Structural basis for antifolate transport by the proton coupled folate transporter PCFT.

Joanne L. Parker[1,*,#], Justin Deme[2,3*], Gabriel Kuteyi[1], Zhiyi Wu[1], Jiangdong Huo[4,5,6]
Ray Owens[4,5,6], Philip C. Biggin[1], Susan Lea[2,3#], Simon Newstead[1,#].

[1]Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, UK; [2]Dunn School of Pathology, University of Oxford, Oxford, OX1 3RE, [3] Central Oxford Structural Molecular Imaging Centre, University of Oxford, South Parks Road, Oxford, OX1 3RE, [4] Structural Biology, The Rosalind Franklin Institute, Harwell Science and Innovation Campus, Didcot, UK, [5]Division of Structural Biology, University of Oxford, The Wellcome Centre for Human Genetics, Headington, Oxford, UK. [6]Protein Production UK, The Research Complex at Harwell, Harwell Science & Innovation Campus, Didcot, UK.

*Equal contribution
#For correspondence:
joanne.parker@bioch.ox.ac.uk;susan.lea@path.ox.ac.uk;simon.newstead@bioch.ox.ac.uk

Folate is an essential vitamin (B9) that plays a critical role in cellular metabolism as the starting point in the synthesis of nucleic acids, amino acids and the universal methylating agent S-adenylsmethionine[1,2]. Folate deficiency is associated with numerous developmental, cardiovascular and neurological disorders[3]. Mammals cannot synthesise folates *de novo*; therefore, several systems have evolved to uptake folates from the diet and distribute these within the body[4]. The proton-coupled folate transporter (PCFT) is responsible for folate uptake across the intestinal brush border membrane and the blood-brain barrier[5] and an important route for antifolate delivery in cancer chemotherapy[6,7]. How PCFT recognizes folates or antifolates however is currently unclear. Here we present the cryo-EM structures of PCFT in both substrate-free and pemetrexed-bound states. Our results provide a structural basis for understanding

antifolate recognition and provide insights into the pH regulated mechanism of cellular folate uptake.

Folate must be acquired from the diet, either through the ingestion of green leafy vegetables, lentils and liver, or more commonly through fortification of wheat with folic acid, a synthetic folate[2]. In 2006 the World Health Organization published guidelines to help countries set targets for folic acid fortification of wheat-based food products to reduce neural tube defects during pregnancy and increase the levels of folic acid in the general population[8]. Folate functions as a coenzyme in one-carbon transfer reactions required for the synthesis of nucleic acids (DNA and RNA) and activated methyl groups that are required for DNA methylation and regulation of chromatin structure[2,9]. Three distinct and essential systems for folate uptake have evolved in humans, which operate through different mechanisms[10]. Folate receptors bind with high affinity to extracellular folates and internalize these within the cell, whereas two integral membrane transport systems, the reduced folate carrier (RFC; SLC19A1) and the proton-coupled folate transporter (PCFT; SLC19A1), function to transport folates into the cell across the plasma membrane (Fig. 1a). Both RFC and PCFT belong to the major facilitator superfamily (MFS) of secondary active transporters. However, whereas RFC operates at neutral pH as a folate cation antiporter[11], PCFT is a folate proton symporter, which functions at acidic pH[12]. Mutations in PCFT [12] cause hereditary folate malabsorption (HFM), a metabolic disorder that results in severe developmental and neurological abnormalities following reduced folate transport into the body[13,14]. The requirement of high levels of folates for rapidly dividing cells has led to several successful anticancer treatments, including methotrexate and pemetrexed[4,6,7,15]. So-called antifolates achieve their mechanism of action by inhibiting enzymes that require folates, such as dihydrofolate reductase and thymidylate synthase, leading to cell cycle arrest[16]. To reduce the cytotoxic effects of antifolates there has been considerable effort in targeting PCFT over the more widely expressed RFC[15]. PCFT is more active in solid tumors due to the increased acidic microenvironment, making this an attractive target for new antifolates[6,16]. However, given the lack of structural information for either RFC or

PCFT, the molecular basis of folate or antifolate recognition and transport remains unclear, hampering efforts in targeted drug delivery.

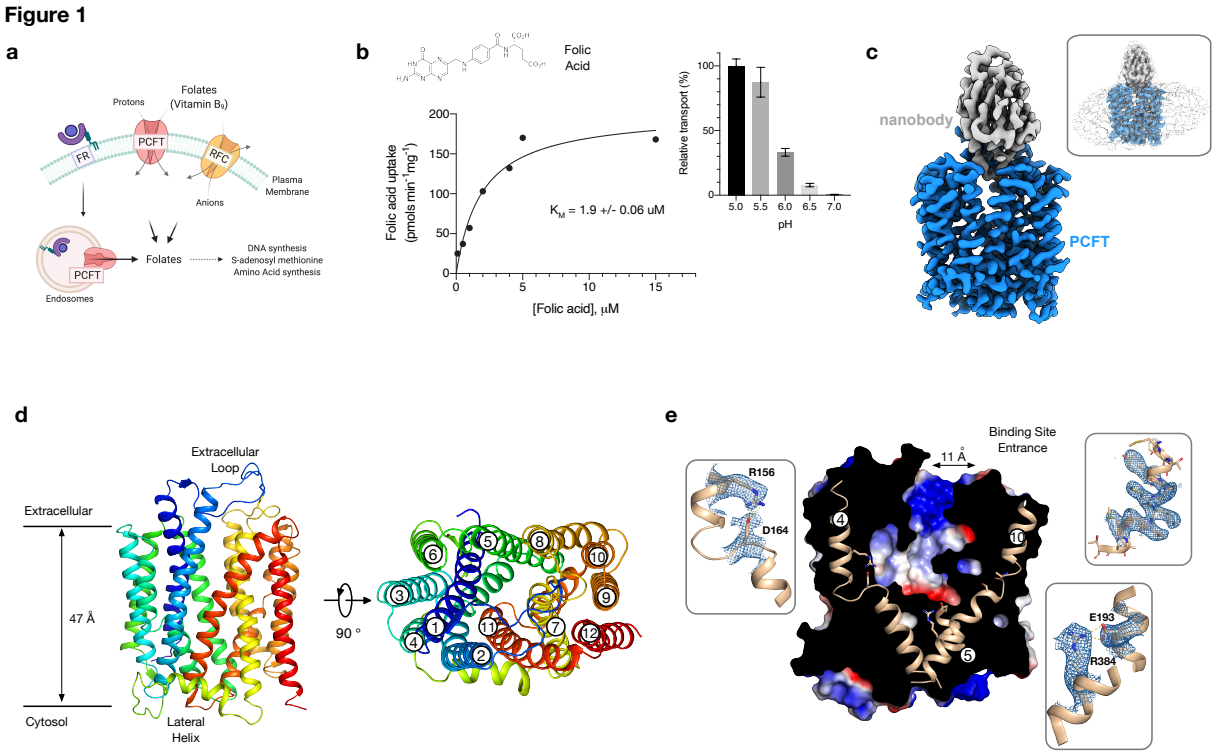


Figure 1. Cryo-EM structure and functional analysis of PCFT. a, Three distinct mechanisms for folate transport exist in mammalian cells. The folate receptor (FR) internalises extracellular folates releasing them in endosomes, whereas PCFT and RFC function to shuttle folates across the plasma membrane. **b,** Functional characterisation of the chicken PCFT reveals identical characteristics to its human homologue. **c,** Cryo-EM density of the PCFT nanobody complex contoured (threshold level of 1.0) to display secondary structure. Inset, superposition with lower contoured (threshold level of 0.22) volume to display detergent micelle. **d,** Cartoon representation of PCFT. **e,** Electrostatic surface representation of PCFT highlighting key structural features. Inset, cryo-EM density for represented side chains (5 sigma).

Cryo-EM structure of PCFT

We identified the chicken PCFT homologue as suitable for structural and biochemical analysis (Extended Data Fig. 1). *Gallus gallus* PCFT displays the characteristic pH-dependent transport of folates and antifolates, with a pH optimum of 5.0-5.5 and a similar K_M for folic acid of 2 μM (Fig. 1b) [12,17]. Attempts at crystallization proved unsuccessful, prompting efforts to obtain nanobodies to aid structural determination. Using a naïve Llama library, a high-affinity inhibitory nanobody with K_D 8 nM was isolated (Extended Data Fig. 2). Although PCFT is only 50kDa, we nevertheless attempted structure determining using single-particle cryo-EM. Following extensive data collection and image processing (Fig. 1c & Extended Data Fig. 3), we successfully determined the atomic structure to 3.3 Å resolution (Extended Data Table 1). The nanobody-PCFT complex is only 62 kDa, demonstrating that 12 kDa nanobodies can also be used as a strategy for the determination of small eukaryotic SLC transporter structures via cryo-EM.

PCFT adopts the canonical MFS fold in an outward open conformation with the nanobody bound at the entrance to the binding site (Fig. 1c, d). In this conformation, TMs1,2 and TMs7,8 splay apart, revealing the entrance to a large polar cavity extending into the interior of the transporter. At the base of this cavity, TM4,5 and TMs10,11 pack together, sealing the binding site from the cytoplasm. These helices form the extracellular and intracellular gates respectively and alternate between open (splayed apart) and closed (packed together) states during transport within members of the MFS[18,19]. Several unique features can be observed from the structure that have important implications for understanding the transport mechanism. A large extracellular loop extends from TM1, reaching over the top of the transporter, before contacting the C-terminal bundle via a disulfide bond[20]. Unbiased molecular dynamics simulations show that in the presence of an antifolate (pemetrexed), discussed below, the loop moves to close the entrance to the large polar cavity that opens out towards the extracellular side of the molecule, and which functions to bind folates (Extended Data Fig. 4). However, breaking the disulfide bond has no measurable effect on transport[20], but does result in destabilization of the transporter (Extended Data Fig. 4d, e), consistent with

a role in stabilizing the protein. The most surprising structural feature is the presence of a substantial discontinuity in TM4 (Fig. 1e). Unusually, the discontinuous region is stabilized through an intrahelical salt-bridge between Arg156 (Arg148 in human PCFT) and Asp164 (Asp156). The discontinuity in TM4 is mirrored in TM10, which although less pronounced functions to create a polar pocket that extends deep within the C-terminal helical bundle and sits close to His289 (His281) on TM7 (Extended Data Fig. 5). His289 is a strictly conserved side chain within the binding site of PCFT, previously identified as playing an important role in proton binding[21]. Molecular dynamics analysis demonstrates this pocket is a stable feature of the structure and is fully solvated in this conformation (Extended Data Fig. 5b) and may function to facilitate the sliding of TM7 against TM10 to facilitate closing of the extracellular gate during transport. Below the water pocket TM10 packs close to TM5, sealing the binding site from the cytoplasmic side of the membrane and stabilized by a conserved salt bridge interaction between Glu193 (Glu185) and Arg384 (Arg376). A recurring structural feature in MFS transporters are salt bridge interactions that function to coordinate transport between the N- and C-terminal bundles[22,23]. These are particularly important for proton-coupled transporters where the protonation state drives the structural rearrangements that coordinate substrate transport across the membrane[24,25]. The presence of an intrahelical salt-bridge in PCFT suggests a novel mechanism within the versatile MFS fold for coupling transport to pH gradients across the membrane, discussed further below. The last notable feature of apo PCFT is the presence of a lateral helix on the cytoplasmic side of the membrane between TM6 and TM7 (Fig. 1b, c). This helix will serve to anchor the transporter in the membrane and resembles a similar structural feature in the human iron efflux transporter ferroportin[26].

Folic acid contains a -2 negative charge at physiological pH due to the presence of glutamic acid, which is attached to the p-aminobenzoate group of the molecule[2] (Fig. 1b). Consistent with this anionic property, the entrance to the transporter binding site contains a distinctive patch of positive charge that would serve to attract folates towards the binding site (Fig. 1e). A similar mechanism of folate capture is observed in the folate receptor [27]. Structural analogues of folate compounds, referred to as folate antimetabolites or simply antifolates, are

important drug molecules for treating a variety of cancers and non-malignant diseases, such as psoriasis and rheumatoid arthritis[16,28]. Clinically relevant antifolates include methotrexate, pemetrexed, raltitrexed and pralatrexate[16], which are transported into cells via both RFC and PCFT[15]. However, solid tumors express high levels of PCFT[29], which becomes the primary carrier for folates due to the increase in acidic environment following the metabolic switch to anaerobic metabolism[7]. Novel antifolates with reduced cytotoxicity are therefore being developed that specifically target uptake by PCFT over RFC[6,15,16,28], but are currently hampered by a lack of structural information.

