Mechanism of substrate binding and transport in BASS transporters.

Patrick Becker[1], Fiona B. Naughton[3#], Deborah H. Brotherton[1], Raul Pacheco-Gomez[2], Oliver Beckstein[3], Alexander D. Cameron[1*]

[1]School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, U.K.

[2] Malvern Panalytical Ltd, Malvern, UK.

[3] Department of Physics, Arizona State University, Tempe, AZ 85287

*Correspondence: a.cameron@warwick.ac.uk

[#]Current address: Cardiovascular Reasearch Institute, University of California San Francisco, San Francisco, CA, USA

Abstract

The Bile Acid Sodium Symporter (BASS) family transports a wide array of molecules across membranes, including bile acids in humans, and small metabolites in plants. These transporters, many of which are sodium-coupled, have been shown to use an elevator mechanism of transport, but exactly how substrate binding is coupled to sodium ion binding and transport is not clear. Here we solve the crystal structure at 2.3 Å of a transporter from *Neisseria Meningitidis* ($ASBT_{NM}$) in complex with pantoate, a potential substrate of $ASBT_{NM}$. The BASS family is characterised by two helices that cross-over in the centre of the protein in an arrangement that is intricately held together by two sodium ions. We observe that the pantoate binds, specifically, between the N-termini of two of the opposing helices in this cross-over region. During molecular dynamics simulations the pantoate remains in this position when sodium ions are present but is more mobile in their absence. Comparison of structures in the presence and absence of pantoate demonstrates that pantoate elicits a conformational change in one of the cross-over helices. This modifies the interface between the two domains that move relative to one another to elicit the elevator mechanism. These results have implications, not only for $ASBT_{NM}$ but for the BASS family as a whole and indeed other transporters that work through the elevator mechanism.

Keywords: Membrane transporter, crystal structure, elevator mechanism, bile acid transporter

Introduction

The Bile Acid Sodium Symporter (BASS) family of secondary transporters is synonymous with its founding members, the apical sodium dependent bile acid transporter (ASBT) and the sodium taurocholate cotransporting polypeptide (NTCP) (Geyer et al., 2006). These proteins harness the sodium gradient to transport bile acids across the plasma membranes of enterocytes of the terminal ileum and hepatocytes, respectively. They are both targets of drugs currently in the clinic; ASBT as the target of drugs to alleviate chronic constipation (Karpen et al., 2020; Khanna & Camilleri, 2021) and NTCP as a target for Hepatitis B and D virus entry inhibitors (Wedemeyer et al., 2022). Both proteins also influence drug distribution. The BASS family, however, transports a wide array of substrates other than bile acids. In mammals, the sodium-dependent organic anion transporter (SOAT) transports sulphated steroids (Grosser et al., 2013) and others are putative neurotransmitter transporters (Burger et al., 2011). In plants sodium-coupled BASS transporters transport small metabolites such as pyruvate (Furumoto et al., 2011) and glycolate (South et al., 2017) across the plastidial membrane.

The first detailed structural information on the BASS transporters came though crystal structures of two bacterial transporters, one from *Neisseria Meningitidis* ($ASBT_{NM}$) (Hu et al., 2011) and one from *Yersinia Frederiksinii* ($ASBT_{YF}$) (Zhou et al., 2014). Though neither protein is likely to transport bile acids physiologically, *in vitro* both transporters have been shown to catalyse the sodium dependent transport of the bile acid taurocholate (TCH) and have provided an initial structural framework through which the extensive site-directed mutagenesis studies carried out on ASBT and NTCP could be mapped. The bacterial transporters are built from 10 transmembrane helices, with a 2-fold inverted repeat arranged in two domains, a core domain, and a panel domain (Fig. 1A). The core domain is characterised by two extended helices that cross over at the centre with residues within the extended region contributing to two sodium ion binding sites (Na1 and Na2). The residues forming the sodium binding site in $ASBT_{NM}$ are conserved in many members of the BASS family including ASBT and NTCP (Fig. S1). $ASBT_{NM}$ was crystallised in the presence of TCH and this bile acid is observed to bind in an inward-facing cavity between the core and panel domains in a binding mode that remains stable during molecular dynamics simulations (Alhadeff et al.,

2015). Secondary transporters function by the alternating access mechanism in which conformational changes to the transporter enable the substrate binding site to switch between the opposing sides of the membrane (Beckstein & Naughton, 2022; Drew & Boudker, 2016). The structure of $ASBT_{YF}$ was solved in both an inward-facing state, similar to $ASBT_{NM}$, and an outward-facing state. Based on these structures an elevator type model of transport was proposed (Zhou et al., 2014). In such mechanisms it is expected that the substrate binds to one domain, which moves with respect to another so that the substrate can be carried across the membrane (Drew & Boudker, 2016). In the $ASBT_{NM}$ structure, however, the position of the TCH is not entirely consistent with such a model, as though there are specific interactions only with residues of the core domain, the TCH is not primarily embedded within that domain and is not set as deeply within the cleft as might be expected. This may be partly due to the protein binding to the inward-facing state of the protein where the substrate should be released, but it is also likely that the bile acids do not bind optimally to the bacterial transporters. More recently, structures of human NTCP have been reported (Asami et al., 2022; Goutam et al., 2022; Liu et al., 2022; Park et al., 2022). Though NTCP lacks the first transmembrane helix of the bacterial transporters, the overall fold of the proteins is the same and similar conformations of outward and inward-facing states consistent with an elevator mechanism of transport are observed (Park et al., 2022). The structure of NTCP was also solved with glyco-chenodeoxycholic (Liu et al., 2022). In this structure two molecules of the bile acid are observed binding to the protein with weak interactions to the core domain. The unusual binding modes led the authors of this paper to propose an alternative to the classical alternating access mechanism that involves the binding of two substrates, only one of which is transported in each cycle. To gain further insight into the mechanism of BASS family of transporters, and in particular the 10-transmembrane helix transporters, we therefore sought to find a likely substrate for the bacterial transporters that would enable us to understand how substrates bind.

