Biological degradation of Polyethylene terephthalate (PET) plastic and assimilation of the corresponding monomers ethylene glycol and terephthalate (TPA) into central metabolism offers an attractive route for bio-based molecular recycling and bioremediation applications. A key step is the cellular uptake of the non-permeable TPA into bacterial cells which has been shown to be dependent upon the presence of the key \textit{tphC} gene. However, little is known from a biochemical and structural perspective about the encoded solute binding protein, TphC. Here, we report the biochemical and structural characterisation of TphC in both open and TPA-bound closed conformations. This analysis demonstrates the narrow ligand specificity of TphC towards aromatic para-substituted dicarboxylates, such as TPA and closely related analogues. Further phylogenetic and genomic context analysis of the \textit{tph} genes reveals homologous operons as a genetic resource for future biotechnological and metabolic engineering efforts towards circular plastic bio-economy solutions.
The global appetite for plastic products has transitioned our planet into an era of the “Plastic Age”\(^1\). Polyethylene terephthalate (PET) plastic is the most commonly used plastic in the packaging of beverages, food and pharmaceuticals. Since its first synthesis back in 1941, PET has gradually emerged as the world’s favorite food-safe plastic due to its robustness, chemical inertness and durability. Although regarded as non-toxic and 100% recyclable, single-use convenience-sized PET bottles have made PET plastic the third most collected debris in beach clean-ups in more than 100 countries\(^2\) and is overwhelmingly omnipresent in the terrestrial ecosystem\(^3,4\). The post-consumer recyclability of plastics is still questionable owing to a number of factors\(^5,6\) that have turned PET from a miraculous material into the scourge of the land and sea. For instance, the PET-recycling rate was estimated to be only ~30% in 2015 in the US, and in Europe 25% of post-consumer plastic waste still went into landfill in 2018 (ref.\(^7\)). This evidently ubiquitous plastic footprint, however, has not impacted upon the demand and production given the essentiality of plastics in our daily lives. The global PET market in 2017 was valued at ~USD 24 billion and is expected to hit USD 39 billion by 2027 (ref.\(^8\)). In 2016, around 485 billion PET bottles were manufactured, and around 583.3 billion are forecast to be produced in 2021 (ref.\(^8\)). The global demand for PET in 2030 is forecast to amount to 42 metric tons\(^10\).

Both mechanical and chemical PET-recycling methods are available; however, widespread use is limited\(^11\). Indeed, some concerns remain over the product quality and environment impact of the mechanical and chemical PET recycling, respectively\(^6\). Microbial and enzyme-based plastic waste biodegradation and recycling offer promising alternative strategies due to the use of benign conditions and potential as a cost-effective environment-friendly approach\(^12–16\). Therefore, there is great interest in finding better strategies for PET bioconversion and recycling through engineering robust enzymes and microbial strains for its degradation, uptake and assimilation. Terephthalic acid (TPA) and ethylene glycol (EG), which together form a polymer chain, are the basic building blocks of PET and can be released by enzymatic hydrolysis via the action of different types of bacterial and fungal origin hydrolysates, such as esterases, lipases, cutinases and carboxylesterases\(^17–20\). TPA is converted to PCA via a pathway encoded by the tphC gene and X-ray crystallography to biochemically and structurally characterise TphC from \(Comamonas\) sp. strain E6. Further phylogenetic and taxonomic analysis was used to reveal homologous proteins that bind to a macromolecular protein through an apparent ligand interactions are poorly understood. So far, the structure of only six TTT-SBPs have been determined, two from \(Rhodopseudomonas palustris\)\(^47,48\), three from \(Bordetella pertussis\)\(^51–53\) and one from \(Polaromonas\) sp. These protein structures show a common “Venus flytrap” fold comprising of two globular domains separated by a cleft that folds around the ligand. Despite cellular uptake of TPA being a key step, the confirmation of the tphC-dependent uptake of TPA at the genetic- and potential for TphC to be used in engineered strains for the cellular uptake of TPA and bioconversion of plastic waste\(^14\), there have been no reports to date on the biochemical or structural characterisation of TphC. Due to the relatively recent xenobiotic introduction of PET and TPA into the environment, we were inspired to explore whether TPA is indeed the cognate ligand for TphC, or whether other closely related xenobiotics naturally found in the environment are also recognised by TphC.

Here, we employed differential scanning fluorescence (DSF) assay, mutational analysis, isothermal titration calorimetry (ITC) and X-ray crystallography to biochemically and structurally characterise TphC from \(Comamonas\) sp. strain E6. Further phylogenetic and taxonomic analysis was used to reveal homologous operons to explore the diversity and origin of these xenobiotic catabolism genes and to provide a genetic resource for future biotechnological and metabolic engineering efforts towards circular plastic bio-economy solutions.

Results

Production and purification of TphC. The pre-protein form of TphC from \(Comamonas\) sp. strain E6 contains a predicted signal peptide sequence and the mature form is composed of 294 amino acids (N29-L322) (Supplementary Fig. 1). In order to characterise TphC, the recombinant gene, encoding the mature form, was cloned and then overexpressed in \(Escherichia coli\) BL21(DE3), and the protein was purified in two consecutive chromatographic steps of metal-affinity chromatography and gel-filtration chromatography. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the eluates shows the presence of an expected ~34 kDa protein band equivalent to an N-terminal 6×His-tag fusion of TphC (Supplementary Fig. 1).