Structure of PCFT in complex with pemetrexed.

To understand the structural basis of antifolate recognition and transport, we determined the structure of PCFT in complex with pemetrexed (PMX) to 3.4 Å resolution (Fig. 2a, Extended Data Fig. 6 and Extended Data Table 1). Pemetrexed is administered to treat non-small cell lung cancer [30] and is a favored PCFT substrate, with a K_M 0.2 - 0.8 μM[31,32] and IC_{50} of 0.2 μM for the human transporter and 0.9 μM for the chicken homologue (Extended Data Fig. 1). The pemetrexed complex was determined at acidic pH to replicate the functional state of the transporter *in vivo*. Pemetrexed was well resolved in the cryo-EM maps and bound in a kinked conformation at the base of the polar cavity observed in the apo structure (Fig. 2b & c). The γ-carboxylate group of the glutamate occupies a positively charged pocket located within the N-terminal bundle. It interacts with Arg156 (Arg148), part of the intrahelical salt-bridge in TM4, and Asn98 (Asn90) on TM2 (Fig. 1d & e). The β-carboxylate sits close to Val44 (Gly40) on TM1 but makes no specific interaction to side chains in the binding site, while the amide and carbonyl groups both interact with Ser411 (Met403) on TM11. The benzoyl group makes a π-π stacking interaction with Phe290 (Phe282) on TM7, whereas the pyrimidine group sits a second polar cavity constructed from side chains in the C-terminal bundle. In particular, Tyr323 (Tyr315) on TM8 interacts with the carbonyl and amide groups; the NH2 amine group interacts with Glu193 (Glu185) on TM5. Interestingly, the pyrole amine group of pemetrexed interacts with Glu407 on TM11, which is an asparagine in the human homolog (Asn399). Alchemical transformations show that a Glu407Asn mutation would stabilize the protein by forming

hydrogen bonds with Asn166, while remaining capable of preserving the interaction with the pryole amine of pemetrexed (Extended Data Fig. 7).

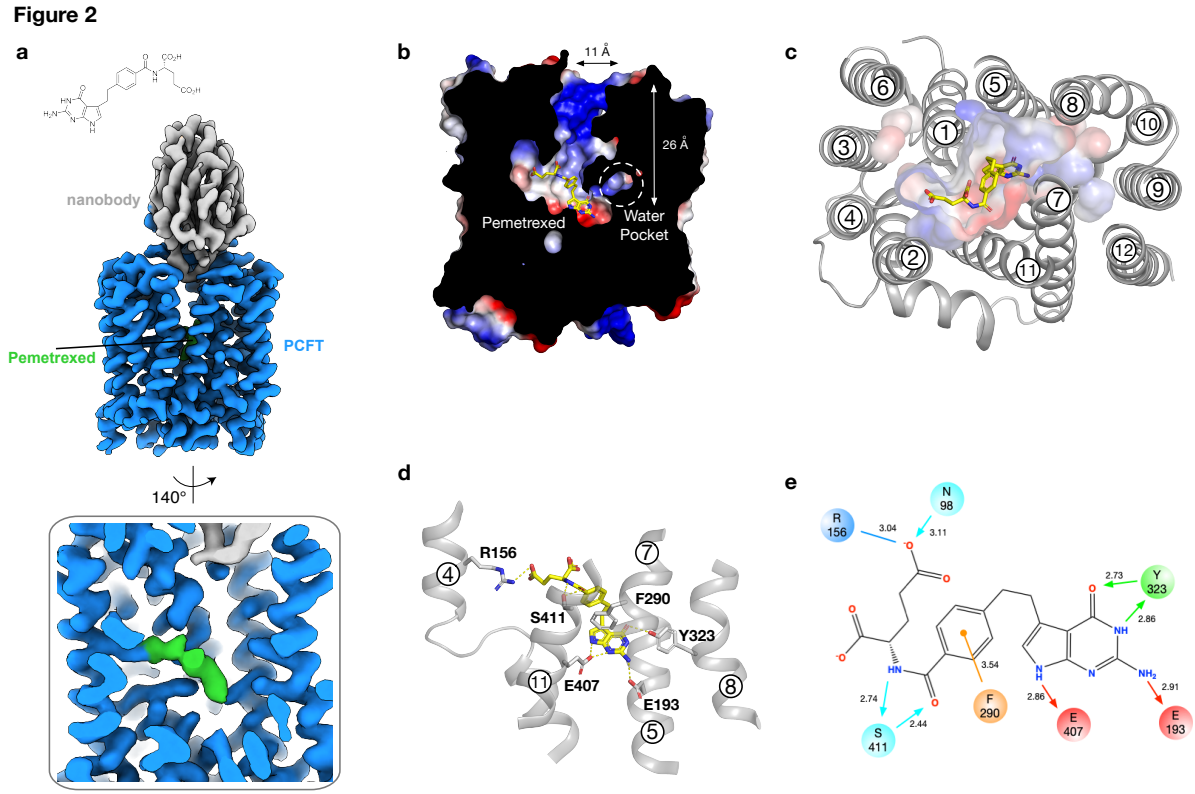


Figure 2. Structure of PCFT in complex with pemetrexed. a, Cryo-EM density of PCFT in complex with pemetrexed and nanobody. **b,** Electrostatic surface representation of PCFT highlighting the central folate binding site, enlarged water pocket and narrow vestibule entrance. The bound pemetrexed molecule shown as sticks. **c,** View of pemetrexed within the binding cavity, viewed 90° relative to **b**, with surface charge highlighted. **d,** Binding site showing the main polar interactions formed with pemetrexed. Key residues interacting with the drug are shown as sticks with hydrogen bonds represented as dashed lines. **e,** Schematic representation of the interactions shown in **d**. Hydrogen bond donors and acceptors are highlighted by directional arrows, the sole charge-charge interaction by a solid line and the pi-pi interaction in orange.

Mechanism of proton coupling

Previous studies have identified essential residues along with missense mutations that cause the genetic disorder hereditary folate malabsorption (HFM) (Extended Data Fig. 1a) [5,10,12]. The structure of PCFT in complex with pemetrexed now allows these mutations to be placed in the context of the binding site and transport mechanism (Extended Data Fig. 8). Of particular interest are Arg384 (TM10) and Asp164 (TM4), which are known HFM mutations[33], and Glu193 (TM5), which was previously reported to play a role in proton coupling[21]. Our structures now reveal that these side chains, together with Arg156, form significant salt bridge interactions that stabilize the transporter in the outward open state (Fig. 1e). Under acidic pH conditions and in the presence of the ligand, we observe repositioning of both arginine side chains, resulting in the breaking of these interactions, thus linking ligand binding to the conformational changes needed to facilitate transport (Fig. 3a). The γ-carboxylate group of pemetrexed functions to replace the interaction of Arg156 with Asp164 at one end of the binding pocket, essentially swapping an intra-protein salt bridge with a protein-ligand salt bridge. At the other end of the binding pocket, the amino group of the nucleotide ring disrupts the interaction between Glu193 and Arg384. These observations strongly suggest a role for both Asp164 and Glu193 in coupling ligand recognition to pH activation and transport of folates and antifolates observed in PCFT.

Molecular dynamics-based free energy calculations estimate the pKa of Asp164 and Glu193 to be 5.6 +/- 0.3 and 6.7 +/- 0.2 respectively in the apo state. Given the working pH of 5.0-5.5, this would suggest that both of these side chains are protonated before pemetrexed binds. To validate the role of Asp164 and Glu193 in the proton coupling mechanism, we analyzed the Asp164Ala, Asp164Asn, Glu193Asn and Glu193Ala variants using an *in vitro* liposome assay to measure transport and a thermal melt assay to assess binding (Fig. 3b, c & Extended Data Fig. 9a, b). The Asp164Ala variant was too unstable and could not be purified, consistent with its role in stabilizing the discontinuous region in TM4. The HFM mutation of this side chain to tyrosine similarly produces unstable protein[33] However, the asparagine variant could still transport [H^3]-folic acid, although at a reduced rate of 45 % of WT levels, identifying

Asp164 as a novel site of proton binding. Molecular dynamics analysis also highlights that protonation of Asp164 reduces the energy required for breaking the interaction with Arg156, following pemetrexed binding (Extended Data Fig. 9c). Unlike Asp164, the Glu193Ala variant was stable but was unable to bind to pemetrexed or folic acid and showed no detectable transport activity. However, we observed the more conservative mutation to asparagine was functional and able to respond to folic acid in both the binding and transport assays, exhibiting uptake ~ 20 % of WT levels. Interestingly, the asparagine variant was functional at both acidic (5.5) and neutral pH (7.5), demonstrating that Glu193 is the principal proton binding site that links proton binding to folate transport in PCFT. Several conserved histidine residues have also been implicated in proton binding, although unlike Glu193, are not essential for transport [21]. His289 is located adjacent to Phe290 on TM7 and makes a hydrogen bond interaction to Asn350 on TM9, stabilizing the extracellular gate in the outward open state (Extended Data Fig. 10). Following pemetrexed binding, His289 rotates away from Asn350, breaking the interaction with TM9. Given the close interaction between the benzoyl group of pemetrexed and Phe290, it is likely that ligand binding promotes the repositioning of TM7 towards TM1, initiating the closure of the extracellular gate of the transporter. This movement would be facilitated by the protonation of His289, a role consistent with our structural data.

Figure 3

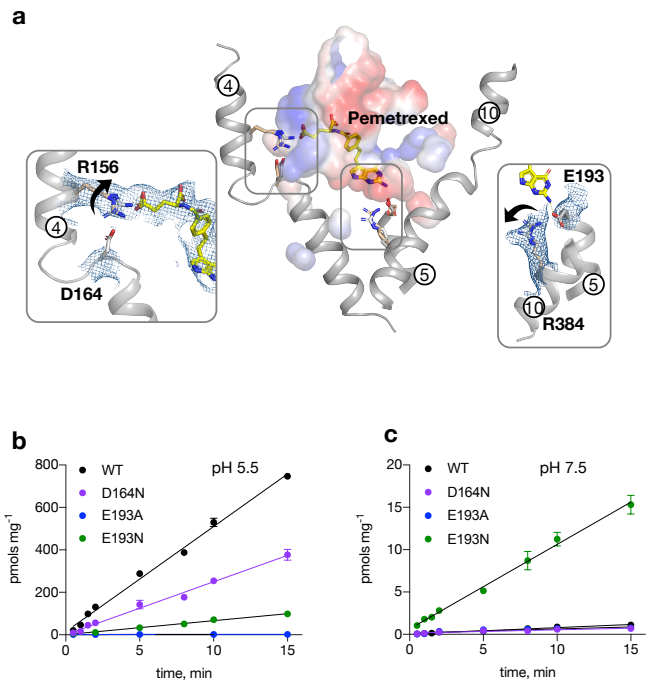


Figure 3. Mechanism of proton coupling. a, Structural comparison of pemetrexed bound and apo structure. Key side chain rearrangements are observed following pemetrexed and proton binding. View is a direct comparison to Fig. 1e. Insets show the cryo-EM density for the two salt bridge interactions which are broken in the presence of substrate. Arrows indicate direction of side chain movement. **b,** Initial transport rates of folic acid uptake liposomes containing wild type (WT) and variant forms of PCFT at pH 5.5. **c,** same experiment as b but performed at pH 7.5 highlight the role of Glu193 in proton coupling.

MFS transporters operate via an alternating access model[18] that requires the coordinated movement of TMs 1-2 and TMs 7-8, which form the extracellular gate with TMs 4-5 and TMs 10-11, that constitute the opposing cytoplasmic gate (Extended Data Fig. 5) [34]. Our data provide a model for proton-coupled folate transport in PCFT that couples proton and ligand binding to structural rearrangements in these gating helices. Specifically, in the outward open state at acidic pH, Asp164, His289 and Glu193 are protonated, and the protein is primed

for transport (Fig. 4a). Binding of the ligand will promote closure of the extracellular gate through the interaction of Phe290 on TM7 and is facilitated by the protonation of His289 and the presence of the water pocket that loosens the packing of the C-terminal bundle helices. Supporting this model, molecular dynamics analysis reveals the movement of the extracellular gate helices in the presence of ligand and the closure of the extracellular loop over the binding site entrance (Extended Data Movie 1). At the cytoplasmic gate, the γ-carboxylate breaks the intrahelical salt bridge in TM4 whereas the amine group of the pyrimidine ring disrupts the interhelical salt bridge between Glu193 and Arg384, which connects TM5 with TM10. The interaction with the amine group would trigger the cytoplasmic half of TM4 and TM5 to open, releasing the folate into the cell. Given the close packing of the lateral helix against TM10 and TM11, it is likely that PCFT operates via a structurally asymmetric movement, with TM4 and TM5 undergoing the larger structural change during transport. Following exposure to the cytoplasmic side of the membrane, Asp164, Glu193 and His289 will deprotonate, favoring reformation of the salt bridge and hydrogen-bond networks and reorientation to the outward-facing state (Fig. 4b). This mechanism explains the extreme pH dependency of PCFT and reveals how pH gradients drive folate transport into the cell.