$ASBT_{NM}$ and $ASBT_{YF}$ have high sequence identity to PanS from *S. enterica* (43% and 83% sequence identity respectively) (Fig. S1). PanS has been implicated in the transport of the coenzyme A precursors, ketopantoate and pantoate (Ernst & Downs, 2015). $ASBT_{NM}$ and $ASBT_{YF}$ are also similar to BASS1 (Fig. S1) and it has been shown that BASS1 from *A. thaliana,*

can transport pantoate, at least *in vitro* (Huang et al., 2018). We therefore decided to investigate whether the coenzyme A precursors would also bind to $ASBT_{NM}$.

Here we demonstrate that pantoate, but not ketopantoate or pantothenate, binds to $ASBT_{NM}$. We solve the crystal structure of the protein in complex with pantoate and show that the pantoate makes specific interactions with residues in the cross-over region of the protein consistent with the elevator mechanism of transport. Molecular dynamics shows that this binding mode is more stable when sodium ions are present in their respective binding sites. Binding of pantoate causes a subtle conformational change within the core region of the protein, which may trigger the more widespread movements of the protein that would enable transport to occur. This suggests a more specific mechanism for $ASBT_{NM}$, much more in line with the classical alternating access model of transport, than has recently been suggested for NTCP.

Results

Pantoate binds to $ASBT_{NM}$

To assess whether pantoate and its derivatives are likely substrates for $ASBT_{NM}$ we first used a dye-based stability assay in which stability is used as a surrogate for binding (Alexandrov et al., 2008). Whereas pantoate stabilised the protein to a similar amount to taurocholate, neither ketopantoate nor pantothenate had any effect under the conditions tested (Fig. 2a). To verify binding and obtain a more reliable estimate of the affinity of pantoate for $ASBT_{NM}$ Isothermal Calorimetry (ITC) was then used, giving a measured K_D of 127 μM (Fig. 2b). Although we were unable to obtain a reliable estimate of the K_D of taurocholate using ITC due to its detergent-like properties, the K_D of pantoate is similar to the K_M reported for TCH (50 μM) (Hu et al., 2011).

Structure of $ASBT_{NM}$ with pantoate

To understand how pantoate binds to $ASBT_{NM}$ we solved the structure of the protein in the presence of pantoate using X-ray crystallography ($ASBT_{NM(Pan)}$). Crystals were grown using the *in-meso* method of crystallisation and the structure was solved by molecular replacement and refined at a resolution of 2.3 Å (Table 1). Density consistent with pantoate, and sodium ions

in both ion binding sites is evident in the resulting maps (Fig. 3B, Fig. S2). There is also evidence of a lipid like molecule within the binding site. The transporter adopts an inward-facing conformation as seen for the structure of $ASBT_{NM}$ with taurocholate present ($ASBT_{NM(TCH)}$; 3ZUX) (Hu et al., 2011) and the two structures can be superposed with a root mean square deviation RMSD of 0.6 Å for 263 out of 308 C_α atoms within 2 Å after superposition (see Methods). The most substantial difference in the conformation of the two structures is seen for TM1. In the $ASBT_{NM(TCH)}$ structure, TM1 bounds one side of the crevice between the panel and core domains (Fig. 1A). In this structure the helix is kinked at residue Thr 14 so that it splays out and enlarges the cavity on the inward-facing side of the protein (Fig. 1A). In the $ASBT_{NM(Pan)}$ structure TM1 is still kinked although the whole helix has moved as an approximate rigid body by ~75° pivoting around Ile 11 such that residues 1 to 10 move over the cytoplasmic entrance to the cavity to partially occlude it from the inward-facing side, and residues 12 to 28 move away from TM10. This creates an opening into the crevice from the membrane between the panel and core domains (Fig. 1B and C, Movie S1).

The pantoate binds between the two cross-over helices TM4b and TM9b (Fig. 3). The carboxylic acid of the pantoate interacts with the main chain nitrogen atoms of Thr 112 and Ala 113 of TM4b and the 2-hydroxyl oxygen is within hydrogen bonding distance of the main chain nitrogen of Gly 267 of TM9b and the amino oxygen of Asn 265. The hydroxyl oxygen of the methyl-propanol moiety also interacts with His 294 and Asn 295, which reside on TM10. These residues are all within the core domain. The closest residues to the pantoate on the panel domain are Ile 203 and Ile 47 that interact with the methyl-propanol moiety. The sodium ion binding sites in $ASBT_{NM}$ are also located at the cross-over region of the two helices, behind the pantoate when viewed from the crevice between the core and panel domains (Fig. 3C). The ions are clearly defined in the electron density (Fig. S2A) and there is very little change in their coordination in the $ASBT_{NM(Pan)}$ structure with respect to that of $ASBT_{NM(TCH)}$ (Fig. S2B).

Within the region of the sodium and pantoate binding sites the most obvious change in the pantoate-bound structure relative to $ASBT_{NM(TCH)}$ is that the main chain nitrogen of Thr 112 is displaced by ~1 Å and the C_γ by 2.4 Å (Fig 3D). In fact, there is a slight movement of the whole of TM4b, which includes the sodium ion ligands, Ser 114 and Asn 115, towards the pantoate

(Fig. 3D, Movie S1). On the panel domain Ile 203, located at the centre of TM7, is also displaced slightly (~1.2 Å) (Fig. 3D and E) enabling the pantoate to be accommodated easily. The conformational change of TM1 may be triggered by this displacement given that Ile 203 would clash with Phe 15 if TM1 had adopted the same conformation as in $ASBT_{NM(TCH)}$ (Fig. 3E).