Ligand specificity screening with differential scanning fluorimetry. In order to determine the ligand recognition profile of TphC, we procured and/or synthesised TPA analogues and their salts (Supplementary Table 1). TphC was probed for ligand binding against the set of 61 compounds by DSF assay (Supplementary Table 2). DSF can be used to rapidly screen compounds that bind to a macromolecular protein through an apparent
stabilisation against thermal denaturation, observed as a shift in the midpoint of the melting curve (Fig. 2 and Supplementary Fig. 2). TphC showed the most stabilised binding interaction with terephthalate (1) with an apparent $\Delta T_m$ of $8.3 \pm 1.1$ °C compared to TphC alone (56.1 ± 0.5 °C). Significant TphC-stabilising interactions were also observed with 2-hydroxyterephthalate (4), 2-aminoterephthalate (7), 2,5-dihydroxyterephthalate (10), and bipheny-4,4′-dicarboxylate (21) with $\Delta T_m$ of $5.9 \pm 0.7$ °C, $2.2 \pm 0.6$ °C, $2.8 \pm 0.4$ °C and 4 ± 0.8 °C, respectively. The presence of 2,5-pyridinedicarboxylate (2) and 2,6-naphthalenedicarboxylate (22) have less pronounced, but still significant, stabilising interactions with TphC, displaying $\Delta T_m$s of 1.4 ± 1.1 °C and 1.4 ± 0.7 °C, respectively. The other para-substituted dicarboxylate analogues (3, 5–6, 8–11) regioisomers (12–13), heteroaromatics (14–19), bicyclic aromatics (20, 23), the monocarboxylate and carboxylate isosteres (24–44), unsaturated phenylpropanoates (45–50), phenols (51–52), aromatic esters (53–56) and aliphatic dicarboxylates (57–61) had either negligible effect on $\Delta T_m$s or their interactions with TphC were found to be slightly destabilising under the assay conditions. This indicates that for optimal interaction a six-membered para-substituted aromatic dicarboxylate is required, plus limited additional substitution by hydroxyl groups around the ring and minor heteroaromatic modifications are tolerated, and also extended aromatic systems are partially tolerated.

**Ligand binding assessment with ITC.** The initial DSF assay screening with a variety of aromatic and aliphatic ligands by thermal-shift analysis revealed the ligand-preference of TphC, with the TPA (1) and its close structural analogues showing the highest thermal-shift among the ligands tested. To further obtain the thermodynamic parameters for the binding affinity of the ligand hits, ITC experiments were performed (Fig. 3 and Supplementary Fig. 3). The determined $K_D$ values for the ligands reveal receptor affinities in the µM range and a ligand preference similar to the one observed in the initial DSF screen, with terephthalate (1) and its derivatives: 2-hydroxyterephthalate (4) and 2,5-dihydroxyterephthalate (10) showing the highest affinities with $K_D$ ranging from 0.36 to 1.9 µM (Table 1). Over two to three orders of magnitude weaker affinities were observed for 2-aminoterephthalate (7) and 2,5-pyridinedicarboxylate (2) than for terephthalate (1), and a considerably lower affinity for 2,6-naphthalenedicarboxylate (22). The TphC-ligand interactions for these six ligands were enthalpically favoured and most ligands (Table 1), with the exception of 22, reported a small entropic penalty (−$T\Delta S$) upon interaction, indicative of conformational change or solvation effect\(^\text{34,35}\). The reduced affinity observed for 2, 7 and 22 emphasises the structural constraint for a six-member aromatic ring with a pair of para-oriented carboxylate groups for optimal coordination in the binding site (Supplementary Fig. 4). This is further evidenced with the attempts to titrate TphC with a bulkier biphenyl-4,4′-dicarboxylate (21), not displaying any heat of binding, suggesting that 21 has a much lower affinity for ITC assessment or displays an entropically driven interaction. To rule out nonpecific interactions, the specific binding of 21 to TphC was indirectly ascertained through a competitive titration of 1 with TphC saturated with a 10-fold excess concentration of 21 (Supplementary Fig. 5), where the presence of 21 completely abolishes TphC–terephthalate interaction.

**Fig. 1 Terephthalate cellular uptake and assimilation operons.** A Schematic representation of tph and tpi operons in the chromosome of Comamonas sp. strain E6. The catabolic genes of the tph operon, encoding a three-component TPA 1,2-dioxygenase (TPADO), are represented in gold (tphA2), orange (tphA2) and pink (tphA1). The tph operon regulator (tphR), a gene encoding a diol dehydrogenase (tphR) and a gene encoding a periplasmic solute-binding protein (tphC), are represented in green, teal and mauve, respectively. The transport genes (tpiBA) in tpi operon, encoding a set of transmembrane transport proteins, are shown in peach and light blue, respectively. B Schematic of TPA transport and catabolism. The tphR encoded activator (TphR) responds to the presence of inducer TPA and leads to the expression of a set of capture, transport and catabolic proteins to metabolise TPA in a successive binding, transport and catabolic steps: TphC in the periplasm first binds to TPA and relays it to a pair of transmembrane proteins, TpiB and TpiA, which in turn transport it to the cytoplasm, where it is converted by a three-component TPADO to 1,6-dihydroxycyclohexa-2,4-diene-dicarboxylate (DCD). The DCD is acted upon by TphB (1,2-dihydroxy-3,5-cyclohexadiene-1,4-dicarboxylate dehydrogenase) and converted to protocatechuate (PCA), which is then funnelled into the central metabolic pathways.
Structure of open apo-TphC structure. In order to understand the open structure of TphC, we sought to crystalise TphC alone. Purified TphC was readily crystallised, apo-TphC crystals diffracted to a resolution of 1.97 Å and structure determination via molecular replacement was performed in Phaser using a model derived from TctC from *Polaromonas* sp (PDB 4X9T). The overall structure of TphC maintains the general architecture of two α/β globular domains linked by a pair of β-strands as previously reported in TTT-SBPs TctC (4X9T) and Bug27 (2QPQ) (Fig. 4A and Supplementary Fig. 6). Data collection and refinement statistics for the apo-TphC structure are summarised (Supplementary Table 3). Residues (P28-D37, PFSAGGTD) map to the motif [P*-F-X-A-G*-G*-X-X-D*] which is highly conserved in TTT-SBPs (conserved residues underlined)\(^{35}\). In the open apo conformation TphC has a solvent accessible surface area of ~13048 Å\(^2\) with a large solvent accessible cavity at the predicted ligand binding site\(^{51}\).

Structure of closed-TPA-bound TphC. In order to better understand recognition of the TPA ligand and any conformational change associated with ligand binding, we sought to crystallise TphC in complex with TPA. Purified TphC was readily co-crystallised in the presence of 2 mM TPA and crystals diffracted to 2.4 Å resolution. The structure was solved using molecular replacement with a search model derived from AdpC from *Rhodopseudomonas palustris* (PDB 5OEI). The overall structure of the TphC in complex with TPA is shown in Fig. 4B. Supplementary Table 4 summarises the data collection and refinement statistics for the closed TphC–TPA complex. The general architecture of the previously reported TTT-SBPs is maintained in the TphC–TPA structure. Sequence alignment, secondary structure features and conserved motifs for TphC, AdpC and the other TTT-dependent SBPs in the closed conformation are shown (Supplementary Fig. 7 and Supplementary Table 5).