Pharmacophore model for antifolate transport

Finally, our model for folate transport indicates that ligands must coordinate both Asp164 and Glu193 to be transported, establishing a minimal distance between these two functional groups in the binding site for antifolates (Fig. 4c). The pharmacophore model established by our structures explains previous structure-activity relationship (SAR) data on antifolates [6]. In particular, structures of novel antifolates where 3 or 4 carbon units lengthen the bridge region between the pyrimidine analogue group and the benzyl group exhibit PCFT selectivity over RFC[35,36]. We can now rationalize these observations, as adjustments in the bridging region would serve to extend the γ-carboxylate into a conserved polar pocket in the N-terminal bundle, which contains two conserved glutamine side chains, Gln43 on TM1 and Gln135 on TM3 (Fig. 4d). This would allow the β-carboxylate to replace the interaction with Arg156, satisfying the minimal length requirement for ligand to trigger transport. Substantially

shortening the bridge region or lengthening greater than 3 carbon units resulted in molecules that were not recognized by PCFT[35,36], likely because these molecules no longer engaged both salt bridge networks correctly. Our structure-based pharmacophore thus establishes a framework for further rational drug design of new antifolates with reduced cytotoxicity and selectivity towards PCFT.

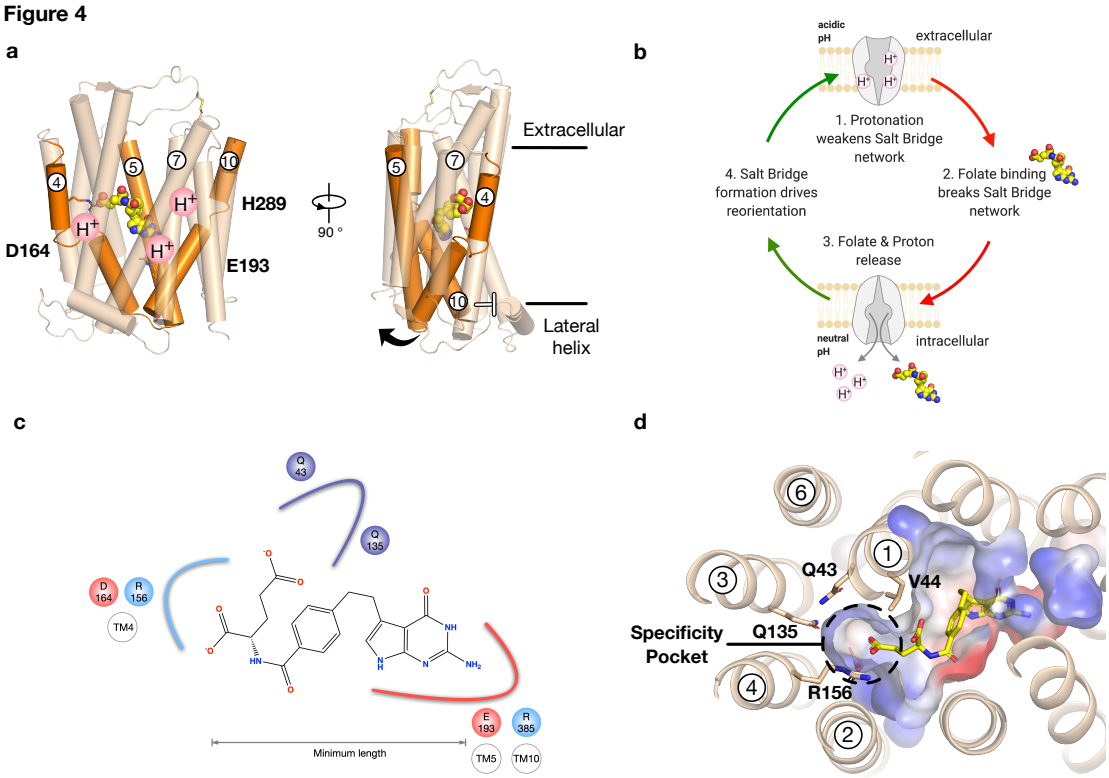
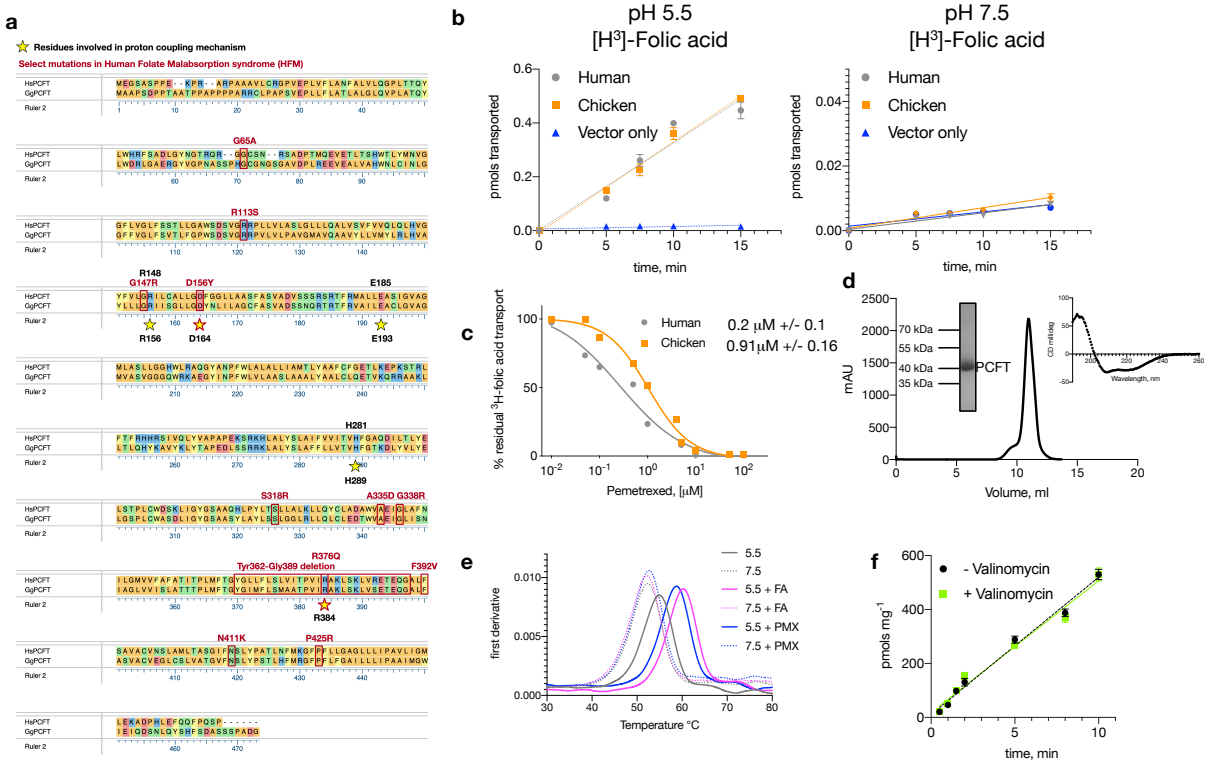


Figure 4. Alternating-access transport mechanism and antifolate binding model. a, Analysis of the gating helices (orange) reveals how proton binding is linked to intracellular gate dynamics. TMs 4 and 5 are likely to undergo the largest structural change to release ligand into the cell, while proton binding at His289 will facilitate closure of the extracellular gate mediated by TM7. **b,** Key steps in the alternating-access model for proton coupled folate transport. **c,** Folates or antifolates must engage both salt bridges at either end of the binding site to open the intracellular gate, establishing a minimum length for substrates. **d,** Polar pocket close to the γ-carboxylate explains how PCFT specific antifolates are recognised.

Extended Data:

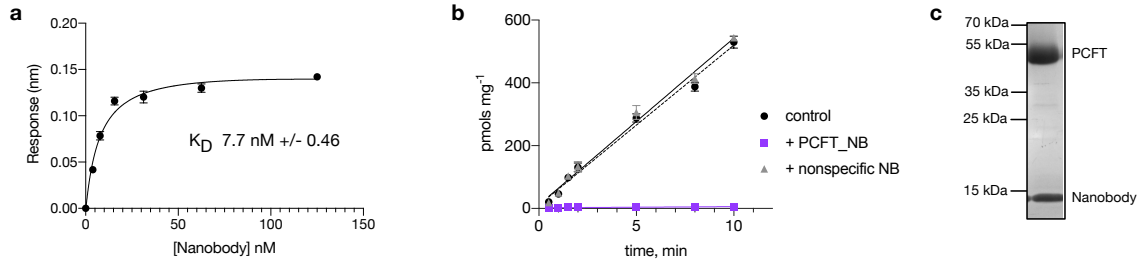
Extended Data Figure 1



Extended Data Figure 1. Comparison of human and chicken PCFT homologues. **a,** Sequence alignment of PCFT from human (Hs, Uniprot Q96NT5) and chicken (Gg, E6Y8U5) coloured via amino acid chemistry. Human and chicken PCFT homologues share an overall 58 % identity and 87 % similarity. Highlighted with yellow stars are functionally relevant residues and in red mutations found causing hereditary folate malabsorption disorder. **b,** Cell based uptake assay comparing the transport of ^{3}H folic acid via human and chicken PCFT at both pH 5.5 and 7.5. The human and chicken homologues transport similar amounts of folic acid to each other and only do so at acidic pH (5.5). **c,** Pemetrexed competition of ^{3}H folic acid uptake into cells, overexpressing either human or chicken PCFT. The calculated mean IC_{50} values are indicated +/- s.d. **d,** Representative gel filtration trace of chicken PCFT in DDM:CHS detergent at pH 6.5. The protein elutes as a monomer, ~ 50 kDa. Insets show coomassie stained SDS-PAGE gel of the purified PCFT protein and circular dichroism analysis. **e,** Analysis of

