP & Eltis LD (2008) Genomic insights into the aerobic pathways for degradation of organic pollutants. In Microbial Biodegradation: Genomics and Molecular Biology (Eduardo D, ed), Caister Academic Press, Norfolk.

23 Janusz G, Pawlik A, Sulej J, Swiderska-Burek U, Jarosz-Wilkolazka A & Paszczynski A (2017) Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution. *FEMS Microbiol Rev* **41**, 941–962.

24 Ralph J, Lundquist K, Brunow G, Lu F, Kim H, Schatz PF, Marita JM, Hatfield RD, Ralph SA, Christensen JH *et al.* (2004) Lignins: natural polymers from oxidative coupling of 4-hydroxyphenyl-propanoids. *Phytochem Rev* **3**, 29–60.

25 Gulvik CA & Buchan A (2013) Simultaneous catabolism of plant-derived aromatic compounds results in enhanced growth for members of the *Roseobacter* lineage. *Appl Environ Microbiol* **79**, 3716–3723.

26 Pan C, Oda Y, Lankford PK, Zhang B, Samatova NF, Pelletier DA, Harwood CS & Hettich RL (2008) Characterization of anaerobic catabolism of p-coumarate in *Rhodopseudomonas palustris* by integrating transcriptomics and quantitative proteomics. *Mol Cell Proteomics* **7**, 938–948.

27 Hirakawa H, Schaefer AL, Greenberg EP & Harwood CS (2012) Anaerobic p-coumarate degradation by *Rhodopseudomonas palustris* and identification of CouR, a MarR repressor protein that binds p-coumaroyl coenzyme A. *J Bacteriol* **194**, 1960–1967.

28 Phattarasukol S, Radey MC, Lappala CR, Oda Y, Hirakawa H, Brittnacher MJ & Harwood CS (2012) Identification of a p-coumarate degradation regulon in *Rhodopseudomonas palustris* using Xpression, an integrated tool for prokaryotic RNA-seq data processing. *Appl Environ Microbiol* **8**, 6812–6818.

29 Salmon RC, Cliff MJ, Rafferty JB & Kelly DJ (2013) The CouPSTU and TarPQM transporters in *Rhodopseudomonas palustris*: redundant, promiscuous uptake systems for lignin-derived aromatic substrates. *PLoS One* **8**, e59844.

30 Cogan DP, Baraquet C, Harwood CS & Nair SK (2018) Structural basis of transcriptional regulation by CouR, a repressor of coumarate catabolism, in *Rhodopseudomonas palustris*. *J Biol Chem* **293**, 11727–11735.

31 Tan K, Chang C, Cuff M, Osipiuk J, Landorf E, Mack JC, Zerbs S, Joachimiak A & Collart FR (2013) Structural and functional characterization of solute binding proteins for aromatic compounds derived from lignin: p-coumaric acid and related aromatic acids. *Proteins* **81**, 1709–1726.

32 Pietri R, Zerbs S, Corgliano DM, Allaire M, Collart FR & Miller LM (2012) Biophysical and structural characterization of a sequence-diverse set of solute-binding proteins for aromatic compounds. *J Biol Chem* **287**, 23748–23756.

33 Mulligan C, Kelly DJ & Thomas GH (2007) Tripartite ATP-independent periplasmic transporters: application of a relational database for genome-wide analysis of transporter gene frequency and organization. *J Mol Microbiol Biotechnol* **12**, 218–226.

34 Gonzalez JM, Mayer F, Moran MA, Hodson RE & Whitman WB (1997) *Sagittula stellata* gen. nov., sp. nov., a lignin-transforming bacterium from a coastal environment. *Int J Syst Bacteriol* **47**, 773–780.

35 Csonka LN, O'Connor K, Larimer F, Richardson P, Lapidus A, Ewing AD, Goodner BW & Oren A (2005) What we can deduce about metabolism in the moderate halophile *Chromohalobacter salexigens* from its genomic sequence. In Adaptation to Life at High Salt Concentrations in Archaea, Bacteria, and Eukarya (Gunde-Cimerman N, Oren A, Plemenitas A, eds), pp. 267–285. Springer, Dordrecht.

36 Copeland A, O'Connor K, Lucas S, Lapidus A, Berry KW, Detter JC, Del Rio TG, Hammon N, Dalin E, Tice H *et al.* (2011) Complete genome sequence of the halophilic and highly halotolerant *Chromohalobacter salexigens* type strain (1H11(T)). *Stand Genomic Sci* **5**, 379–388.

37 Frank AM, Chua MJ, Gulvik CA & Buchan A (2018) Functional redundancy in the hydroxycinnamate catabolism pathways of the salt marsh bacterium *Sagittula stellata* E-37. *Appl Environ Microbiol* **84**, e02027-18.

38 Diaz E, Ferrandez A, Prieto MA & Garcia JL (2001) Biodegradation of aromatic compounds by *Escherichia coli*. *Microbiol Mol Biol Rev* **65**, 523–569.

39 Deangelis KM, Sharma D, Varney R, Simmons B, Isern NG, Markilllie LM, Nicora C, Norbeck AD, Taylor RC, Aldrich JT *et al.* (2013) Evidence supporting dissimilatory and assimilatory lignin degradation in *Enterobacter lignolyticus* SCF1. *Front Microbiol* **4**, 280.

40 Darby JF, Hopkins AP, Shimizu S, Roberts SM, Brannigan JA, Turkenburg JP, Thomas GH, Hubbard RE & Fischer M (2019) Water networks can determine the affinity of ligand binding to proteins. *J Am Chem Soc* **141**, 15818–15826.