Fig. 2 TphC unfolding profile and melting temperature determination. A The presence of different ligand induced shifts of the TphC melting curve (light blue) to the right (mauve to dark blue melting curves) resulting in a maximum stabilisation of -8 °C. The first derivative of melting curves enables the determination of melting temperature \(T_m\) (inflection point in the raw fluorescence traces, see Supplementary Fig. 2. Fl fluorescence intensity). B Chemical structures of the ligands producing a thermal shift terephthalate (1), 2,5-pyridinedicarboxylate (2), 2-hydroxyterephthalate (4), 2-aminoterphthalate (7), 2,5-dihydroxyterephthalate (10), bipheny-4,4’-dicarboxylate (21) and 2,6-naphthalenedicarboxylate (22). C The degree of protein stabilisation in the presence of various ligands was calculated as \(\Delta T_m\). The data shown are the mean of independent experiments (\(n=3\)) and the error bars show the standard deviation (SD) of the mean. The statistical difference between ligand groups and the ligand-free TphC control was assessed using one-way ANOVA using Dunnett's multiple comparisons tests (\(p < 0.05\)), where asterisks denote statistically significant difference (* \(p < 0.02\); **** \(p < 0.0001\)) compared to the control. Noted p values 1 < 0.0001, 2 0.0184, 4 < 0.0001, 7 < 0.0001, 10 < 0.0001, 21 < 0.0001, 22 0.0115. Each well of a 96-well plate contained 50 µl of total reaction buffer (25 mM Tris-HCl (pH 7.5)/200 mM NaCl), 60 µM of TphC, 1200 µM of ligand and 1× SYPRO orange dye, and fluorescence was monitored at each 1 °C rise in temperature from 20 to 95 °C.
Superposition of the open and closed TphC structures highlights an ~36° closure between the two domains upon TPA binding. DynDom analysis identifies residues 115–117 and 241–242 as the hinge region. Upon closure the solvent accessible surface area of TphC decreases from 13,048 to 12,495 Å² encapsulating a closed cavity of 280 Å³ in which the TPA is bound. Hydrogen bonds between both the proximal and distal carboxylate groups of the TPA stabilise the closed conformation of the protein along with a network of direct and water-mediated hydrogen bonds highlighted in Fig. 5. The bound TPA is recognised through a number of direct and water-mediated hydrogen bond interactions. The TPA carboxylate group buried deepest within the pocket is recognised and stabilised through direct (T155, T242) and water-mediated hydrogen bonds (T155, Q150, N197). The second carboxyl group of the TPA hydrogen bonds to the backbone amino groups of A36 and with additional water-mediated interactions with the backbones of G34, S31 and T179. The interactions at this second carboxylate group span between the two TphC domains stabilising the closed conformation (Fig. 5). The binding pocket is bordered on one side by a conserved aromatic residue (F30); the positioning of this sidechain is facilitated by the proximity of a conserved glycine (G178). Unlike other TTT-SBPs characterised to date, the ligand site of TphC contains bulky amino acids P86 and Q150, the sidechains of which create a very narrow pocket 6.4 Å across, which would appear to confer the observed selectivity toward planar aromatic ligands. In addition, the Q150 sidechain is closely bounded by S154 (in addition to S156 and F194 not shown), all of which together would prevent alternative rotamers being sampled to increase the diversity of ligands accommodated.

Docking and TPA depletion. In order to visualise their potential binding site interaction, the seven ligands were docked into the binding site of the closed TphC, terephthalate 1.
2,5-pyridinedicarboxylate 2, 2-hydroxyterephthalate 4, 2-aminoyterephthalate 7, 2,5-dihydroxyterephthalate 10, biphenyl-4,4’-dicarboxylate 21 and 2,6-naphthalenedicarboxylic acid 22 (Supplementary Fig. 8 and Supplementary Table 6). To investigate the transport of the selected analogues in vivo, we performed substrate depletion assays with resting cells of *P. putida* KT2440 ΔpcaGH harbouring the plasmid pJCBAAtG45. As previously shown, resting cells were able to transport and thus deplete TPA by ~20% after 60 min (Supplementary Fig. 9). However, due to the lack of their enzymatic conversion, neither depletion nor transport of the selected analogues (2, 4, 7, 10, 21 and 22) could be determined. Further experimentation would help investigate the transport of these ligands by TphC in vivo.