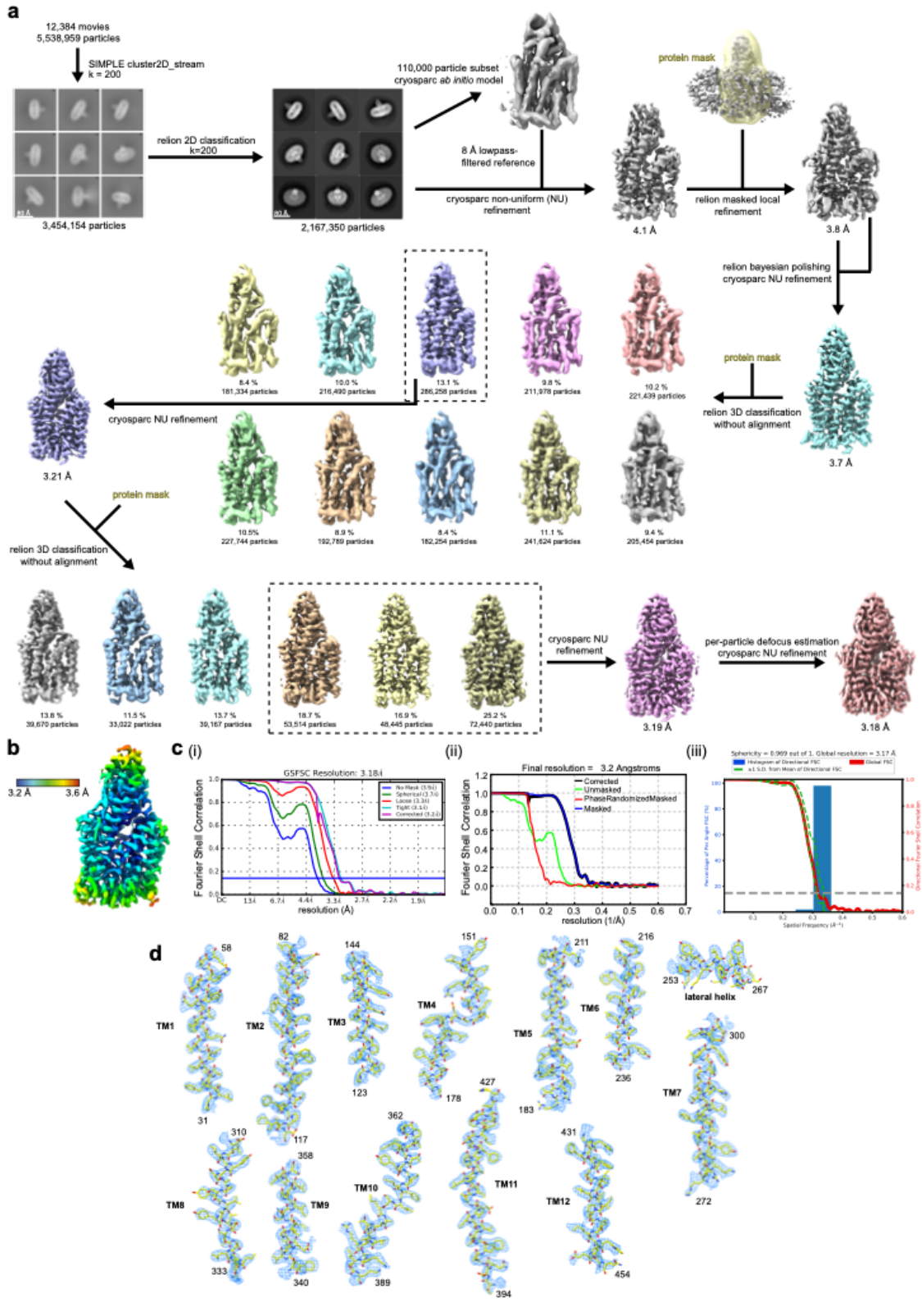
thermostability of PCFT under different pH conditions (7.5 and 5.5) indicate that acidic pH stabilises the protein. The presence of both folic acid (FA) or pemetrexed (PMX) further stabilise the protein but only at acidic pH. **f,** The transport of folic acid by chicken PCFT in a reconstituted liposome uptake assay is not affected by a membrane potential generated by valinomycin (negative inside).

Extended Data Figure 2



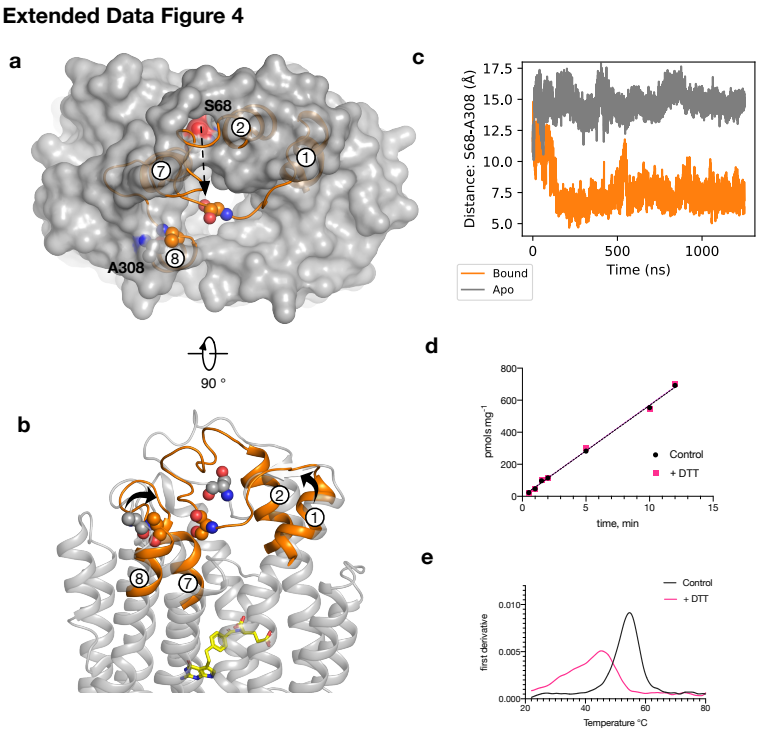
Extended Data Figure 2. Characterisation of high affinity nanobody to PCFT. a, The nanobody identified from a naïve llama library has a K_D of ~ 8 nM for PCFT binding. **b,** The nanobody blocks uptake of folic acid into liposomes containing PCFT where as a non-specific nanobody has no affect. **c,** SDS-PAGE analysis of the PCFT:nanobody complex post size exclusion.

Extended Data Figure 3



297

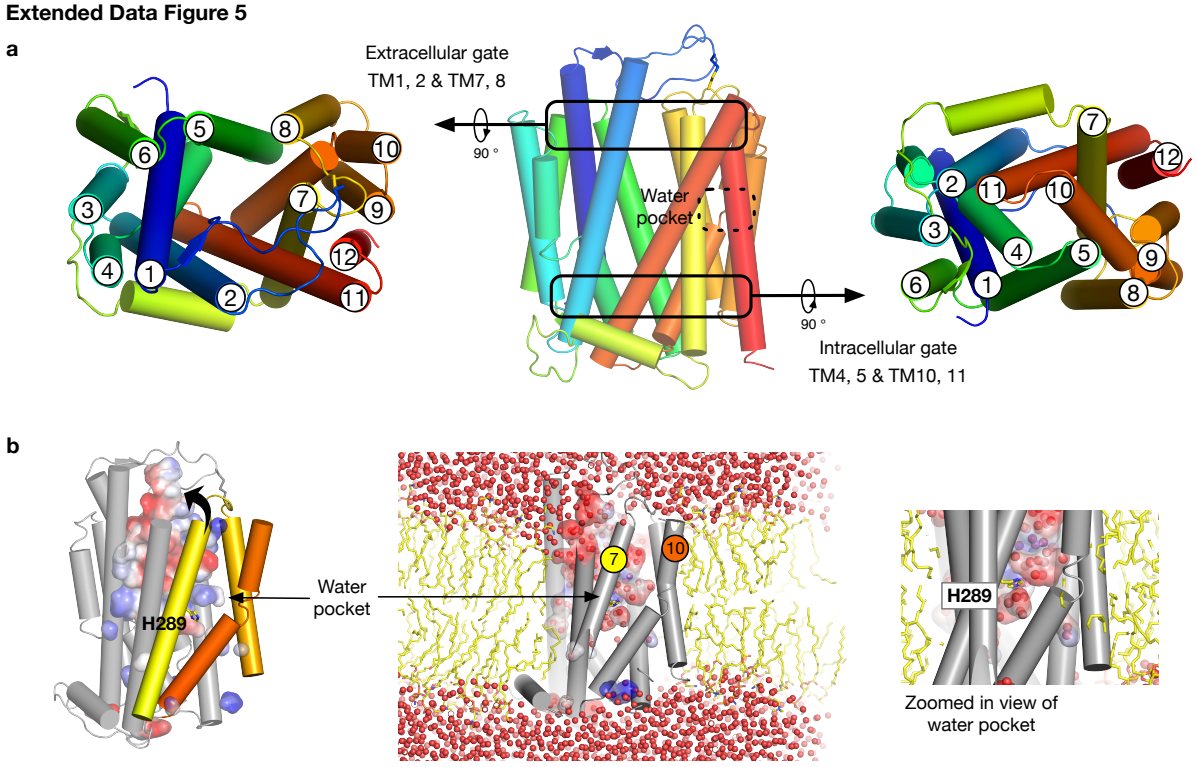
Extended Data Figure 3. Cryo-EM processing workflow and local/global map quality for PCFT-NB. a, Image processing workflow for PCFT-NB. **b,** Local resolution estimation of reconstructed map as determined within RELION. Detergent density omitted for clarity. **c,** Gold-standard Fourier Shell Correlation (FSC) curves used for global resolution estimates within (i) cryoSPARC, (ii) RELION, or (iii) 3DFSC. **d,** Close-up of map/side chain density for transmembrane helices and lateral helix. Volume contoured at threshold level of 0.805.



Extended Data Figure 4. Closure of the TM1/2 loop upon pemetrexed binding. a, In MD simulations of the apo structure (grey), the TM1/TM2 loop stays away from the TM7/TM7 loop, revealing the entrance to the binding pocket. In simulations with pemetrexed bound however (orange), the same loop closes quickly with Ser68 of the TM1/TM2 loop coming into close proximity to Ala308 on the TM7/TM8 loop. **b,** View of a, rotated 90 °. **c,** Time course of the distance between the Cα atoms of Ser68 and Ala308 for both apo (grey line)

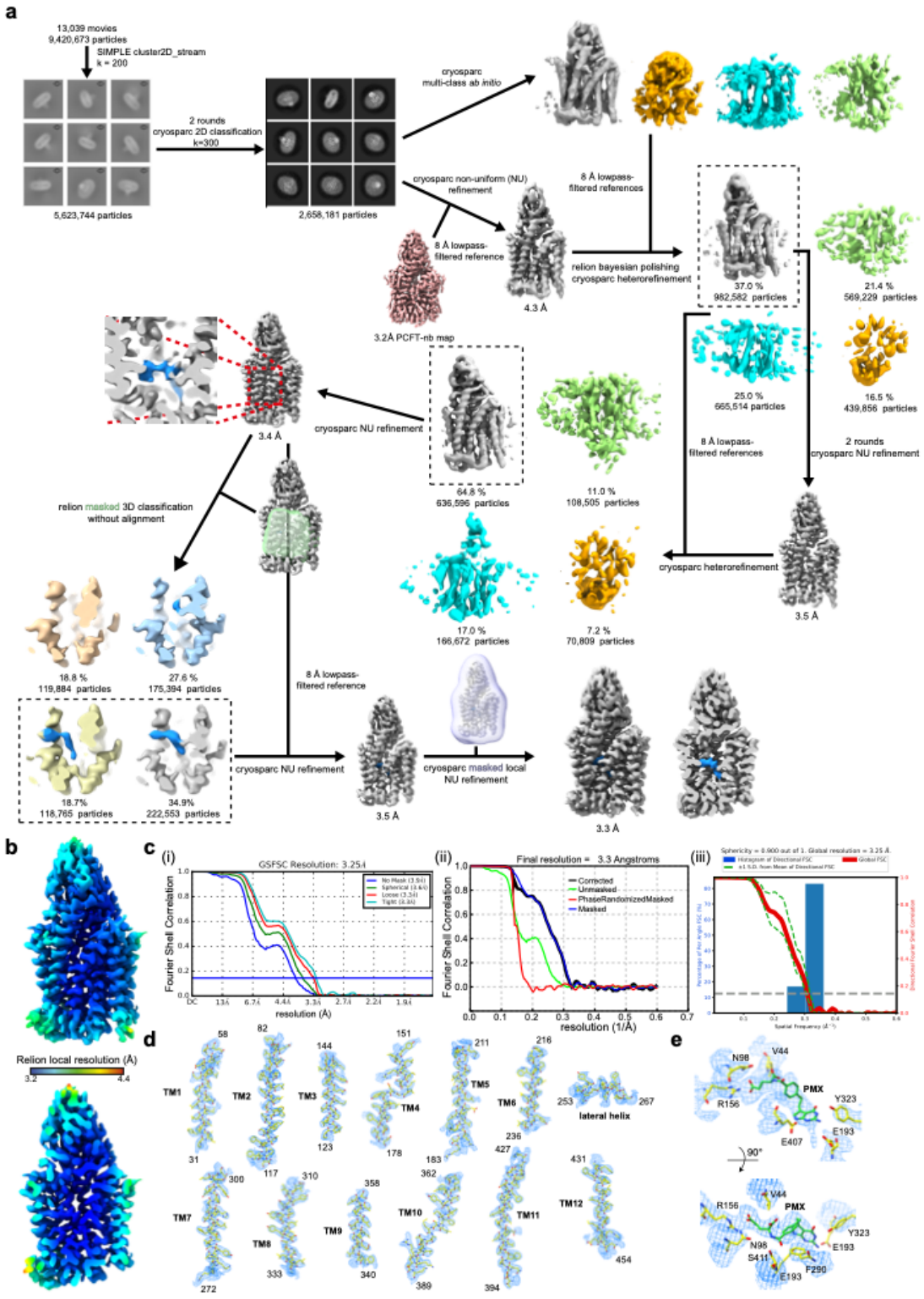
and pemetrexed bound (orange) simulations. The closure event observed with pemetrexed bound occurs around 250 ns and this closed conformation remains stable for the remainder of the simulation. **d,** The presence of 2 mM DTT does not alter the uptake of folic acid into liposomes by PCFT. **e,** The presence of 2 mM DTT does lead to a destabilisation of the protein as determined by differential scanning fluorimetry.