Structure of $ASBT_{NM}$ without substrate

Given that the subtle conformational changes between the pantoate and taurocholate bound structures would be consistent with mechanistic changes upon substrate binding, with $ASBT_{NM(TCH)}$ representing a non-substrate bound structure, we also solved the structure without taurocholate or pantoate present ($ASBT_{NM(ns)}$) at 2.1Å using the *in-meso* method of crystallisation (Table 1). Overall, $ASBT_{NM(ns)}$ is very similar to $ASBT_{NM(TCH)}$ with an RMSD of 0.5 Å for 293 out of 308 C_α atoms (see Methods) and an almost identical coordination of the sodium ions (Fig 3A-C). There are only two regions where there are slightly larger changes. The first, again, centres on TM1. However, the change is much more subtle than to the pantoate bound structure, pivoting ~15° at Thr 14 (Fig. S3A and B). The second is in the loop between TM5 and TM6, which links the core to the panel domain where the loop takes a conformation more similar to that seen in the $ASBT_{NM(Pan)}$ structure. In flexing between the inward-and outward-facing structures, as reported for $ASBT_{YF}$ (Zhou et al., 2014), this loop changes conformation as it allows the panel to move with respect to the core. Overall, therefore, it appears that the binding of pantoate, rather than either the absence of TCH or the difference in crystallisation method, causes the change in position of TM4b.

Specificity of pantoate binding

To probe the specificity of binding we used two approaches. Firstly, we mutated the two residues for which the side chains are within hydrogen bonding distance of the pantoate in the $ASBT_{NM(Pan)}$ structure and tested the affinity of pantoate for the resultant proteins by ITC. Whereas the mutation of Asn 265 to alanine caused the binding of pantoate to be abolished (Fig. 4A), mutation of Thr 112 to either valine or alanine, surprisingly resulted in an increase in affinity to 86μM (Fig. 4B) and 11μM (Fig. 4C) respectively, though noticeably the latter was entropy driven. Secondly, because much of the molecular recognition involves the main chain

atoms, we used a structure activity relationship approach, testing whether a panel of similar compounds would stabilise the protein (Fig. S4). None of the compounds tested stabilised the protein as much as pantoate, showing the importance of the hydroxy-acetate group, which interacts with the main-chain atoms. Just replacing the hydroxyl oxygen with a ketone as in ketopantoate appears to disrupt binding, likely due to an unfavourable interaction with Asn 265. The only other residues that possess the hydroxy-acetate moiety are isocitrate and D-malate. Both compounds have additional charged groups that may make them less favourable for binding.

Molecular Simulations show greater stability of pantoate when sodium ions are present

To gain insight into the effect of sodium ions on the binding of pantoate, molecular simulations were carried out. Over 0.5 µs of MD simulations, pantoate remains in the crystallographic binding position (Fig 5). With sodium ions present in both the Na1 and Na2 sites the hydrogen bonds between pantoate and the main chain nitrogen atoms of T112 and A113 at the N-terminus of TM4b and G267 at the N-terminus of TM9b remain intact (Fig. 5A) with more fluctuating interactions with the side chain atoms. On the other hand, in the absence of sodium ions the pantoate is more mobile (Fig. 5B, Fig S5A) with the interactions with the main-chain atoms more intermittent (Fig 5A). Over the course of the simulations with the sodium ions, the ions remain stably bound in or close to the Na1 and Na2 sites although the simulations indicate that there is an alternative sodium ion binding position close to the crystallographic Na1 site (labelled Na1* in Fig. 5C). In the simulations without bound sodium ions, ions enter the inward-facing funnel and approach the Na1* binding site, but do not settle into the same binding mode seen in the bound simulations (Fig. S6), although complete binding events may occur on longer time scales.

Discussion

$ASBT_{NM}$ binds pantoate, consistent with this compound being suggested as a substrate for the homologous PanS (Ernst & Downs, 2015) and BASS1 (Huang et al., 2018) proteins. The results from the molecular dynamics simulations demonstrate the importance of the sodium ions in stabilising the binding mode of the pantoate at the N-termini of the cross-over helices (Fig. 5) and might suggest that the ions structure the region in readiness for substrate binding. This

conclusion is supported by the sodium-free wild-type structures of $ASBT_{YF}$, (Wang et al., 2021; Zhou et al., 2014) where the region equivalent to 110-116 in $ASBT_{NM}$ and the pantoate interacting residues Thr 112 and Ala 113 adopt varying positions . While it would appear from the molecular dynamics simulations that the interaction that the pantoate makes with the side chain of Asn 265 is less conserved than those involving the main chain atoms, the mutagenesis studies highlight the importance of the residue in binding. Asn 265 also caps TM4b so could potentially stabilise the structure of the binding site as well as interacting with the substrate.

The pantoate is firmly nestled within the core domain and the only interactions it makes with the panel domain are van der Waals interactions with Ile 47 and Ile 203. These residues are conserved in PanS and $ASBT_{YF}$ but in BASS1 are replaced with a valine and a threonine respectively (Fig. S1). The binding of the pantoate appears to cause a displacement of Ile 203, which in turn displaces Phe 15 on TM1. The novel position of TM1 appears to partly occlude the pantoate in the binding site. It would be tempting to think that the partially occluded conformation we observe here is mechanistic as transporters often go through one or more occluded conformations during their mechanistic cycle (Beckstein & Naughton, 2022). However, while Phe 15 is conserved in PanS, $ASTB_{YF}$ and BASS1, and there is some flexing of this region in $ASBT_{YF}$ as the protein changes conformation from outward to inward-facing (Zhou et al., 2014), there is little conservation at the N-terminus amongst the proteins, suggesting that the position of the N-terminus may not be critical for transport.

Zhou et al have demonstrated that $ASBT_{YF}$ is likely to go through an elevator mechanism as is also observed for other proteins with the same fold (Fang et al., 2021; Lee et al., 2013; Park et al., 2022; Ung et al., 2022). As a sodium-coupled symporter, the pantoate should bind to the outward-facing form of the protein and trigger movement to the inward-facing state where it can be released. Modelling the outward-facing state of the $ASBT_{NM(Pan)}$ structure based on the $ASBT_{YF}$ structure shows that the pantoate could easily be accommodated in the outward-facing form (Fig. S7). The position of the pantoate observed here, therefore, is consistent with an elevator mechanism. It is possible that with the constraints of the membrane, the interaction of Ile 207 with the pantoate as the protein moves to the inward-facing state would trigger the release of the pantoate, rather than resulting in the movement

of TM1. In binding the pantoate the position of Thr 112 moves by 2.4 Å (Movie S2). The position of this residue is intriguing because in morphing between the putative outward and inward states of the protein, Thr 112 comes within 2 Å of Met 48 on the panel domain. The interaction between Thr112 and Met48 may, therefore, block the protein from switching conformations. It can be speculated that the movement of the threonine side-chain may unlock the transporter, allowing it to switch from outward to inward-facing (Fig. 6).