**Mutational analysis of amino acids involved in ligand binding.**

In order to confirm the involvement of the amino acids in TPA and analogue recognition, and ligand-induced conformation change a series of mutant proteins were created and screened with D8F. This included mutation of amino acids involved in recognition of TPA via direct H-bond interactions (Fig. 5), T155A, T242A, T155A/T242A, those involved in creating a narrow binding pocket Q150S, P86T/Q150S, along with an amino acid within the hinge region V116P. Mutation of the amino acids involved in direct ligand recognition had little or no impact (<3 °C) on the intrinsic stability of the TphC as measured by $T_m$; however, the P86T/Q150S and Q150S displayed a higher $T_m$ (>3 °C) indicating higher intrinsic thermal stability (Supplementary Fig. 10A). In addition, the hinge mutant V116P displayed a significantly lower $T_m$ (<10 °C) indicating lower intrinsic thermal stability. The seven most active ligands (1, 2, 4, 7, 10, 21 and 22) were then screened against the protein mutants (Supplementary Fig. 10B). The loss of H-bonding potential for T155A is clearly...
Phylogenetic analysis of TphC homologues and genomic context analysis. The gene encoding TphC is located in the tph operon in Comamonas sp. strain E6, which is composed of the transcriptional regulator (tphR), solute-binding protein (tphC), TPA dioxygenase subunit A2 (tphA2), TPA dioxygenase subunit A3 (tphA3), TPA dioxygenase subunit B (tphB), TPA dioxygenase subunit A1 (tphA1) (Fig. 1A). Phylogenetic analysis of TphC from Comamonas sp. strain E6 identified SBP homologues (n = 100), with sequence identity ranging from 39 to 98% relative to the query (Fig. 7). Taxonomic analysis reveals that the entries originate from 80 different species, 46 genera, 12 families, and majority of the entries (81) originate from species belonging to the Burkholderiales order (β-proteobacteria), 13 to the Rhizobiales order (α-proteobacteria), and 6 to the Rhodospirillales order (α-proteobacteria) (Supplementary Data 1). Due to the environmental and biotechnological importance of TphC and the associated genes in the uptake and catabolism of TPA, we analysed the gene content, operon architecture and encoded protein homology of the operons flanking the tph homologues (Fig. 7 and Supplementary Fig. 11). The tph operon content was conserved in a subset of these selected entries (22/100), including from Ideonella sakaiensis, representing species from five different families within the α- and β-proteobacteria classes (Supplementary Data 1). TphC homologues encoded by the tph-like operons contained highly conserved amino acids at the positions where the sidechains make direct contact with TPA (T155, T242) or create the planar binding pocket (P86, Q150) (Fig. 7). A small subset of operons encoding TphC-like proteins contain A242 rather than T242, consistent with the observed non-essential nature of T242 in TPA recognition (Supplementary Fig. 10). Interestingly the same conserved binding site amino acids are also present in TphC homologues not associated with tph-like operons (Fig. 7A), indicating alternative genomic loci as sources of potential novel TPA catabolic genes.

Homology alignment of the TphC sequences associated with the tph-like operons indicated the highly conserved amino acids within the ligand binding site (Supplementary Fig. 12). Sequence identity, represented by calculating the percentage identity for each protein type across the tph-like operons, was generally modest across all proteins (62%), with TphR the least conserved (55%) and TphA2 the most conserved (77%) (Supplementary Fig. 13). Within these operons the predominate operon architecture was the same as that found in Comamonas sp. strain E6. Here tphC is located at the 5’ end of the operon and tphR divergently located (head-to-head), with respect to the corresponding promoter (5’H). This 5’H operon structure is observed in both α- and β-proteobacteria classes (e.g. Comamonas thiooxidans and Roseomonas deserti) (Fig. 7B and Supplementary Fig. 11). Within an alternative operon structure, tphC is located at the 3’ end of the operon, this operon architecture was observed exclusively in α-proteobacteria (e.g. Bradyrhizobium ivoirensense) (Supplementary Data 1). Within these genomic loci the presence of transposable elements flanking the tph operon within Comamonas thiooxidans and other operons is indicative of horizontal gene transfer of the tph operon reflecting the accessory nature of these genomic loci (Fig. 7B) (Supplementary Fig. 11). Additionally a number of operons contained gene insertions within the tph operon encoding for a hydrolase (MHETase), observed for all substrates where no significant ligand-induced stabilisation was observed. Unexpectedly the T242A mutant still bound to TPA and ligand 4 indicating the less stringent requirement of T242 in substrate recognition. The double-mutant T155A/T242A also displayed no ligand-induced stabilisation in the presence of TPA or the analogues consistent with the single mutant T155A. A small, but still significant shift was observed in the presence of TPA towards Q150S, although the other binding pocket mutant, P86T/Q150S, was no longer able to bind TPA. Finally, the hinge mutant V116P displayed ligand-induced stabilisation in the presence of TPA and five of the analogues (4, 7, 10, 21 and 22); for TPA this stabilisation (∆Tm) was greater than that observed with the wild-type TphC. This greater shift can be explained by the lower intrinsic thermal instability of the V116P (Supplementary Fig. 10A).
oxidoreductase, and a transporter (e.g. Comamonas composti, Ideonella sakaiensis, and Piscinibacter defluvii), presumably involved in TPA recognition and catabolism.

**Discussion**

In this study, we characterised the TphC SBP that is a key component in the cellular uptake of TPA, the breakdown product of PET plastic45. TphC belongs to the TTT family in Comamonas sp. strain E6 and through biochemical and structural analysis, the ligand-binding properties and 3D structure were determined in both open and ligand-bound closed forms. TTT-dependent SBPs, such as TphC belong to the Cluster E-II according to the classification of Scheepers et al.57. As with the tripartite ATP-independent periplasmic (TRAP) class of SBP-dependent transporters, they are symporter proteins driven by an electrochemical ion-gradient, rather than ATP as per the SBP-dependent class of ABC transporters46.

Screening with DSF showed TphC binds to para-substituted aromatic dicarboxylates. Substitution of the aromatic ring with hydroxyl and amino groups, and the presence of hetero-atoms within the ring, is tolerated, but with a significantly lowered ligand-induced stabilisation. This seems to indicate that the binding pocket of TphC is structurally quite constrained. Furthermore, the efficient binding of the terephthalate 1 ligand but not the C6 aliphatic or unsaturated aliphatic ligands, adipate 59 and trans, trans-muconate 57, revealed the need for an aromatic structure for the binding ligand. The TTT-SBPs characterised so far are specific for aliphatic ligands47,49,50 with the sole exception of the B. pertussis Bug27 protein which is reported to bind an aromatic mono-carboxylated nicotinate51. Our model proposes that TphC has a narrower more selective binding site, whereas AdpC is more relaxed, which is consistent with biochemical data in the literature that shows that AdpC binds to C6–C8 aliphatic dicarboxylates plus trans, trans-muconate47. The ITC determined TphC affinities for the ligand-hits were in µM range with terephthalate 1 and its derivatives: 2-hydroxyterephthalate 4 and 2,5-dihydroxyterephthalate 10 showing the greatest affinities (0.36 – 1.9 µM). The sub-micromolar TphC affinity observed for 1 (0.36 µM) is comparable to other high-affinity SBPs reported for Bug27-Nicotinate/Nicotinamide (0.36 µM) 51, AdpC-Adipate (0.55 µM)47 and MatC-Malate (~0.02 µM)48.