Extended Data Figure 5



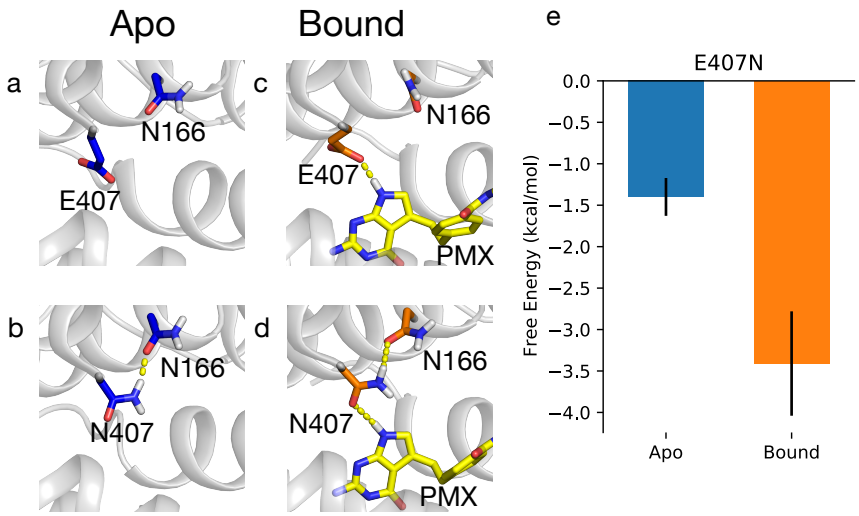
Extended Data Figure 5. Structural relationship between the observed water pocket and the gating helices in PCFT. a, Cartoon representation of PCFT showing the open and closed states of the extracellular and intracellular gates respectively. **b,** Molecular dynamic simulations demonstrate the water pocket is both stable and accessible to bulk solvent via the substrate binding cavity. The space occupied by the polar pocket results in fewer interactions between TM7 and 10 which is likely to facilitate the movement of the helices against one another during transport.

Extended Data Figure 6



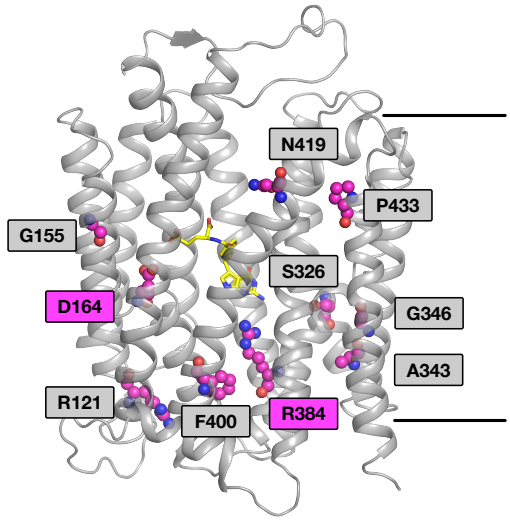
Extended Data Figure 6. Cryo-EM processing workflow and local/global map quality for PCFT-NB bound to pemetrexed. a, Image processing workflow for pemetrexed-bound PCFT-NB. **b,** Local resolution estimation of reconstructed map as determined within RELION. Detergent density omitted for clarity. Top, full map; bottom, central slabthrough map. **c,** Gold-standard Fourier Shell Correlation (FSC) curves used for global resolution estimates within (i) cryoSPARC, (ii) RELION, or (iii) 3DFSC. **d,** Close-up of side chain density for all transmembrane helices. Volume contoured at threshold level of 0.3. **e,** density for pemetrexed (PMX) and interacting PCFT residues. Volume contoured at threshold level of 0.4.

Extended Data Figure 7



Extended Data Figure 7. The effect of E407N on apo and pemetrexed bound PCFT. a, In wildtype Apo state E407 does not interact with N166. The E407N mutant however can readily hydrogen-bond with N166 (**b**), resulting in the E407N mutant being 1.5 kcal/mol more stable (**e**: blue). In the wildtype Apo state, E407 makes a hydrogen bond with the pryole amine group of pemetrexed (PMX) (**c**), which is also preserved in the E407N mutant (**d**). The coordinated hydrogen bind network among N166, E407N and pemetrexed further stabilises the protein-bound state by 2 kcal/mol compared to the apo state (**e**: orange).

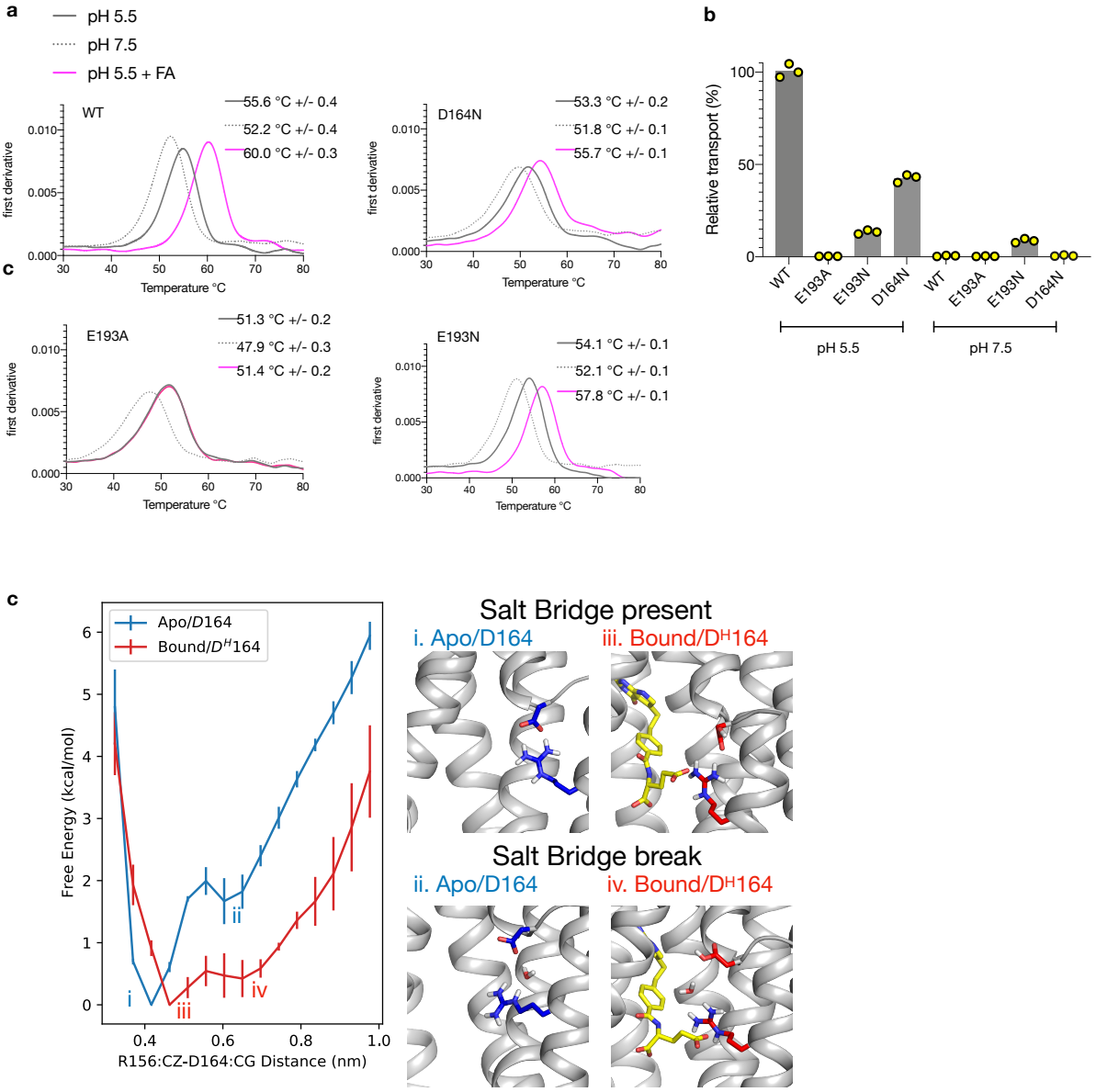
Extended Data Figure 8



Human PCFT	Phenotype	Chicken PCFT
R113C/S	Trafficking impaired	R121
G147R	Impaired stability/no expression	G155
G147R	Impaired stability/no expression	G155
D156Y	Impaired stability/no expression	D164
S318R	Impaired stability/no expression	S326
A335D	Impaired stability/no expression	A343
G338R	Impaired stability/no expression	G346
R376Q/W	Impaired proton binding/reduced transport	R384
F392V	Transport defective	F400
N411K	Impaired transport, 10% WT	N419
P425R	Trafficking impaired	P433

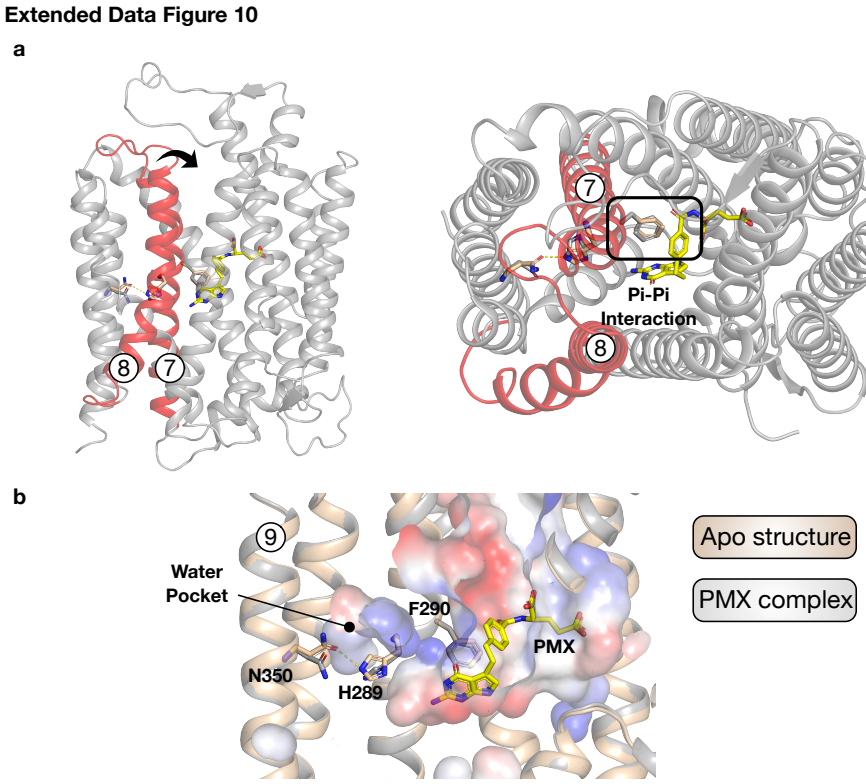
Extended Data Figure 8. Hereditary folate malabsorption mutations. Cartoon representation of chicken PCFT structure with residues involved in hereditary folate malabsorption shown as spheres. Asp164 and Arg384 are highlighted as these residues play an important role in the transport mechanism. The table shows the corresponding human residue number and the associated phenotype.

Extended Data Figure 9



Extended Data Figure 9. Analysis of residues involved in proton binding in PCFT. a, Thermal stability assays using a Prometheus NT4.8 shows that both the Asp164 and Glu193 asparagine variants are stabilised by the presence of folic acid (FA) at pH 5.5 similarly to the WT protein. The Glu193Ala variant however is no longer stabilised by substrate and is also less thermostable than WT protein. The mean melting temperatures calculated are indicated +/- s.d. **b,** Bar chart showing the percentage transport of folic acid into reconstituted liposomes by the variants of PCFT compared to the amount transported by WT at pH 5.5. Data are mean

values after 8 minutes. **c,** Protonation of D164 results in easier breakage of the salt bridge interaction to R156. There is a free energy barrier of 2 kcal/mol associated with the breakage of the R156/D164 salt bridge in the apo state (b>c). The presence of the Pemetrexed as well as the protonation of the D164 ($D^{H}164$) would lower this free energy barrier to a level lower than 0.5 kcal/mol and thus, bring it under the level of thermal fluctuation.