The pantoate binding mode seen for $ASBT_{NM}$ can easily be extrapolated to the plant BASS transporters, which transport similar substrates. In addition to BASS1, which has been shown to transport pantoate *in vitro* as discussed above (Huang et al., 2018), BASS2 transports pyruvate (Furumoto et al., 2011), BASS6 glycolate (South et al., 2017) and BASS5 chain-elongated 2-keto acids (Gigolashvili et al., 2009). The sodium binding sites are conserved throughout these BASS transporters with high conservation within the cross-over regions. Each of these molecules would be able to form hydrogen bonds with the main chain nitrogen atoms of TM4b and TM9b as observed with pantoate. While the pantoate transporter BASS1 contains an asparagine and threonine at the positions of Asn265 and Thr112 respectively, in the other BASS transporters these are replaced by serine and glutamine respectively. This may allow the keto-acids to bind. It would be expected, therefore, that a similar mechanism and binding mode may be seen throughout the BASS transporters.

It also seems plausible that in the human bile acid transporters, the substrate would also form specific interactions with the main chain nitrogen atoms of the cross-over helices. As we observe for pantoate binding to $ASBT_{NM}$, for human ASBT it has been shown that uptake of TCH is abolished when the equivalent of Asn 265 (Asn 266) is mutated to a cysteine (Banerjee et al., 2008) and increases when Thr 112 (Thr 110 in $ASBT_{NM}$) is changed to the same residue (Hussainzada et al., 2008). In NTCP mutation of Asn 262 also abolishes uptake of TCH (Yan et al., 2014). This suggests there may be some similarity in the mechanisms but exactly how these proteins are able to bind to the wide variety of primary, secondary and conjugated bile acids (Grosser et al., 2021) that have been reported as substrates is difficult to say. Liu and co-workers have modelled two bile acids into density that they observe when they solved the structure of human NTCP (Liu et al., 2022). In this structure, while the glycine head group of the bile acid is near to the cross-over helices, the interactions with it are rather weak and the

density associated with this moiety is also rather poorly defined. As the authors of this study have pointed out, it is hard to reconcile this binding mode with the conformational changes associated with an elevator mechanism, which others have demonstrated since that paper was submitted (Park et al., 2022). It may well be that the binding mode observed for NTCP is a non-productive mode and a greater interaction with the residues of the cross-over region, linking the sodium ions with substrate binding, will be required to elicit a conformational change. Given that pantoate does not interact with TM1, it seems unlikely that the absence of this helix in the bile acid transporters would affect the elevator mechanism unduly.

In conclusion the elucidation of pantoate-bound $ASBT_{NM}$ provides new insight into the mechanism of the BASS family of transporters. Pantoate binding to the cross-over region of the sodium-bound protein causes subtle changes to Thr 112 and TM4b. Thr 112 is located in the centre of the protein near to the panel domain and its repositioning could unlock the transporter, enabling it to swap between outward and inward facing states. In the absence of sodium ions these residues are likely to be more flexible, which may enable the transporter to switch from one conformation to another without requiring the conformational change mediated by the substrate. While the exact binding mode of bile acids in the human proteins remains unclear, the high conservation of residues involved in this area suggests that the interaction with residues on the cross-over region may follow a similar mechanism.

Materials and Methods

Expression and Purification

$ASBT_{NM}$ (Hu et al., 2011) was subcloned into a modified version of the expression vector, pWaldo GFPd (Drew et al., 2006) in which the TEV protease site had been altered to a 3C protease recognition site. Site-directed mutations were introduced by PCR (Quikchange II, Agilent Technologies). Cultures were grown in Lemo21 (DE3) cells in PASM-5052 media following the MemStar protocol (Lee et al., 2014). Briefly, the cells were grown at 37 °C with shaking at 200 rpm. At an OD_{600} of 0.5, 0.4mM IPTG and 0.25mM L-rhamnose were added and the temperature was decreased to 25°C for overnight induction. Cell pellets were harvested by centrifugation at 5000 *g* for 15min at 4°C and resuspended in PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) with 1 mM $MgCl_2$, DNaseI, and 0.5 mM Pefabloc (Roche). Cells were lysed by passing them twice through a cell disruptor at a pressure of 25 kpsi. Unbroken cells and cell debris were pelleted by centrifugation at 15,000 *g* for 13 min and the supernatant was subjected to ultracentrifugation at 200,000 *g* at 4°C for 1h to pellet the membranes. Membrane pellets were resuspended in PBS, 15 ml per 1 l of culture, snap frozen in liquid nitrogen, and then stored at -80°C.

Membranes were solublised in 1x PBS, 150 mM NaCl, 10 mM imidazole and 1% (w/v) DDM supplemented with 0.5 mM Pefabloc (Roche) for 2hrs at 4°C. Insolubilized material was removed by centrifugation at 200,000 *g* for 45 min and the supernatant was added to HisPur Ni-NTA superflow agarose , (ThermoFisher) (1ml per 1 mg GFP-tagged protein). The slurry was gently stirred for 3 h to allow binding and then loaded into a glass Econo-Column (Bio-Rad). The column was washed with 10 Column Volumes (CV) of wash buffer (1x PBS, 150 mM NaCl, 0.1% DDM) containing 20 mM imidazole, followed by 10 CV with the imidazole augmented to 30 mM. 3C protease (1:1 stoichiometry with $ASBT_{NM}$-GFP) was added to the resin and cleavage was performed overnight at 4°C. The protein was eluted with 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.03% DDM and passed over a 5-ml HisTrap HP column (GE Healthcare) equilibrated with the same buffer. The flow through was collected and concentrated to 6-10 mg/ml using a 100 kDa molecular weight (MW) cutoff centrifugal concentrator (Sartorius) and loaded onto a Superdex 200 Increase 10/300 GL column equilibrated with 20 mM Tris-HCl (pH

7.5), 150 mM NaCl, 0.03% DDM. Fractions containing protein were pooled together and concentrated to ~ 25 mg/ml as above.