In comparison with other open/apo structures, even though TphC has low similarity and identity to other apo SBP structures TctC (26%) and Bug27 (29%), the two domains of TphC have RMSD of 0.83 and 0.44 for the corresponding domains within TctC (PDB 4X9T), and 0.47 and 0.89 within the Bug27 (PDB 2QPQ) domains (Supplementary Table 5). All three SBPs have the same architecture; however, there is significant differences...
between the orientation of domains between TphC and TcC, specifically in the upper domains around α5-6 and β5-8, and the ligand-binding loop between β6 and α5 (Supplementary Fig. 6). TphC shows a 36.3° closure angle, whereas previously resolved apo- and ligand-bound structures of Bug27 reported a closure angle of 24.7°. Other periplasmic binding proteins show closure angles ranging from 15° to 60°. Similar to MatC and AdpC, the absence of positively charged residues (to counteract the negative charges of the ligand carboxylate groups) in the ligand-binding pocket of TphC is observed. The TphC–TPA ligand-binding site appears similar to the aliphatic-SBP complex (oxoadipate-AdpC), with comparable distances between the corresponding residues responsible for H-bonding with the dicarboxylate groups. Within SBPs that recognise the smaller aspartate and malate ligands (BugD and MatC), the C5 ligands are positioned 90° rotated along the dicarboxylate axis relative to TPA (and the other C5 ligands), with the carboxylate H-bonding to a neighbouring serine residue (S151). While this serine is conserved in TphC (S156), a bulky glutamine residue in TphC (Q150) occludes the alternative H-bonding site and thus provides the rationale for why the smaller dicarboxylate ligands are not recognised by TphC. The bulky residues P86 and Q150 present within TphC create a narrow pocket, 6.4Å across, which appears to exclude aliphatic C5 ligands.

Along with the majority of TTT-SBPs reported to date, TphC and its homologues identified here are not located in the same operon as the TTT-dependent membrane transporter subunit genes (tpiA and tpiB), but are located within the tph operon along with the genes encoding transcriptional regulator (TphR) and the TPA dioxygenase subunits (TphA1 and TphB) (Fig. 7). Due to the genomic co-localisation of the tphC and the catabolic pathway for TPA (tph), we sought to use TphC as genomic fish-bait to locate other homologous gene clusters of biotechnological interest. This approach has previously been employed to identify arylmalonate decarboxylase. TphC was not observed is unclear. One possibility is that TphC evolved from an ancestral or pre-industrial SBP that displayed side-activity or promiscuity, and TphC has since lost affinity to its original ligand, as has been proposed for the evolution of enzymes. Considering the evolutionary time span of ~80 years since the first production of TPA, and the low conservation observed between the SBP’s associated with the putative tph operons (Fig. 7), it would seem likely that TphC homologues might have evolved from a pool of multiple ancestral SBPs rather than from a common ancestor. This notion is also consistent with the low degree of sequence conservation and global distribution of xenobiotic degradation pathways, e.g., those involved in plastic and organohalide degradation. In this study, we provide evidence for the structural basis of TPA recognition by TphC through DSF, ITC and X-ray crystallography. Together with TPA, we report the binding potential of TphC towards some other TPA derivatives and analogues which have varied applications. For instance, 2,5-dihydroxyterephthalic acid is a useful monomer for the synthesis of high strength fibres, 2-aminoterephthalic acid and 2,5-pyridinedicarboxylic acid have application in the synthesis of lanthanide coordination polymers and 2,6 naphthalene dicarboxylic acid is used as a monomer for the production of polynethylene naphthalate (PEN) esters which are considered superior to PET for certain applications. In summary this biochemical, structural and phylogenetic analysis provides useful insights into the ligand-recognition potential of TphC and opens up a new array of opportunities to engineer heterologous hosts for the uptake and assimilation of the breakdown products from PET and other polymers.

Methods

Chemicals, genes and reagents. All reagents were procured at the highest purity available from Sigma Aldrich Ltd. (Dorset, UK) unless otherwise specified. Genes were synthesised by IDT DNA technologies, and molecular biology reagents were purchased from NEB unless otherwise stated.

Cloning, Expression and Purification of TphC. The tphC gene from Comamonas sp. E6 (BAE47076.1), excluding the first 26 amino acids of the TphC protein (WP_019043844.1) comprising of the N-terminal signal sequence, was chemically synthesised with the codons optimised for expression in . The cloning of tphC was undertaken using Gibson ligation of linearised plasmid pET44a+ with the synthetic truncated DNA sequence of tphC. The resultant plasmid was transformed into E. coli DH5α and subsequently into E. coli expression host BL21(DE3) for overexpression. The cells bearing the plasmid PET44a+his-TphC were initially grown at 37 °C with shaking (200 r.p.m.) in tert broth medium supplemented with 100 µg/mL ampicillin and 0.2% (w/v) glycine. The overnight seed culture was then equilibrated to 25 °C and then 200 mL was used to inoculate 22 L auto-induction medium, pre-equilibrated to 25 °C, consisting of (w/v): 2% tryptone, 0.5% yeast extract and 0.5% NaCl and supplemented with 0.4% (v/v) glycerol, 0.002% (w/v) glucose and 0.01% lactose and 100 µg/mL of ampicillin. The culture was grown at 25 °C for 48 h. The culture was centrifuged at 5000g at 4 °C for 1 h and the cell pellet was re-suspended (100 g wet cells/200 mL buffer) in the binding buffer (25 mM Tris–HCl, pH 7.5/500 mM NaCl/10 mM imidazole). The cells were lysed using a pre-chilled CF1 cell disrupter (Constant Systems Ltd, UK) in the presence of EDTA-free protease-inhibitor cocktail (Roche) and endonuclease (Benzonase) and the lysate was centrifuged at 125,000g for 1 h. The resulting supernatant was loaded onto a 5 mL Ni-NTA agarose column, pre-equilibrated with 10-column volumes of binding buffer. The column was then washed with 5-column volumes of washing buffer (25 mM Tris–HCl, pH 7.5/500 mM NaCl/20 mM imidazole) and the bound TphC was step-eluted with 3 x 2-column volumes of elution buffers containing increasing concentrations of imidazole (25 mM Tris–HCl, pH 7.5/500 mM NaCl/40-150 mM imidazole). To remove imidazole and to concentrate the protein, the eluted TphC fractions were pooled and buffer exchanged into a Tris buffer (25 mM Tris–HCl, pH 7.5/200 mM NaCl) using PD-10 columns. The buffer-exchanged TphC protein was concentrated with a centrifugal filter concentrator (Amicon, 10kD MWCO, Merk Millipore) and subjected to gel-filtration chromatography on a high-performance gel-filtration column (AKTA, Superdex-200 26/600 GL, GE Healthcare) previously equilibrated with the Tris buffer. The peak fractions from gel-filtration column were collected and assessed for purity on SDS-PAGE. The protein concentrations were determined from the calculated extinction coefficient of TphC at 280 nm. The mid-peak fractions were pooled and concentrated, as described above, to 5–10 mg/mL and the concentrated protein was stored in 100 µL aliquots at ~80 °C until later use.