Extended Data Figure 10. Model for ligand induced closure of the extracellular gate. a, Structure of PCFT highlighting the extracellular gate helices, TM7 and 8 (red) and their relationship to the bound pemetrexed molecule (yellow). The arrow indicates the movement required to seal the binding site from the extracellular side of the membrane. The interaction of Phe290 with the benzyl group of pemetrexed is likely to play an important role in triggering gate closure. **b,** Structural comparison between the apo and pemetrexed bound states reveals repositioning of His289 resulting in the breakage of its interaction with Asn350 facilitating the

movement of TM7. The water pocket substantially enlarges in the pemetrexed bound state consistent with greater flexibility in the C-terminal bundle under acidic conditions.

Extended Data Table 1.

Table 1. Cryo-EM data collection, refinement and validation statistics

	PCFT-NB (EMDB-xxxx) (PDB xxxx)	PCFT-NB + pemetrexed (EMDB-xxxx) (PDB xxxx)
Data collection and processing		
Magnification	105,000	105,000
Voltage (kV)	300	300
Electron exposure (e–/Å^2)	59.1	58.5
Defocus range (μm)	0.8 - 2.5	0.8 - 2.5
Pixel size (Å)	0.832	0.832
Symmetry imposed	C1	C1
Initial particle images (no.)	5,538,959	9,420,673
Final particle images (no.)	174,399	341,318
Map resolution (Å)	3.2	3.3
FSC threshold	0.143	0.143
Map resolution range (Å)	3.2-3.6	3.2-4.4
Refinement		
Initial model used (PDB code)	None	XXXX
Model resolution (Ă)	3.2	3.3
FSC threshold	0.143	0.143
Model resolution range (Å)	3.2-3.6	3.2-4.4
Map sharpening *B* factor (Ă^2)	-95.3	-81.8
Model composition		
Non-hydrogen atoms	4178	4209
Protein residues	549	549
Ligands	0	LYA: 1
B factors (Å^2)		
Protein	39.19	55.49
Ligand	N/A	60.24
R.m.s. deviations		
Bond lengths (Å)	0.008	0.008
Bond angles (°)	0.893	0.870
Validation		
MolProbity score	1.95	2.13
Clashscore	9.52	12.19
Poor rotamers (%)	0.92	0.23
Ramachandran plot		
Favored (%)	93.03	90.46
Allowed (%)	6.97	9.36
Disallowed (%)	0.00	0.18

384

Acknowledgments

The Central Oxford Structural Microscopy and Imaging Centre is support by the Wellcome Trust (201536), The EPA Cephalosporin Trust, and a Royal Society/Wolfson Foundation Laboratory Refurbishment Grant (WL160052). Computing was supported via the Advanced Research Computing facility, Oxford, the ARCHER UK National Supercomputing Service and JADE (EP/ P020275/1) granted via the High-End Computing Consortium for Biomolecular Simulation, (HECBioSim - http://www.hecbiosim.ac.uk), supported by EPSRC (EP/L000253/1). This research was supported by Wellcome awards to SML (209194;100298), PCB (219531) and SN (215519;219531) and through MRC grants to SML (MR/M011984/1) and JLP (MR/S021043/1). ZW is a Wellcome Trust PhD student (203741).

Contributions

J.L.P. S.N. conceived the project. G.K. maintained cell stocks, undertook large scale expression and tissue culture. S.N. J.L.P performed all protein preparation. J.H. R.O. screened the Abcore library. J.D S.M.L. performed all cryo-EM sample processing, data collection and image analysis. J.D. S.M.L. S.N. constructed the atomic models. J.L.P. conducted all transport and biochemical assays. Z.W. P.C.B. performed all molecular dynamics analysis. J.L.P. S.N. wrote the manuscript and prepared figures with input from the other authors.

Materials & Methods

Cloning, expression and purification of PCFT. The gene encoding chicken *PCFT* was codon optimised for expression in *S.cerevisiae* and synthesised as a DNA fragment containing homologous recombination sites for insertion into pDDGFP-LEU2d (Addgene 58352) [37]. Wild-type and mutant forms of PCFT were expressed by growing an overnight culture in medium minus leucine with 2 % glucose. This culture was diluted tenfold in medium minus leucine with 2 % lactate in 12 or 24 TunAir flasks and incubated for at least 8 hours at 30 °C at 260 rpm. After this time expression was induced through the addition of 1.5 % galactose from a 25 % (w/v) stock. The yeast was collected after a further 16 h and membranes were prepared. PCFT

and mutants were purified to homogeneity using standard immobilized metal-affinity chromatography protocols in *n*-dodecyl-β-d-maltopyranoside (DDM) (Glycon, DE) detergent with cholesterol hemisuccinate (5:1 ratio DDM:CHS) as described previously. Following TEV cleavage the protein was subjected to size exclusion chromatography (Superdex 200; Cytiva, UK) in a buffer consisting of 20 mM Tris pH 7.5, 150 mM NaCl with 0.03 % DDM and 0.006 % CHS. Biotinyalted PCFT was produced by adding a C- terminal Avi-tag before the TEV cleavage site within the vector. The protein was purified as for WT, following biotintlaytion by Glutathione S-transferase-BirA overnight, the protein was subjected to a further size exclusion run.

Reconstitution into liposomes. PCFT was reconstituted into liposomes consisting of a 3:1 POPE:POPG using biobeads. Chloroform was removed from the lipids (Avanti polar lipids, USA) through the use of a rotary evaporator to obtain a thin film. The lipids were washed twice in pentane and then resuspended at 10 mg ml^{-1} in lipid buffer (50 mM potassium phosphate at pH 6.5). These lipid vesicles were frozen and thawed twice in liquid nitrogen and stored at −80 °C until required. For reconstitution, the lipids were thawed and then extruded first through a 0.8-μm filter and then through a 0.4-μm filter. Purified PCFT in DDM:CHS (at 0.5 μg $μl^{-1}$ concentration) was added to the lipids at a final lipid:protein ratio of 80:1 and incubated for 1 h at room temperature, then for a further 1 h on ice; for the no-protein liposome control, the same volume of gel filtration buffer containing 0.03% DDM, 0.006% CHS was added. After this time, biobeads were added in batches over 24 hours. Biobeads (BioRad, USA) were removed and the proteoliposomes harvested by centrifugation at 120,000 x g for 40 minutes before resuspension in lipid buffer at a final protein concentration of 0.25 μg $μl^{-1}$. They were subjected to three rounds of freeze–thawing in liquid nitrogen before storage at −80 °C. The amount of protein (both wild-type and mutant variants) reconstituted into the lipids was quantified by SDS–PAGE and densitometry. For immunisation into Llamas to raise nanobodies, 0.6 mg of PCFT was reconstituted into POPE:POPG (3:1) ratio at a lipid:protein ratio of 40:1 and resuspended at a final protein concentration of 1 mg/ml using the same method outlined above.

Transport assays. To analyse transport activity, ^{3}H Folic acid (Moravek, USA) was used to monitor uptake and the folic acid was used within 6 months of the activity date due to instability observed of the product over time. Proteoliposomes were thawed and the required amount of proteoliposomes were harvested through centrifugation before resuspending in inside buffer (120 mM NaCl, 2 mM $MgSO_4$ and 20 mM HEPES pH 7.5). The liposomes were subjected to four rounds of freeze thawing in liquid nitrogen to fully distribute the buffer and then extruded through a 0.4-µm filter. To initiate transport the proteoliposomes were diluted into the desired external buffer containing folic acid and trace amounts of ^{3}H Folic acid. The reaction was incubated at 30 °C and stopped by rapidly filtering onto 0.22 µm filters, which were then washed with 2 × 2 ml cold water. The amount of folic acid transported inside the liposomes was calculated by scintillation counting in Ultima Gold (Perkin Elmer, USA) with comparison to a standard curve for the substrate. Experiments were performed a minimal of three times to generate an overall mean and s.d. For the pH optimum experiments 0.5 µg of protein was used per time points and a final concentration of 0.5 µM Folic acid. The external buffers used consisted of 120 mM NaCl, 2 mM $MgSO_4$ and 20 mM cacodylate (pH 5.0) or MES (pH 5.5-6.5) or HEPES 7.0. The data shown in Fig. 1b shows the percent of folic acid transported after 5 minutes as a percentage of the amount at pH 5.0.

The kinetic analysis was performed at pH 5.5 due to the toxicity of cacodylate and also to be able to compare to the literature. The assays used 0.2 µg protein per time point and to allow for the increasing concentration of folic acid and its slight stickiness to the liposomes for each time point a background was removed which was liposomes reconstituted with buffer only. The rate at each concentration was plotted form the linear range of the data and the whole experiment was repeated in triplicate to obtain the K_M +/- s.d. To analyse the effect of a membrane potential the internal buffer used consisted of 120 mM KCl, 2 mM $MgSO_4$ and 20 mM HEPES pH 7.0 and the external buffer was 120 mM NaCl, 2 mM $MgSO_4$ and 20 mM pH 5.5 with or without 1 µM valinomycin. To test the effect on transport of the nanobody a 3x molar excess of nanobody was added to the PCFT liposomes prior to the freeze thaw steps and the liposomes were subjected to 5 rounds of freeze thawing in liquid nitrogen. To act as a

nanobody negative control a nanobody raised against a different protein was used which showed no affinity for PCFT.

Cell based folic acid transport assays. HeLa-R1-11 cells, which lack PCFT and RFC expression and were a generous gift from Prof. I. David Goldman, M.D. (Albert Einstein College of Medicine, NY, USA), were maintained in RPMI 1640 medium supplemented with 10 % foetal bovine serum and 2 mM L-glutamine under 5% CO2 at 37 °C. For transport assays 2 x 10^5 cells per well were seeded into 12 well plates and 24 hours later transfected using lipofectamine 2000 with PCFT constructs in the vector pCDNA3.1 for 36 hours. Cells were washed twice with PBS before application of 0.5 ml assay buffer containing ^{3}H folic acid. The buffer consisted of 120 mM NaCl with either 20 mM MES pH 5.5 or HEPES pH 7.5. After the desired time the assay buffer was removed, and the cells quickly washed twice with 0.5 ml assay buffer with no folic acid. Cells were removed using trypsin, placed in a scintillation vial containing 100 ul 1M NaOH and lysed for 5 minutes prior to the addition of scintillation fluid. The amount of folic acid taken up by the cells was calculated by scintillation counting in Ultima Gold (Perkin Elmer) with comparison to a standard curve for the substrate. Experiments were performed a minimal of three times to generate an overall mean and s.d.

Identification and screening of nanobodies. To identify PCFT specific nanobodies both a naïve llama library (Abcore, USA) and a library raised from immunisation of a Llama with reconstituted PCFT and screened using C-terminally biotinylated protein as described previously [38,39]. ELISA positive clones were sequenced and unique nanobodies identified. These were further characterised using biolayer interferometry using an Octet Red385 (Sartorius, USA) and streptavidin biosensors loaded with biotinyated PCFT at 100 nM in 20 mM Tris pH 7.5, 150 mM NaCl, 0,03% DDM, 0.006% CHS to compare k_{on} and k_{off}. To calculate K_D a serial dilution of the nanobody from 125 nM to 3.9 nM was made and following a 60s baseline step the nanobody was allowed to associate for 300 seconds followed by a dissociation step of 300 seconds. Data were analysed in the Octet v9.0 software package and fit to a single binding site model in Prism. All raw data was baseline and reference subtracted, in-step

corrected, y-axis aligned and filtered with a Savitzky-Golay filter. The nanobodies exhibiting the slowest off rate were analysed for co-elution with PCFT via size exclusion chromatography on a Superdex 200 column at pH 7.5 at 4 °C.

Thermal stability measurements. Thermal stability in the presence and absence of 0.1 mM ligand was analysed using a Prometheus NT.48 (NanoTemper Technologies, DE). The proteins were diluted to a final concentration of 0.3 mg ml^{-1} into buffer containing 20 mM Mes pH 5.5 or 20 mM Hepes pH 7.5, 150 mM NaCl and 0.03% DDM 0.006% CHS before the addition of ligand and incubated on ice for 15 minutes. Thermal measurements were carried out in a range from 20 to 90 °C with 1 °C per min steps. The resulting melting curves were generated by plotting the first derivative of the fluorescence ratio at 330 nm/350 nm against temperature.