Protein Crystallisation and Structure Solution

$ASBT_{NM(Pan)}$ Crystals were grown using the lipidic cubic phase method (Caffrey & Cherezov, 2009). The protein was mixed with monoolein at 60:40 (w/w) ratio using a coupled syringe device (SPT Labtech) and crystallisation trials were set up at 20°C using glass sandwich plates using a Mosquito Robot (SPT Labtech). The protein was preincubated with 1mM pantoate for 30 mins at room temperature. Crystals appeared after 1 week. Crystals were harvested from MemGold2 (Molecular Dimensions) condition A1, (0.2 M Magnesium chloride hexahydrate, 0.005 M Cadmium chloride hemi-(pentahydrate), 0.1 M Tris (pH 7.5), and 14% v/v PEG 500 MME). Crystals were harvested into MicroMounts (MiTeGen) and snap-cooled in liquid nitrogen. X-ray diffraction data were collected at I24 at Diamond Light Source. Diffraction images were integrated and scaled using DIALS (Waterman et al., 2016) with further processing in CCP4 (Collaborative Computational Project Number 4, 1994). The structures were solved by molecular replacement in Phaser (McCoy et al., 2007) through the Phenix suite of programs (Adams et al., 2010) from a model derived from the deposited structure of $ASBT_{NM}$ (3zuy) that had been crystallised by vapour diffusion (Hu et al., 2011). Refinement was performed in Phenix.refine (Afonine et al., 2012) interspersed with manual rebuilding in Coot (Emsley & Cowtan, 2004). Pantoate and sodium ions were built into clear density in the maps. Lipids and metal ions were tentatively assigned to other features in these maps. Given that both structures contained metal ions from the crystallisation or purification the structures were refined against I^+/I^-.

$ASBT_{NM(ns)}$

Using the *in-meso* method of crystallisation as above, several structures were solved where taurocholate was not added to the crystallisation mixture. The highest resolution data were obtained from a single crystal harvested from condition C3 of the MemMeso screen (Molecular Dimensions) with 0.1 M Sodium chloride 0.1 M HEPES 7, 30 % v/v PEG 300, and 0.1 M Calcium chloride dihydrate. The drop also contained (4R-cis)-1-[4-[4-[3,3-Dibutyl-7-(dimethylamino)-2,3,4,5-tetrahydro-4-hydroxy-1,1-dioxido-1-benzothiepin-5-yl]-

phenoxy]butyl]-4-aza-1-azoniabicyclo[2.2.2]octane Methanesulfonate, dissolved in dimethyl sulphoxide (DMSO). This is an inhibitor of human ASBT (Huang et al., 2005) that did not show any effect in our stability assays with $ASBT_{NM}$. The data were processed as above. As the ASBT inhibitor could not be observed in the electron density maps and the resultant structure was consistent with lower resolution structures where this compound was not added we treat this as a good representative of the non-substrate bound structures. Density, present in the cavity was modelled as monoolein (Fig 3D).

Superpositions were performed in Chimera (Pettersen et al., 2004) and structural images were prepared in PyMol (Delano, 2002). Images involving electron density were prepared in CCP4mg (McNicholas et al., 2011) Movies were made with Chimera.

Stability Assay

Screening of compounds for binding was carried out using a stability assay based on binding of 7-Diethylamino-3-(4'-Maleimidylphenyl)-4-Methylcoumarin (CPM) to the protein (Alexandrov et al., 2008; Sonoda et al., 2011). CPM (ThermoFisher) was dissolved in DMSO to a final concentration of 4 mg/ml. The assay was performed in 0.2 ml non-skirted low profile 96-well PCR plates (ThermoFisher). 50 µl of protein (2.5 µg in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.03% DDM) was added to each well supplemented with 1 µl (final concentration 1mM) of each of the compounds of interest and the plate incubated for 30 min at room temperature. The CPM dye was diluted 1:100 in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.03% DDM and 2.5 µl of the diluted dye was added to each well. The assay was performed using a Stratagene Mx3005P Real-Time PCR machine (Strategene) and samples were heated from 25°C to 95°C in 1°C/min steps. Data were analyzed using GraphPad Prism.

Isothermal calorimetry

The protein sample was dialyzed overnight at 4 °C against 20 mM Tris-HCl (pH 7.5), 150 mM NaCl and 0.03% DDM and centrifuged at 16,000 g at 4 °C for 30 min. Isothermal calorimetry (ITC) experiments were performed on a MicroCal PEAQ-ITC (Malvern Panalytical, UK). The protein solution (220 µM) was filled into the sample cell and the pantoate solution (5mM in the dialysis buffer) into the syringe. The cell temperature was set to 10°C with a stirring speed

of 750 rpm and a reference power of 10 µcal/sec. 20 injections were performed with an initial delay of 250 s. The initial injection was performed for 0.8 s with an injection volume of 0.4 µl. The later injections were performed for 4 s with an injection volume of 2 µl. 180 s spacing was left between each injection. The data were analysed using the "one set of sites" model within the MicroCal PEAQ-ITC software (Malvern) iterated using the Lavenberg-Marquardt algorithm after subtraction of the control experiment (pantoate titrated into buffer). The thermodynamic and binding parameters were derived from the nonlinear least squares fit to the binding isotherm.

Molecular Dynamics Simulations

The pantoate-bound ASBT structure was embedded in in a 80:20 POPE:POPG bilayer and solvated with neutralizing ions (0.15 M NaCl) to a final box size of 9.1 x 9.1 x 9.6 nm^3 using CHARMM-GUI (Jo et al., 2008; Lee et al., 2016; Wu et al., 2014). An initial structure was generated without bound sodium. Two sodium ions were moved back to the binding sites manually to generate a sodium-bound initial structure.