Ligand synthesis. Most of the TPA analogues are insoluble in aqueous solution, but are soluble in DMSO which makes it difficult to assess them using ITC. As a result, a majority of the ligands were sourced as the sodium salt or the sodium salt of the free acid when the salt was not commercially available. Sodium as sodium as described: organic acid (3 mmol) was dissolved in NaOH solution (3 mL, 2 M), this solution was slowly added to acetone (40 mL), precipitating the sodium salt. The solid was isolated by filtration, washing with acetone (3 x 5 mL), and remaining solvent removed in vacuo, affording dry solid product. Elemental Analysis was performed and the purity calculated (Supplementary Table 1).

DSF assay. TphC protein–ligand interactions were investigated using a DSF assay. Preliminary assays were conducted to determine a minimum TphC concentration required in the assay to achieve a fluorescence signal strength above the noise in the assay. The assays were conducted at a fixed protein to ligand ratio of 1:10 between the protein concentration range of 3–1000 µM in a 96-well plate in a total reaction volume of 50 µL buffer (25 mM Tris–HCl, pH 7.5/200 mM NaCl) and assessed for fluorescence signal strength above the noise in the assay. The assays were run on a QuantStudio 3 Real-Time PCR System (ThermoFischer Scientific, UK) with melt curve experiments equilibrating to 20 °C before ramping to 95 °C while
Ligand depletion assays. The concentration of 60 µM was determined to be the optimal concentration for the assay. The ligand and the ligand solution titrated was set up as follows: ligand (5 µL, 1200 µM), protein (10 µL, 60 µM), SYPRO orange dye (5 µL, 10x) and Tris-Cl buffer to a total volume of 50 µL. The ligand was dissolved in DMSO and the same concentration (v/v) of DMSO was added to the parallel controls. Data were processed in excel and analysis in Prism 8 (GraphPad).

Isothermal titration calorimetry. All ITC measurements were carried out using a MicroCal Auto-iTC200 (Malvern Panalytical Ltd, UK) in a passive mode with a microcalorimeter cell containing 300 µL of Tris-HCl buffer to a total volume of 50 µL. Where the ligand was dissolved in DMSO. The active reaction cell contained 150 µM of 1 mM of ligand. The concentration and the titrating syringe. The active reaction cell contained 300 µL of protein and the syringe contained 120 µL of 1 mM ligand. The protein–ligand titrations were carried out at 22 °C with an initial pilot injection of reduced volume (0.5 µL, 1 x) followed by 15 successive injections (2.5 µL, 5 x) spaced 300 s apart. Reference buffer to buffer titrations were run in series and were subtracted from the test titrations. The resulting data were processed and an independent binding model was fit in ITC data analysis software (OriginLab Corporation).

Ligand depletion assays. The plasmid pCBA6TC200 (Malvern Panalytical Ltd, UK) set in a passive mode with a reference power of 4 µcal/s. For isothermal calorimetric titration, TphC concentration was adjusted to either 100 or 500 µM and the ligand solution titrated 1 to 6 in the same stock of Tris buffer as was used during the size exclusion chromatographic purification. The buffer (800 µL/mM protein, 400 µL/mM ligand and 200 µL/mM ligand) solutions were dispensed into a 96-well sample tray and added to the reaction cell and the titrating syringe. The active reaction cell contained 300 µL of protein prior to screening as described above for the apo crystal form. The co-crystallisation of TphC with the ligand was performed using a cubic search volume with 30 Å sides centred on the geometric centre of the protein with an exhaustiveness of 25. The average docking for all ligands in the library is ~6.1 kcal/mol with a standard deviation of 1.02, compared to the scores for the ligands shown in Supplementary Table 6. To illustrate the putative binding mode for biphenyl-4,4′-dicarboxylate 21, which was not successfully docked in the crystal structure, the ligand was first positioned in the active site by aligning to the docked pose of the similarly sized 2-6-naphthalenedicarboxylate 22, after which the system was energy minimised with 1000 steps of steepest descent followed by 10 steps of conjugate gradient in UCSF Chimera. The Amber FF14SB force field55 for the protein the AM1-BCC atom and bonding definitions for the ligand, and biphenyl-4,4′ dicarboxylate was then docked into the resulting protein structure. After energy minimisation, the main change in the protein structure was in the loop formed by residues 30–35, with a resulting RMSD of 0.302 Å relative to the crystal structure for these residues (and 0.095 Å for the whole protein).

Phylogenetic, taxonomic and genomic context analysis. Searching for TphC-like proteins was performed by Basic Local Alignment Search Tool (BLAST) in National Center for Biotechnology Information (NCBI) server using the BLAST-P method. Multiple alignment was performed with Clustal omega, and visualised using iTOL75. Taxonomic information for these sequence files was then retrieved using the NCBI Taxonomy page. If there were multiple entries/species associated with a particular TphC-like protein homologue, then the first NCBI database entry was selected. Genomic context analysis was performed by retrieving the corresponding protein sequences from the genes flanking the tphC-like genes (~7000 nts) using NCBI database. The resulting sequence files were parsed using the biopython and pandas packages57,58 in python and further processed using the tidyverse package R78 (Supplementary Table 7).