Secondary structure analysis. Circular dichroism was used to assess secondary structure of the purified PCFT. Concentrated (20 mg ml^{-1}) protein was diluted to 0.2 mg ml^{-1} into buffer consisting of 10 mM potassium phosphate pH 6.5, 50mM sodium sulphate and 0.03 % DDM:CHS (5:1 ratio). CD spectrum was obtained using a JASCO 815 spectrophotometer (JASCO, JP).

Cryo-EM sample preparation and data acquisition. PCFT post size exclusion was mixed with a 1.2 molar excess of the PCFT nanobody and incubated on ice for at least 30 minutes and the complex separated via size exclusion chromatography as above. For the pemetrexed PCFT NB complex, PCFT was buffer exchanged into 20 mM MES pH 6.0, 150 mM NaCl with 0.03 % DDM and 0.006 % CHS via size exclusion chromatography. The peak fraction was incubated with 0.1 mM pemetrexed for one hour on ice prior to the incubation with a 1.2 molar excess of nanobody before subsequent separation by size exclusion chromatography at pH 6.0.

PCFT with nanobody (6 mg/ml) or pemetrexed-bound PCFT with nanobody (5.8 mg/ml) were adsorbed to glow-discharged holey carbon-coated grids (Quantifoil 300 mesh, Au R1.2/1.3) for 10s. Grids were then blotted for 2s at 100% humidity at 8°C and frozen in liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific). Data were collected in counted super-

resolution mode on a Titan Krios G3 (FEI) operating at 300 kV with a BioQuantum imaging filter (Gatan) and K3 direct detection camera (Gatan) at 105,000x magnification, physical pixel size of 0.832 Å. 12,384 movies were collected for PCFT NB complex at a dose rate of 22.2 e-/Å2/s, exposure time of 2.66 s, corresponding to a total dose of 59.1 e-/Å2over 40 fractions. 13,039 movies were collected for pemetrexed-bound PCFT NB complexat a dose rate of 20.9 e-/Å2/s, exposure time of 2.80 s, corresponding to a total dose of 58.5 e-/Å2over 40 fractions.

Cryo-EM data processing. Initial micrograph processing was performed in real time using the SIMPLE pipeline [40], using SIMPLE-unblur for patched (15 x 10) motion correction, SIMPLE-CTFFIND for patched CTF estimation and SIMPLE-picker for particle picking. After initial 2D classification in SIMPLE to remove junk particles, all subsequent processing was performed in either cryoSPARC [41] or RELION-3.1 [42] using the csparc2star.py script within UCSF pyem [43] to convert between formats. Resolution estimates were derived from gold-standard Fourier shell correlations (FSCs) using the 0.143 criterion as calculated within RELION, cryoSPARC or the remote 3DFSC [44] processing server. Local resolution estimations were calculated within RELION.

For PCFT-NB (Extended Data Fig. 3), 3,454,154 particles were re-extracted in 256 x 256 boxes following initial 2D classification in SIMPLE and subjected to reference-free 2D classification in RELION (200 classes). Selected particles (2,167,350) were used for non-uniform refinement in cryoSPARC against an 8 Å lowpass-filtered *ab initio* model generated from a 110,000 particle subset within cryoSPARC, using an initial lowpass filter of 35 Å, to yield a 4.1 Å map. The corresponding orientation file was imported into RELION using csparc2star.py and focused local refinement was performed using a soft "protein" mask encompassing the transmembrane helices and nanobody, which generated a 3.8 Å volume. Non-uniform refinement in cryoSPARC after Bayesian polishing in RELION (320 x 320 box) improved map quality and resolution to 3.7 Å. A class consisting of 286,258 particles was recovered after focused 3D classification without alignment in RELION using a protein mask; subsequent non-uniform refinement of this particle

548 set in cryoSPARC generated a 3.2 Å map with enhanced side-chain density. Following another
549 focused 3D classification without alignment in RELION, particles from three high resolution
550 classes (174,399 particles total) were subjected to non-uniform refinement within cryoSPARC to
551 generate a 3.2 Å map. B-factors were further improved by per-particle defocus refinement and
552 another round of non-uniform refinement to generate a final map with global resolution of 3.2
553 Å.

554 For pemetrexed-bound PCFT-NB (Extended Data Fig. 4), 5,623,744 particles in 256 x 256
555 boxes were recovered following initial 2D classification in SIMPLE and subjected to 2 rounds
556 of reference-free 2D classification in cryoSPARC (300 classes each round). Selected particles
557 (2,658,181) were used for non-uniform refinement in cryoSPARC against an 8 Å lowpass-
558 filtered PCFT-NB map to yield a 4.3 Å volume. These particles were further Bayesian-polished
559 in RELION (320 x 320 box) and subjected to a multi-class heterorefinement in cryoSPARC using
560 four 8 Å lowpass-filtered *ab initio* models generated from the non-polished particle dataset
561 (starting initial lowpass filter of 35 Å) as input. Particles (982,582) from the only sensible PCFT-
562 NB class were subjected to two rounds of cryoSPARC non-uniform refinement to yield an
563 improved 3.5 Å volume. This particle set was subjected to another round of multi-class
564 heterorefinement using 8 Å lowpass-filtered maps generated from the previous multi-class
565 heterorefinement as references. A class containing 636,596 particles was subjected to one
566 round of non-uniform refinement yielding a 3.4 Å volume with clear density within the
567 transporter cavity. Focused 3D classification without alignment was performed in RELION
568 using a soft mask encompassing the transporter cavity from which two similar but highly
569 occupied classes (totaling 341,318 particles) were selected. This particle set was subjected to
570 non-uniform refinement to generate a 3.5 Å volume with strong pemetrexed density. This
571 density was further improved to 3.3 Å by local non-uniform refinement in cryoSPARC using
572 prior orientations, an 8 Å lowpass-filtered reference, and a soft mask encompassing only
573 protein.

574

Model building and refinement. The atomic model of PCFT-NB (Table 1) was built *de novo* from the globally-sharpened 3.2 Å map following multiple rounds of manual building using Coot v. 0.9[45] and real-space refinement in PHENIX v. 1.18.2-3874 [46] using secondary structure, rotamer and Ramachandran restraints.

The atomic model of pemetrexed-bound PCFT-NB (Table 1) was generated first by rigid body fitting the PCFT-NB model into the globally-sharpened pemetrexed-bound PCFT-NB 3.3 Å map followed by multiple rounds of real-space refinement in Coot and PHENIX. Pemetrexed ligand ("LYA" monomer) was imported into COOT [47] and rigid body fit into map density. Hydrogens were removed from pemetrexed in COOT and corresponding ligand restraint file was generated by eLBOW [48] within the ReadySet! module of PHENIX followed by multiple rounds of real-space refinement in PHENIX and COOT. Both PCFT-NB and PCFT-NB + pemetrexed models were validated using MolProbity [49] within PHENIX. Figures were prepared using UCSF ChimeraX v.1.1 [50], PyMOL v.2.4.0 (The PyMOL Molecular Graphics System, v.2.0; Schrödinger) and BioRender.com (2020).

Molecular Dynamics. The coordinates of PCFT were obtained and the associated antibody was removed. PCFT was embedded into a pure 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membrane via the self-assembly protocol at the coarse-grain level with the Martini 3.0.b.3.2 forcefield as previously described by us [51]. The system was then trimmed to include 90 lipids on both leaflets.

The head group of some of the lipid molecules were then edited to give 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-ethanolamine (POPE) lipids resulting in an overall ratio of POPE:POPC = 3:1 across the entire lipid bilayer system. This was equilibrated for 10 μs before converting to an atomistic representation with cg2at [52]. The protein was represented with Amber14SB [53] and the lipid was represented with the Slipid forcefield [54]. Hydrogens atoms were added to pemetrexed according to [55]. Pemetrexed was represented with the GAFF force-field [56] and the

charge was derived using RESP [57]. The system was then solvated with TIP3P water [58] and neutralised to an ionic concentration of 0.15 M with NaCl.

The MD engine was Gromacs 2020.4 [59]. After energy minimisation, the system was equilibrated with 1000 kj/mol/nm positional restraint for 200 ps with a V-rescale thermostat [60] at 310 K in the NVT ensemble. The system was further equilibrated with the same positional restraint in the NPT ensemble with a berendsen barostat [61] at 1 atm and a Nose-Hoover thermostat at 310 K for 1 ns. Production runs were performed for 100 ns with the Parrinello-Rahman barostat [62]. Five independent repeats were performed and trajectories were clustered via the linkage method. The centre of a cluster was extracted with MDAnalysis [63] and was the starting frame of the following calculations.

Protonation free energy calculation. The free energy of protonating D164 and E193 were calculated for both apo and the pemetrexed-bound state. The topology for the protonation free energy calculation was prepared with pmx [64]. The alchemical transformation was performed via 11 equally spaced lambda windows, where coulombic, van der waals and bonded interactions were changed at the same time with no soft-core. The stochastic dynamics integrator was used for integration and an alchemical chloride ion was positionally restrained to the edge of the box to balance the charge. Each lambda window was equilibrated in the NVT ensemble for 200 ps and in the NPT ensemble for 1 ns followed by a 30 ns production run with replica exchange performed at 1 ps intervals. Three independent repeats were performed and the data was analysed with alchemical-analysis [65].

Alchemical transformation free energy calculation. The free energy of E407N and K293Q mutations were calculated for both apo and the pemetrexed-bound state. The topology for the alchemical transformation was prepared with pmx [64] and then processed by alchemicalitp [66] such that the transformation was done in three stages: i) turning off the partial charges on the atoms that are dummy at state B, ii) transformation of the van der waals and bonded potential and iii) turning on the partial charge on the new atoms in state B. The turning on and off of the

partial charge was performed via 10 equally spaced lambda windows, while van der Waals and bonded interactions were transformed via 20 equally spaced lambda windows. Stochastic dynamics integrator was used for integration and an alchemical chloride ion was positionally restrained to the edge of the box to balance the charge. Each lambda window was equilibrated in the NVT ensemble for 200 ps and then in the NPT ensemble for 1 ns followed by a 30 ns production run with replica exchange performed at 1 ps intervals. Three independent repeats were performed and the data was analysed with alchemical-analysis [65].

Umbrella sampling calculations. The collective variable was defined as the distance between the CZ atom of R156 and CG atom of D164. 15 equally spaced windows were used to separated the collective variable from 0.3 nm to 1 nm and the CV was restrained at 1000 kj/mol/nm. The initial frames were generated using metadynamics with plumed 2.6.2 [67]. Each window underwent 200 ps NVT equilibrium and 1 ns NPT equilibrium before a production run between 150 ns to 250 ns with replica exchange of 1 ps until the results converged. Three independent repeats were performed and the free energy profile was recovered using WHAM [68].

References

1 Zheng, Y. & Cantley, L. C. Toward a better understanding of folate metabolism in health and disease. *J Exp Med* **216**, 253-266, doi:10.1084/jem.20181965 (2019).

2 Bailey, L. B. *Folate in health and disease*. 2nd edn, (Taylor & Francis, 2010).

3 Alam, C., Kondo, M., O'Connor, D. L. & Bendayan, R. Clinical Implications of Folate Transport in the Central Nervous System. *Trends Pharmacol Sci* **41**, 349-361, doi:10.1016/j.tips.2020.02.004 (2020).

4 Hou, Z. & Matherly, L. H. Biology of the major facilitative folate transporters SLC19A1 and SLC46A1. *Current topics in membranes* **73**, 175-204, doi:10.1016/B978-0-12-800223-0.00004-9 (2014).

5 Visentin, M., Diop-Bove, N., Zhao, R. & Goldman, I. D. The intestinal absorption of folates. *Annu Rev Physiol* **76**, 251-274, doi:10.1146/annurev-physiol-020911-153251 (2014).

6 Matherly, L. H., Hou, Z. & Gangjee, A. The promise and challenges of exploiting the proton-coupled folate transporter for selective therapeutic targeting of cancer. *Cancer Chemother Pharmacol* **81**, 1-15, doi:10.1007/s00280-017-3473-8 (2018).

7 Desmoulin, S. K., Hou, Z., Gangjee, A. & Matherly, L. H. The human proton-coupled folate transporter: Biology and therapeutic applications to cancer. *Cancer Biol Ther* **13**, 1355-1373, doi:10.4161/cbt.22020 (2012).