From each starting structure, simulations were performed using Gromacs 2018.6 (Abraham et al., 2015) with the CHARMM-36 forcefield (Best et al., 2012) and TIP3 water. Parameters for pantoate were generated using CGenFF (Vanommeslaeghe et al., 2010) and converted to a Gromacs format using the cgenff_charmm2gmx.py script. Energy minimization and 5 ns multi-step equilibration were performed following the CHARMM-GUI protocol, followed by three 500 ns production runs using different initial velocities. Simulation timestep was 2 fs; temperature and pressure were maintained using v-rescale (at 303.15 K) and a Parrinello-Rahman semi-isotropic barostat (Parrinello & Rahman, 1981) (at 1 atm), respectively. The Particle-Mesh Ewald method (Darden et al., 1993) was used for long-range electrostatic interactions, and non-bonded interactions were reduced from 1 nm to a 1.2 nm cut-off using potential-shift. Hydrogen bonds were restrained using the LINCS algorithm (Hess et al., 1997). Three repeats of 500 ns were carried out for both the structures with no sodium bound (-Na^+) and sodium ions bound in both the Na1 and Na2 sites (+ Na^+). All simulation analysis was performed using MDAnalysis (Gowers et al., 2016; Michaud-Agrawal et al., 2011). For

hydrogen bond analysis, a 3.5Å distance and 145° angle cut-off were used. Visualisations of structures were made using VMD (Humphrey et al., 1996).

Acknowledgments: We thank the staff at beamline I24 at DLS and the technical support in the School of Life Sciences, University of Warwick. Funding: The project was supported by the National Institutes of Health (Award No R01GM118772). DHB was supported by MRC (MR/P010393/1).

Author contributions: The project was initiated and supervised by AC and OB. Cloning, expression, purification, crystallisation and assays were carried out by PB, with initial help from DB. ITC was carried out by RPG. Data collection, processing and structural analysis were carried out by PB and AC. Simulations were carried out by FN. AC wrote the manuscript with contributions from all authors.

Competing interests: Authors declare no competing interests.

Data and materials availability: Data and coordinates have been deposited in the RCSB Protein Data Bank under accession numbers 8OYG ($ASBT_{NM(Pan)}$) and 8OYF ($ASBT_{NM(ns)}$). MD trajectories in the GROMACS XTC format were deposited in the OSF.io repository under DOI 10.17605/OSF.IO/KFDT5 under the open CC-BY Attribution 4.0 International license. The trajectories contain all atoms and were subsampled at 5-ns intervals. GROMACS run input files (TPR format) and initial coordinate files (GRO format) together with topology files (GROMACS format) are also included.

References

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1-2*, 19-25. https://doi.org/10.1016/j.softx.2015.06.001

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D*, *66*(2), 213-221. https://doi.org/doi:10.1107/S0907444909052925

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., & Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr*, *68*(Pt 4), 352-367. https://doi.org/10.1107/S0907444912001308

Alexandrov, A. I., Mileni, M., Chien, E. Y., Hanson, M. A., & Stevens, R. C. (2008). Microscale fluorescent thermal stability assay for membrane proteins. *Structure*, *16*(3), 351-359. https://doi.org/10.1016/j.str.2008.02.004

Alhadeff, R., Ganoth, A., & Arkin, I. T. (2015). Mechanistic studies of the apical sodium-dependent bile acid transporter. *Proteins*, *83*(6), 1107-1117. https://doi.org/10.1002/prot.24796

Asami, J., Kimura, K. T., Fujita-Fujiharu, Y., Ishida, H., Zhang, Z., Nomura, Y., Liu, K., Uemura, T., Sato, Y., Ono, M., Yamamoto, M., Noda, T., Shigematsu, H., Drew, D., Iwata, S., Shimizu, T., Nomura, N., & Ohto, U. (2022). Structure of the bile acid transporter and HBV receptor NTCP. *Nature*, *606*(7916), 1021-1026. https://doi.org/10.1038/s41586-022-04845-4

Banerjee, A., Hussainzada, N., Khandelwal, A., & Swaan, P. W. (2008). Electrostatic and potential cation-pi forces may guide the interaction of extracellular loop III with Na+ and

bile acids for human apical Na+-dependent bile acid transporter. *Biochem J*, *410*(2), 391-400. https://doi.org/10.1042/BJ20071300

Beckstein, O., & Naughton, F. (2022). General principles of secondary active transporter function. *Biophys Rev (Melville)*, *3*(1), 011307. https://doi.org/10.1063/5.0047967

Best, R. B., Zhu, X., Shim, J., Lopes, P. E., Mittal, J., Feig, M., & Mackerell, A. D., Jr. (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J Chem Theory Comput*, *8*(9), 3257-3273. https://doi.org/10.1021/ct300400x

Burger, S., Doring, B., Hardt, M., Beuerlein, K., Gerstberger, R., & Geyer, J. (2011). Co-expression studies of the orphan carrier protein Slc10a4 and the vesicular carriers VAChT and VMAT2 in the rat central and peripheral nervous system. *Neuroscience*, *193*, 109-121. https://doi.org/10.1016/j.neuroscience.2011.06.068

Caffrey, M., & Cherezov, V. (2009). Crystallizing membrane proteins using lipidic mesophases. *Nat Protoc*, *4*(5), 706-731. https://doi.org/10.1038/nprot.2009.31

Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D*, *50*(Pt 5), 760-763. https://doi.org/10.1107/S0907444994003112

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: AnN·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, *98*(12), 10089-10092. https://doi.org/10.1063/1.464397

Delano, W. L. (2002). The PyMOL Molecular Graphics System. *DeLano Scientific, Palo Alto, CA, USA*.