41 Fischer M, Zhang QY, Hubbard RE & Thomas GH (2010) Caught in a TRAP: substrate-binding proteins in secondary transport. *Trends Microbiol* **18**, 471–478.

42 Fischer M, Hopkins AP, Severi E, Hawkhead J, Bawdon D, Watts AG, Hubbard RE & Thomas GH (2015) Tripartite ATP-independent periplasmic (TRAP) transporters use an arginine-mediated selectivity filter for high affinity substrate binding. *J Biol Chem* **290**, 27113–27123.

43 Singh J, Thornton JM, Snarey M & Campbell SF (1987) The geometries of interacting arginine-carboxyls in proteins. *FEBS Lett* **224**, 161–171.

44 Donald JE, Kulp DW & DeGrado WF (2011) Salt bridges: geometrically specific, designable interactions. *Proteins* **79**, 898–915.

45 Mattes TE, Alexander AK, Richardson PM, Munk AC, Han CS, Stothard P & Coleman NV (2008) The genome of *Polaromonas* sp. strain JS666: insights into the evolution of a hydrocarbon- and xenobiotic-degrading bacterium and features of relevance to biotechnology. *Appl Environ Microbiol* **20**, 6405–6416.

46 Unuofin JO, Okoh AI & Nwodo UU (2019) Utilization of agroindustrial wastes for the production of laccase by *Achromobacter xylosoxidans* HWN16 and *Bordetella bronchiseptica* HSO16. *J Environ Manage* **231**, 222–231.

47 Waterhouse AM, Procter JB, Martin DMA, Clamp M & Barton GJ (2009) Jalview version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191.

48 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797.

49 Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC & Jahn D (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* **33**, W526–W531.

50 Emanuelsson O, Brunak S, von Heijne G & Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**, 953–971.

51 Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd & Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343–345.

52 Stols L, Millard CS, Dementieva I & Donnelly MI (2004) Production of selenomethionine-labeled proteins in two-liter plastic bottles for structure determination. *J Struct Funct Genomics* **5**, 95–102.

53 Winter G (2010) xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43**, 186–190.

54 Sheldrick GM (2010) Experimental phasing with *SHELXC/D/E*: combining chain tracing with density modification. *Acta Crystallogr D* **66**, 479–485.

55 Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, Hintze B, Hung LW, Jain S, McCoy AJ *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D* **75**, 861–877.

56 Emsley P, Lochkamp B, Scott WG & Cowtan K (2010) Features and development of *Coot*. *Acta Crystallogr D* **66**, 486–501.

57 Lebedev AA, Young P, Isupov MN, Moroz OV, Vagin AA & Murshudov GN (2012) *JLigand*: a graphical tool for the *CCP*4 template-restraint library. *Acta Crystallogr D* **68**, 431–440.

58 Kovalevskiy O, Nicholls RA, Long F, Carlon A & Murshudov GN (2018) Overview of refinement procedures within REFMAC5: utilizing data from different sources. *Acta Crystallogr D* **74**, 215–227.

59 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *J Appl Cryst* **40**, 658–674.

60 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG,

McCoy A *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D* **67**, 235–242.

61 Rodríguez DD, Grosse C, Himmel S, González C, de Ilarduya IM, Becker S, Sheldrick GM & Usón I (2009) Crystallographic *ab initio* protein structure solution below atomic resolution. *Nat Methods* **6**, 651–653.

62 Tian W, Chen C, Lei X, Zhao J & Liang J (2018) CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res* **46**, W363–W367.

63 Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB *et al.* (2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293–315.

64 de Groot BL, Hayward S, van Aalten D, Amadei A & Berendsen HJC (1998) Domain motions in bacteriophage T4 lysozyme; a comparison between molecular dynamics and crystallographic data. *Proteins* **31**, 116–127.

65 Gaudreault F, Morency LP & Najmanovich RJ (2015) NRGsuite: a PyMOL plugin to perform docking simulations in real time using FlexAID. *Bioinformatics* **31**, 3856–3858.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Phylogenetic tree of TRAP transporter SBPs in diverse marine bacteria.

Fig. S2. Fluorescence titrations of TarP_Sse, TarP_Csal and TarP_Rhp with ligninderived aromatic ligands.

Fig. S3. Amino acid sequence alignment of TarP_Rhp, TarP_Csal and TarP_Sse.

Fig. S4. Details of structural features from the TarP_Rhp complex with 4HPA.

Fig. S5. The domain organisation of TarP SBPs.

Fig. S6. Analysis of plane distortions in the coumarate ligand.

Fig. S7. Superposition of TarP_Csal with possible related TRAP SBPs.

Table S1. Locus tag identifiers of the TRAP SBPs used in this study for the phylogenetic analysis in Fig. S1.

Video S1. Overview of the conformational changes associated with coumarate binding in TarP_Csal, with major residues surrounding the ligand binding pocket (E190, Y74 and R148) highlighted. The video was generated using Chimera by interpolating between the open apo-structure and the closed ligand-bound structure of TarP_Csal. The protein backbone is shown as pale green ribbons with important residues and secondary structure elements shown as sticks. Coumarate is shown as darker green sticks.

Video S2. Conformational changes around the key arginine (R148) in TarP_Csal in more detail, including the changes in the hydrogen bonding network between the arginine and the surrounding beta sheet during domain closure (see also Figure 5A and 6). Hydrogen bonds and distances are shown in orange. The protein backbone is shown as pale green ribbons with important residues and secondary structure elements shown as sticks. Coumarate is shown as darker green sticks. Video generated using Chimera as for Video S1.