Homologous proteins were selected on the basis of NCBI annotation, homology analysis was subsequently performed between the putative homologues and Comamonas strain sp. Es using Clustal omega, and genomic loci containing four or more tph-like genes were classified as tph-like operons. Genes within these operons that could not be classified as tph-like had their function inferred using Blast2GO69. The operon configuration was annotated on basis of the location of the genes encoding TphR and TphC relative to the corresponding intergenic/promoter region. Predominately tphC is located at the 5′ end of the operon and tphR located head-to-head with respect to the corresponding promoter (5′-1′). These data were plotted using R and the tidyverse, gggenes56 and RColorBrewer81 packages. The graphs produced were finalised using GNU Image Manipulation Program82.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The X-ray datasets generated during the current study are available in the Worldwide Protein Data Bank (wwPDB) repository with accession codes 7NDR and 7NDS. The phylogenetic data generated in this study are provided in the Supplementary Data file. Source data are provided within this paper.

Received: 24 March 2021; Accepted: 6 October 2021; Published: 29 October 2021

References
1. Thompson, R. C., Swan, S. H., Moore, C. J., & Vom Saal, F. S. Our plastic. *Philos. Trans. R. Soc. B* **364**, 1973–1976 (2009).
2. Ocean Conservancy. *Building a Clean Swell—International Coastal Cleanup 2018 Report.* https://oceanconservancy.org/wp-content/uploads/2018/07/Building-A-Clean-Swell.pdf (2018).
3. Chen, Y., Awasthi, A. K., Wei, F., Tan, Q., & Li, J. Single-use plastics: production, usage, disposal, and adverse impacts. *Sci. Total Environ.* **752**, 141772 (2021).
4. Blasing, M. & Amend, W. Plastics in soil: analytical methods and possible sources. *Sci. Total Environ.* **612**, 422–435 (2018).
5. Klemes, J. J., Fan, Y. V. & Jiang, P. Plastics: friends or foes? The circularity and plasticity of the UK. *Nature* **2020**, 1–17. https://doi.org/10.1038/s13778-020-01904-9 (2020).
6. Hopewell, J., Dvorak, R. & Kosior, E. Recyclers: challenges and opportunities. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 2115–2126 (2009).
7. PlasticsEurope. *Plastics—the Facts 2019.* https://www.plasticsEurope.org/en/news/market-data (2019).
8. ResearchNester. *Global Polyethylene Terephthalate (PET) Market Analysis & Opportunity Outlook 2027.* https://www.researchnester.com/reports/global-polyethylene-terephthalate-pet-market-analysis-opportunity-outlook-2027/ (2020).

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-021-26508-0 | www.nature.com/naturecommunications
Global PET Bottle Production 2004-2021

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-021-26508-0 | ARTICLE

26. Joo, S. et al. Structural insight into molecular mechanism of poly(ethylene
terephthalate) degradation by a bacterium that degrades and assimilates poly(ethylene
terephthalate). Environ. Lett. 117, 25476–25485 (2020).

27. Skarvini, K. et al. Characterization and engineering of a two-enzyme system for plastics depolymerization. Proc. Natl Acad. Sci. USA 115, 25476–25485 (2020).

28. Austin, H. P. et al. Characterization and engineering of a plastic-degrading aromatic polyesterase. Proc. Natl Acad. Sci. USA 115, E4350–E4357 (2018).

29. Wei, R. et al. Conformationally fitting of a flexible oligomeric substrate does not explain the enzymatic PET degradation. Nat. Commun. 10, 1–4 (2019).

30. Mükcsel, B. et al. Ethylene glycol metabolism by Pseudomonas putida. Appl. Environ. Microbiol. 78, 8531–8539 (2012).

31. Kasai, D. et al. Uncovering the protocatechuate 2,3-cleavage pathway genes. J. Bacteriol. 191, 6758–6768 (2009).

32. Murayama, K. et al. Cloning and characterization of the genes encoding enzymes for the protocatechuate meta-degradation pathway of Pseudomonas aeruginosa NCIB 10272. Biosci. Biotechnol. Biochem. 68, 1434–1441 (2004).

33. Wang, Y. Z., Zhou, Y. & Zhi, G. J. Molecular analysis of isophthalate and terephthalate degradation by comomonas testosteroni YZW-D. Environ. Health Perspect. 103, 9–12 (1995).

34. Satoh, M. et al. Characterization of the terephthalate degradation genes of Comamonas sp. strain E6. Appl. Environ. Microbiol. 72, 1825–1832 (2006).

35. Fukuhara, Y., Kasai, D., Kataya, Y., Fukuda, M. & Masi, E. Enzymatic properties of terephthalate 1,2-dioxigenase of Comamonas sp. strain E6. Biosci. Biotechnol. Biochem. 72, 1–7 (2008).

36. Chong, C. et al. Molecular and biochemical analysis of phthalate and terephthalate degradation by Rhodococcus sp. strain DK17. FEMS Microbiol. Lett. 252, 207–213 (2005).

37. Li, W. J. et al. Laboratory evolution reveals the metabolic and regulatory basis of ethylene glycol metabolism by Pseudomonas putida KT2440. Environ. Microbiol. 21, 3669–3682 (2019).

38. Tso, T. et al. Towards bio-upcycling of polyethylene terephthalate. Metal. Eng. 66, 167–178 (2021).

39. Kang, M. J. et al. A chemo-microbial hybrid process for the production of 2-pyrone-4,6-dicarboxylic acid as a promising bioplastic monomer from PET. Denger, K. et al. Independent periplasmic (TRAP) transporters and tricarboxylic acid transporters (TTR): from uptake to pathogenicity. Front. Cell. Infect. Microbiol. 8, 33 (2018).

40. Roths, L. T., Bianconi, M. E., Thomas, G. H. & Kelly, D. J. Tripartite ATP-independent periplasmic (TRAP) transporters and tricarboxylic acid transporters (TTR): from uptake to pathogenicity. Front. Cell. Infect. Microbiol. 8, 33 (2018).

41. Tiso, T. et al. Towards bio-upcycling of polyethylene terephthalate. Biotechnol. Appl. Environ. Microbiol. 79, 6148–6155 (2013).

42. Rosa, L. T., Bianconi, M. E., Thomas, G. H. & Kelly, D. J. Structural basis for high-affinity adiopate binding to AdpC (P4A515), an orphan periplasmic-binding protein from the tripartite tricarboxylic transporter (TTR) family in Pseudomonas aeruginosa. FEBS Lett. 587, 862–877 (2013).