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

Drew, D., Lerch, M., Kunji, E., Slotboom, D. J., & de Gier, J. W. (2006). Optimization of membrane protein overexpression and purification using GFP fusions. *Nat Methods*, *3*(4), 303-313. https://doi.org/10.1038/nmeth0406-303

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, *60*(Pt 12 Pt 1), 2126-2132. https://doi.org/10.1107/S0907444904019158

Ernst, D. C., & Downs, D. M. (2015). The STM4195 gene product (PanS) transports coenzyme A precursors in Salmonella enterica. *J Bacteriol*, *197*(8), 1368-1377. https://doi.org/10.1128/JB.02506-14

Fang, S., Huang, X., Zhang, X., Zhang, M., Hao, Y., Guo, H., Liu, L. N., Yu, F., & Zhang, P. (2021). Molecular mechanism underlying transport and allosteric inhibition of bicarbonate transporter SbtA. *Proc Natl Acad Sci U S A*, *118*(22). https://doi.org/10.1073/pnas.2101632118

Furumoto, T., Yamaguchi, T., Ohshima-Ichie, Y., Nakamura, M., Tsuchida-Iwata, Y., Shimamura, M., Ohnishi, J., Hata, S., Gowik, U., Westhoff, P., Brautigam, A., Weber, A. P., & Izui, K. (2011). A plastidial sodium-dependent pyruvate transporter. *Nature*, *476*(7361), 472-475. https://doi.org/10.1038/nature10250

Geyer, J., Wilke, T., & Petzinger, E. (2006). The solute carrier family SLC10: more than a family of bile acid transporters regarding function and phylogenetic relationships. *Naunyn Schmiedebergs Arch Pharmacol*, *372*(6), 413-431. https://doi.org/10.1007/s00210-006-0043-8

Gigolashvili, T., Yatusevich, R., Rollwitz, I., Humphry, M., Gershenzon, J., & Flugge, U. I. (2009). The plastidic bile acid transporter 5 is required for the biosynthesis of methionine-derived glucosinolates in Arabidopsis thaliana. *Plant Cell*, *21*(6), 1813-1829. https://doi.org/10.1105/tpc.109.066399

Goutam, K., Ielasi, F. S., Pardon, E., Steyaert, J., & Reyes, N. (2022). Structural basis of sodium-dependent bile salt uptake into the liver. *Nature*, *606*(7916), 1015-1020. https://doi.org/10.1038/s41586-022-04723-z

Gowers, R. J., Linke, M., Barnoud, J., Reddy, T. J. E., Melo, M. N., Seylor, S. L., Dotson, D. L., Domanski, J., Buchoux, S., Kenney, I. M., & Beckstein, O. (2016). A Python package for the

rapid analysis of molecular dynamics simulations. *Proceedings of the 15th Python in Science Conference*, 98-105. https://doi.org/DOI:10.25080/Majora-629e541a-00e

Grosser, G., Fietz, D., Gunther, S., Bakhaus, K., Schweigmann, H., Ugele, B., Brehm, R., Petzinger, E., Bergmann, M., & Geyer, J. (2013). Cloning and functional characterization of the mouse sodium-dependent organic anion transporter Soat (Slc10a6). *J Steroid Biochem Mol Biol*, *138*, 90-99. https://doi.org/10.1016/j.jsbmb.2013.03.009

Grosser, G., Muller, S. F., Kirstgen, M., Doring, B., & Geyer, J. (2021). Substrate Specificities and Inhibition Pattern of the Solute Carrier Family 10 Members NTCP, ASBT and SOAT. *Front Mol Biosci*, *8*, 689757. https://doi.org/10.3389/fmolb.2021.689757

Hess, B., Bekker, H., Berendsen, H. J. C., & Fraaije, J. G. E. M. (1997). LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, *18*(12), 1463-1472. https://doi.org/10.1002/(sici)1096-987x(199709)18:12<1463::Aid-jcc4>3.0.Co;2-h

Hu, N. J., Iwata, S., Cameron, A. D., & Drew, D. (2011). Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT. *Nature*, *478*(7369), 408-411. https://doi.org/10.1038/nature10450

Huang, H. C., Tremont, S. J., Lee, L. F., Keller, B. T., Carpenter, A. J., Wang, C. C., Banerjee, S. C., Both, S. R., Fletcher, T., Garland, D. J., Huang, W., Jones, C., Koeller, K. J., Kolodziej, S. A., Li, J., Manning, R. E., Mahoney, M. W., Miller, R. E., Mischke, D. A., . . . Reitz, D. B. (2005). Discovery of potent, nonsystemic apical sodium-codependent bile acid transporter inhibitors (Part 2). *J Med Chem*, *48*(18), 5853-5868. https://doi.org/10.1021/jm0402162

Huang, L., Pyc, M., Alseekh, S., McCarty, D. R., de Crecy-Lagard, V., Gregory, J. F., 3rd, Henry, C. S., Fernie, A. R., Mullen, R. T., & Hanson, A. D. (2018). A plastidial pantoate transporter with a potential role in pantothenate synthesis. *Biochem J*, *475*(4), 813-825. https://doi.org/10.1042/BCJ20170883

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph*, *14*(1), 33-38, 27-38. https://doi.org/10.1016/0263-7855(96)00018-5

Hussainzada, N., Da Silva, T. C., Zhang, E. Y., & Swaan, P. W. (2008). Conserved aspartic acid residues lining the extracellular loop 1 of sodium-coupled bile acid transporter ASBT Interact with Na+ and 7alpha-OH moieties on the ligand cholestane skeleton. *J Biol Chem*, *283*(30), 20653-20663. https://doi.org/10.1074/jbc.M802885200

Jo, S., Kim, T., Iyer, V. G., & Im, W. (2008). CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem*, *29*(11), 1859-1865. https://doi.org/10.1002/jcc.20945

Karpen, S. J., Kelly, D., Mack, C., & Stein, P. (2020). Ileal bile acid transporter inhibition as an anticholestatic therapeutic target in biliary atresia and other cholestatic disorders. *Hepatol Int*, *14*(5), 677-689. https://doi.org/10.1007/s12072-020-10070-w

Khanna, L., & Camilleri, M. (2021). Review article: Elobixibat: a novel treatment for chronic constipation. *Aliment Pharmacol Ther*, *53*(2), 234-242. https://doi.org/10.1111/apt.16143

Lee, C., Kang, H. J., Hjelm, A., Qureshi, A. A., Nji, E., Choudhury, H., Beis, K., de Gier, J. W., & Drew, D. (2014). MemStar: a one-shot Escherichia coli-based approach for high-level bacterial membrane protein production. *FEBS Lett*, *588*(20), 3761-3769. https://doi.org/10.1016/j.febslet.2014.08.025