8 de Bree, A., van Dusseldorp, M., Brouwer, I. A., van het Hof, K. H. & Steegers-Theunissen, R. P. Folate intake in Europe: recommended, actual and desired intake. *Eur J Clin Nutr* **51**, 643-660, doi:10.1038/sj.ejcn.1600467 (1997).

9 Lan, X., Field, M. S. & Stover, P. J. Cell cycle regulation of folate-mediated one-carbon metabolism. *Wiley Interdiscip Rev Syst Biol Med* **10**, e1426, doi:10.1002/wsbm.1426 (2018).

10 Zhao, R., Matherly, L. H. & Goldman, I. D. Membrane transporters and folate homeostasis: intestinal absorption and transport into systemic compartments and tissues. *Expert Rev Mol Med* **11**, e4, doi:10.1017/S1462399409000969 (2009).

11 Matherly, L. H. & Hou, Z. Structure and function of the reduced folate carrier a paradigm of a major facilitator superfamily mammalian nutrient transporter. *Vitam Horm* **79**, 145-184, doi:10.1016/S0083-6729(08)00405-6 (2008).

12 Qiu, A. *et al.* Identification of an intestinal folate transporter and the molecular basis for hereditary folate malabsorption. *Cell* **127**, 917-928, doi:10.1016/j.cell.2006.09.041 (2006).

13 Kronn, D. & Goldman, I. D. in *GeneReviews((R))* (eds M. P. Adam *et al.*) (1993).

14 Salojin, K. V. *et al.* A mouse model of hereditary folate malabsorption: deletion of the PCFT gene leads to systemic folate deficiency. *Blood* **117**, 4895-4904, doi:10.1182/blood-2010-04-279653 (2011).

15 Matherly, L. H., Wilson, M. R. & Hou, Z. The major facilitative folate transporters solute carrier 19A1 and solute carrier 46A1: biology and role in antifolate chemotherapy of cancer. *Drug Metab Dispos* **42**, 632-649, doi:10.1124/dmd.113.055723 (2014).

16 Goldman, I. D., Chattopadhyay, S., Zhao, R. & Moran, R. The antifolates: evolution, new agents in the clinic, and how targeting delivery via specific membrane transporters is driving the development of a next generation of folate analogs. *Curr Opin Investig Drugs* **11**, 1409-1423 (2010).

17 Qiu, A. *et al*. Rodent intestinal folate transporters (SLC46A1): secondary structure, functional properties, and response to dietary folate restriction. *Am J Physiol Cell Physiol* **293**, C1669-1678, doi:10.1152/ajpcell.00202.2007 (2007).

18 Drew, D. & Boudker, O. Shared Molecular Mechanisms of Membrane Transporters. *Annu Rev Biochem* **85**, 543-572, doi:10.1146/annurev-biochem-060815-014520 (2016).

19 Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D.-N. Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. *Science (New York, NY)* **301**, 616-620, doi:10.1126/science.1087619 (2003).

20 Zhao, R., Unal, E. S., Shin, D. S. & Goldman, I. D. Membrane Topological Analysis of the Proton-Coupled Folate Transporter (PCFT-SLC46A1) by the Substituted Cysteine Accessibility Method. *Biochemistry* **49**, 2925-2931, doi:10.1021/bi9021439 (2010).

21 Unal, E. S. *et al*. The Functional Roles of the His247 and His281 Residues in Folate and Proton Translocation Mediated by the Human Proton-coupled Folate Transporter SLC46A1. *Journal of Biological Chemistry* **284**, 17846-17857, doi:10.1074/jbc.M109.008060 (2009).

22 Yan, N. Structural Biology of the Major Facilitator Superfamily Transporters. *Annu Rev Biophys* **44**, 257-283, doi:10.1146/annurev-biophys-060414-033901 (2015).

23 Quistgaard, E. M., Low, C., Guettou, F. & Nordlund, P. Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nat Rev Mol Cell Biol* **17**, 123-132, doi:10.1038/nrm.2015.25 (2016).

24 Solcan, N. *et al*. Alternating access mechanism in the POT family of oligopeptide transporters. *The EMBO journal* **31**, 3411-3421, doi:10.1038/emboj.2012.157 (2012).

25 Zhang, X. C., Zhao, Y., Heng, J. & Jiang, D. Energy coupling mechanisms of MFS transporters. *Protein Sci* **24**, 1560-1579, doi:10.1002/pro.2759 (2015).

26 Billesbolle, C. B. *et al*. Structure of hepcidin-bound ferroportin reveals iron homeostatic mechanisms. *Nature* **586**, 807-811, doi:10.1038/s41586-020-2668-z (2020).

27 Chen, C. *et al*. Structural basis for molecular recognition of folic acid by folate receptors. *Nature* **500**, 486-489, doi:10.1038/nature12327 (2013).

28 Visentin, M., Zhao, R. & Goldman, I. D. The antifolates. *Hematol Oncol Clin North Am* **26**, 629-648, ix, doi:10.1016/j.hoc.2012.02.002 (2012).

29 Gonen, N. & Assaraf, Y. G. Antifolates in cancer therapy: structure, activity and mechanisms of drug resistance. *Drug Resist Updat* **15**, 183-210, doi:10.1016/j.drup.2012.07.002 (2012).

30 Newman, A. C. & Maddocks, O. D. K. One-carbon metabolism in cancer. *Br J Cancer* **116**, 1499-1504, doi:10.1038/bjc.2017.118 (2017).

31 Diop-Bove, N. K., Wu, J., Zhao, R., Locker, J. & Goldman, I. D. Hypermethylation of the human proton-coupled folate transporter (SLC46A1) minimal transcriptional regulatory region in an antifolate-resistant HeLa cell line. *Mol Cancer Ther* **8**, 2424-2431, doi:10.1158/1535-7163.MCT-08-0938 (2009).

32 Zhao, R. *et al*. The proton-coupled folate transporter: impact on pemetrexed transport and on antifolates activities compared with the reduced folate carrier. *Mol Pharmacol* **74**, 854-862, doi:10.1124/mol.108.045443 (2008).

33 Zhao, R., Aluri, S. & Goldman, I. D. The proton-coupled folate transporter (PCFT-SLC46A1) and the syndrome of systemic and cerebral folate deficiency of infancy: Hereditary folate malabsorption. *Mol Aspects Med* **53**, 57-72, doi:10.1016/j.mam.2016.09.002 (2017).

34 Fowler, P. W. *et al*. Gating topology of the proton-coupled oligopeptide symporters. *Structure* **23**, 290-301, doi:10.1016/j.str.2014.12.012 (2015).

35 Desmoulin, S. K. *et al*. Targeting the proton-coupled folate transporter for selective delivery of 6-substituted pyrrolo[2,3-d]pyrimidine antifolate inhibitors of de novo purine biosynthesis in the chemotherapy of solid tumors. *Mol Pharmacol* **78**, 577-587, doi:10.1124/mol.110.065896 (2010).

36 Wang, L. *et al*. Synthesis and antitumor activity of a novel series of 6-substituted pyrrolo[2,3-d]pyrimidine thienoyl antifolate inhibitors of purine biosynthesis with selectivity for high affinity folate receptors and the proton-coupled folate transporter over the reduced folate carrier for cellular entry. *J Med Chem* **53**, 1306-1318, doi:10.1021/jm9015729 (2010).

37 Parker, J. L. & Newstead, S. Method to increase the yield of eukaryotic membrane protein expression in Saccharomyces cerevisiae for structural and functional studies. *Protein Sci* **23**, 1309-1314, doi:10.1002/pro.2507 (2014).

38 Huo, J. *et al*. Neutralizing nanobodies bind SARS-CoV-2 spike RBD and block interaction with ACE2. *Nat Struct Mol Biol* **27**, 846-854, doi:10.1038/s41594-020-0469-6 (2020).

39 Pardon, E. *et al*. A general protocol for the generation of Nanobodies for structural biology. *Nat Protoc* **9**, 674-693, doi:10.1038/nprot.2014.039 (2014).

40 Caesar, J. *et al*. SIMPLE 3.0. Stream single-particle cryo-EM analysis in real time. *J Struct Biol X* **4**, 100040, doi:10.1016/j.yjsbx.2020.100040 (2020).

41 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).

42 Zivanov, J. *et al*. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, doi:10.7554/eLife.42166 (2018).

43 Asarnow, D., Palovcak, E. & Cheng, Y. UCSF pyem v0.5. Zenodo. doi:https://doi.org/10.5281/zenodo.3576630 (2019).

44 Tan, Y. Z. *et al*. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods* **14**, 793-796, doi:10.1038/nmeth.4347 (2017).

45 Brown, A. *et al.* Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta crystallographica. Section D, Biological crystallography* **71**, 136-153, doi:10.1107/S1399004714021683 (2015).

46 Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544, doi:10.1107/S2059798318006551 (2018).

47 Emsley, P. Tools for ligand validation in Coot. *Acta Crystallogr D Struct Biol* **73**, 203-210, doi:10.1107/S2059798317003382 (2017).

48 Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta crystallographica. Section D, Biological crystallography* **65**, 1074-1080, doi:10.1107/S0907444909029436 (2009).

49 Prisant, M. G., Williams, C. J., Chen, V. B., Richardson, J. S. & Richardson, D. C. New tools in MolProbity validation: CaBLAM for CryoEM backbone, UnDowser to rethink "waters," and NGL Viewer to recapture online 3D graphics. *Protein Sci* **29**, 315-329, doi:10.1002/pro.3786 (2020).

50 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82, doi:10.1002/pro.3943 (2021).

51 Wu, Z., Newstead, S. & Biggin, P. C. The KDEL Trafficking Receptor Exploits pH to Tune the Strength of an Unusual Short Hydrogen Bond. *bioRxiv*, doi:https://doi.org/10.1101/2020.07.18.209858 (2020).

52 owen vickery, R. C. owenvickery/cg2at: CG2AT2 a fragment based conversion (Version v0.2). *Zenodo*, doi:http://doi.org/10.5281/zenodo.3994618 (2020, August 21).

53 Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **11**, 3696-3713, doi:10.1021/acs.jctc.5b00255 (2015).

54 Jämbeck, J. P. M. & Lyubartsev, A. P. An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes. *Journal of Chemical Theory and Computation* **8**, 2938-2948, doi:10.1021/ct300342n (2012).

55 Petrova, J. *et al*. Molecular simulation of the structure of folate and antifolates at physiological conditions. *Journal of Molecular Graphics and Modelling* **87**, 172-184, doi:https://doi.org/10.1016/j.jmgm.2018.11.018 (2019).

56 Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case, 'Development and testing of a general amber force field'Journal of Computational Chemistry(2004) 25(9) 1157–1174. *Journal of Computational Chemistry* **26**, 114-114, doi:https://doi.org/10.1002/jcc.20145 (2005).

57 Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* **97**, 10269-10280, doi:10.1021/j100142a004 (1993).

58 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935, doi:10.1063/1.445869 (1983).

59 Abraham, M. J. *et al*. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25, doi:https://doi.org/10.1016/j.softx.2015.06.001 (2015).

60 Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *Journal of Chemical Physics* **126**, doi:Artn 014101
10.1063/1.2408420 (2007).

61 Berendsen, H. J., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J Chem Phys* **81**, 3684-3690, doi:10.1063/1.448118 (1984).

62 Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182-7190, doi:10.1063/1.328693 (1981).

63 Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* **32**, 2319-2327, doi:10.1002/jcc.21787 (2011).

64 Gapsys, V., Michielssens, S., Seeliger, D. & de Groot, B. L. pmx: Automated protein structure and topology generation for alchemical perturbations. *Journal of computational chemistry* **36**, 348-354, doi:10.1002/jcc.23804 (2015).

65 Klimovich, P. V., Shirts, M. R. & Mobley, D. L. Guidelines for the analysis of free energy calculations. *Journal of Computer-Aided Molecular Design* **29**, 397-411, doi:10.1007/s10822-015-9840-9 (2015).

66 Wu, Z. *Alchemicalitp: A Gromacs parser for alchemical transformation (Version 0.1)*, 2020, December 16).

67 Bonomi, M. *et al*. Promoting transparency and reproducibility in enhanced molecular simulations. *Nature Methods* **16**, 670-673, doi:10.1038/s41592-019-0506-8 (2019).

68 Grossfield, A. "WHAM: the weighted histogram analysis method", version 2.0.10.2. doi:http://membrane.urmc.rochester.edu/wordpress/?page_id=126.