43. Rosa, L. T., Dix, S. R., Rafferty, J. B. & Kelly, D. J. A new mechanism for high-affinity uptake of C4-dicarboxylates in bacteria revealed by the structure of Rhodopseudomonas palustris MatC (P4A3494), a periplasmic binding protein of the tripartite tricarboxylic transporter (TTR) Family. J. Mol. Biol. 431, 351–367 (2019).

44. Denger, K. & Cook, A. M. Racemase activity effected by two dehydrogenases in sulfolactate degradation by Chromohalobacter sakerensis: purification of (S)- sulfolactate dehydrogenase. Microbiology 156, 967–974 (2010).

45. Wübbeler, J. H. et al. Unravelling the complete genome sequence of Advenella minigardeformis strain DP107 and novel insights in the catalabolism of the xenobiotic polythioester precursor 3,3′-dithiodipropionate. Microbiology 160, 1401–1416 (2014).

46. Herrou, J. et al. Structure-based mechanism of ligand binding for periplasmic solute-binding protein of the bug family. J. Mol. Biol. 373, 954–964 (2007).

47. Huvent, I. et al. Crystal structure of Bordetella pertussis BugD solute receptor unveils the basis of ligand binding in a new family of periplasmic binding proteins. J. Mol. Biol. 356, 1014–1026 (2006).

48. Huvent, I. et al. Structural analysis of Bordetella pertussis BugD solute receptor in a bound conformation. Acta Crystallogr. Sect. D Biol. Crystallogr. 62, 1375–1381 (2006).

49. Salmon, R. C., Cliff, M. J., Rafferty, J. B. & Kelly, D. J. The CouPS1TU and TarPQFM transporters in Rhodopseudomonas palustris: redundant, promiscuous uptake systems of the derived aromatic substrates. PLoS ONE. https://doi.org/10.1371/journal.pone.0059844 (2013).

50. Redhead, M., Satchell, R., McCarthy, C., Pollack, S. & Unitt, J. Thermal shift as a new paradigm for protein stability. Science 318, 1200–1203 (2007).

51. Schepers, G. H., Lycklama a Nijeholt, J. A. & Poolman, B. An updated phylogenetic distribution of plastic-degrading microorganisms. FEMS Microbiol. Lett. 194, 124–131 (2001).

52. Chong, C. et al. Molecular and biochemical analysis of phthalate and terephthalate degradation by Rhodococcus sp. strain DK17. FEMS Microbiol. Lett. 252, 207–213 (2005).

53. Li, W. J. et al. Laboratory evolution reveals the metabolic and regulatory basis of ethylene glycol metabolism by Pseudomonas putida KT2440. Environ. Microbiol. 21, 3669–3682 (2019).

54. Tso, T. et al. Towards bio-upcycling of polyethylene terephthalate. Metal. Eng. 66, 167–178 (2021).

55. Tomás, R. F., Bordado, J. C. M. & Gomes, J. J. P. Xylene oxidation to terephthalic acid: a literature review oriented toward process optimization and development. Chem. Rev. 113, 7421–7469 (2013).

56. Coltro, L., Pitta, J. B. & Madaleno, E. Performance evaluation of new plasticizers for stretch PVC films. Polym. Degrad. Stab. 96, 1541–1543 (2012).

57. Trakhanov, S. et al. Ligand-free and -bound structures of the binding protein (Liv) of the Escherichia coli ABC leucine/isoleucine/valine transport system: trajectory and dynamics of the interdomain rotation and ligand specificity. Biochemistry 50, 4393–4401 (2011).

58. Maimanakos, J. et al. Sequence-based screening for rare enzymes: new insights into the world of AMDas enzymes revealed a conserved motif and 58 novel enzymes clustering in eight distinct families. Front. Microbiol. 7, 1332 (2016).

59. Tomás, R. F., Bordado, J. C. M. & Gomes, J. J. P. Xylene oxidation to terephthalic acid: a literature review oriented toward process optimization and development. Chem. Rev. 113, 7421–7469 (2013).

60. Coltro, L., Pitta, J. B. & Madaleno, E. Performance evaluation of new plasticizers for stretch PVC films. Polym. Test. 32, 272–278 (2013).

61. Ling, C., Shi, S., Hou, W. & Yan, Z. Separation of waste polyester/cotton blended fabrics by photocatalytic degradation and preparation of terephthalic acid. Polym. Degrad. Stab. 161, 157–165 (2019).

62. Sörensen, L. et al. UV degradation of natural and synthetic microplastics for stretch PVC films. Polym. Test. 32, 272–278 (2013).

63. Gambarini, V. et al. Phylogenetic distribution of plastic-degrading microorganisms. mSystems 6, e01112-20 (2021).
Acknowledgements

We would like to thank Dr. Adokiye Berepiku for providing the tphC gene and Dr. Christopher Robinson for technical support with the ITC. We acknowledge Diamond Light Source for time on Beamlines i03 and i04 under Proposal [MX24447-09 and MX24447-19]. We would like to acknowledge Profs. Eiji Masai and Naofumi Kamimura for sharing strains and plasmids for use as controls. This research was supported by EPSRC [EP/M013219/1 and EP/023755/1] and Commonwealth Scholarship Commission (CSC Reference Number INCN-2018-57).

Author contributions

T.G. performed the genetic construct design, cloning, expression, purification, crystallisation of the WT TphC, and initial ITC analysis with TPA. D.D. performed ligand screening for hit identification with DSF, ligand binding validation and ligand competition assays through ITC, designed all TphC point mutants, plus cloned and expressed T242A, Q150S and P86T/Q150S, mutants, and also performed initial phylogenetic analysis. C.L. performed crystallisation, X-ray data collection and structure determination. T.B. performed the phylogenetic analysis. G.A.G performed the depletion assays. P.L.R. created, expressed and performed DSF with the remaining point mutants. L.C. prepared the salts of TPA analogues. K.F performed protein production and purification. T.G. performed the docking analysis. N.D. supervised the study. T.G., D.D and N.D. wrote the manuscript, and all authors contributed to the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-26508-0.

Correspondence and requests for materials should be addressed to Neil Dixon.

Peer review information Nature Communications thanks Bert Poolman, Wolfgang Streit and Ren Wei for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021