Lee, C., Kang, H. J., von Ballmoos, C., Newstead, S., Uzdavinys, P., Dotson, D. L., Iwata, S., Beckstein, O., Cameron, A. D., & Drew, D. (2013). A two-domain elevator mechanism for sodium/proton antiport. *Nature*, *501*(7468), 573-577. https://doi.org/10.1038/nature12484

Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Brooks, C. L., 3rd, MacKerell, A. D., Jr., Klauda, J. B., & Im, W. (2016). CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput*, *12*(1), 405-413. https://doi.org/10.1021/acs.jctc.5b00935

Liu, H., Irobalieva, R. N., Bang-Sorensen, R., Nosol, K., Mukherjee, S., Agrawal, P., Stieger, B., Kossiakoff, A. A., & Locher, K. P. (2022). Structure of human NTCP reveals the basis of

recognition and sodium-driven transport of bile salts into the liver. *Cell Res*. https://doi.org/10.1038/s41422-022-00680-4

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *J Appl Crystallogr*, *40*(4), 658-674. https://doi.org/doi:10.1107/S0021889807021206

McNicholas, S., Potterton, E., Wilson, K. S., & Noble, M. E. (2011). Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr*, *67*(Pt 4), 386-394. https://doi.org/10.1107/S0907444911007281

Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., & Beckstein, O. (2011). MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem*, *32*(10), 2319-2327. https://doi.org/10.1002/jcc.21787

Park, J. H., Iwamoto, M., Yun, J. H., Uchikubo-Kamo, T., Son, D., Jin, Z., Yoshida, H., Ohki, M., Ishimoto, N., Mizutani, K., Oshima, M., Muramatsu, M., Wakita, T., Shirouzu, M., Liu, K., Uemura, T., Nomura, N., Iwata, S., Watashi, K., . . . Park, S. Y. (2022). Structural insights into the HBV receptor and bile acid transporter NTCP. *Nature*, *606*(7916), 1027-1031. https://doi.org/10.1038/s41586-022-04857-0

Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, *52*(12), 7182-7190. https://doi.org/10.1063/1.328693

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, *25*(13), 1605-1612. https://doi.org/10.1002/jcc.20084

Sonoda, Y., Newstead, S., Hu, N. J., Alguel, Y., Nji, E., Beis, K., Yashiro, S., Lee, C., Leung, J., Cameron, A. D., Byrne, B., Iwata, S., & Drew, D. (2011). Benchmarking membrane protein detergent stability for improving throughput of high-resolution X-ray structures. *Structure*, *19*(1), 17-25. https://doi.org/10.1016/j.str.2010.12.001

South, P. F., Walker, B. J., Cavanagh, A. P., Rolland, V., Badger, M., & Ort, D. R. (2017). Bile Acid Sodium Symporter BASS6 Can Transport Glycolate and Is Involved in Photorespiratory Metabolism in Arabidopsis thaliana. *Plant Cell*, *29*(4), 808-823. https://doi.org/10.1105/tpc.16.00775

Ung, K. L., Winkler, M., Schulz, L., Kolb, M., Janacek, D. P., Dedic, E., Stokes, D. L., Hammes, U. Z., & Pedersen, B. P. (2022). Structures and mechanism of the plant PIN-FORMED auxin transporter. *Nature*, *609*(7927), 605-610. https://doi.org/10.1038/s41586-022-04883-y

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., & Mackerell, A. D., Jr. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem*, *31*(4), 671-690. https://doi.org/10.1002/jcc.21367

Wang, X., Lyu, Y., Ji, Y., Sun, Z., & Zhou, X. (2021). Substrate binding in the bile acid transporter ASBTYf from Yersinia frederiksenii. *Acta Crystallogr D Struct Biol*, *77*(Pt 1), 117-125. https://doi.org/10.1107/S2059798320015004

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189-1191. https://doi.org/10.1093/bioinformatics/btp033

Waterman, D. G., Winter, G., Gildea, R. J., Parkhurst, J. M., Brewster, A. S., Sauter, N. K., & Evans, G. (2016). Diffraction-geometry refinement in the DIALS framework. *Acta Crystallogr D Struct Biol*, *72*(Pt 4), 558-575. https://doi.org/10.1107/S2059798316002187

Wedemeyer, H., Schoneweis, K., Bogomolov, P., Blank, A., Voronkova, N., Stepanova, T., Sagalova, O., Chulanov, V., Osipenko, M., Morozov, V., Geyvandova, N., Sleptsova, S., Bakulin, I. G., Khaertynova, I., Rusanova, M., Pathil, A., Merle, U., Bremer, B., Allweiss, L., . . . Urban, S. (2022). Safety and efficacy of bulevirtide in combination with tenofovir disoproxil fumarate in patients with hepatitis B virus and hepatitis D virus coinfection (MYR202): a multicentre, randomised, parallel-group, open-label, phase 2 trial. *Lancet Infect Dis*. https://doi.org/10.1016/S1473-3099(22)00318-8

Wu, E. L., Cheng, X., Jo, S., Rui, H., Song, K. C., Davila-Contreras, E. M., Qi, Y., Lee, J., Monje-Galvan, V., Venable, R. M., Klauda, J. B., & Im, W. (2014). CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J Comput Chem*, *35*(27), 1997-2004. https://doi.org/10.1002/jcc.23702

Yan, H., Peng, B., Liu, Y., Xu, G., He, W., Ren, B., Jing, Z., Sui, J., & Li, W. (2014). Viral entry of hepatitis B and D viruses and bile salts transportation share common molecular determinants on sodium taurocholate cotransporting polypeptide. *J Virol*, *88*(6), 3273-3284. https://doi.org/10.1128/JVI.03478-13

Zhou, X., Levin, E. J., Pan, Y., McCoy, J. G., Sharma, R., Kloss, B., Bruni, R., Quick, M., & Zhou, M. (2014). Structural basis of the alternating-access mechanism in a bile acid transporter. *Nature*, *505*(7484), 569-573. https://doi.org/10.1038/nature12811

Table 1: Data Processing and Refinement Statistics.

	$ASBT_{NM(Pan)}$	$ASBT_{NM(ns)}$
Wavelength (Å)	0.9999	0.9999
Resolution range	43.78 - 2.3 (2.382 - 2.3)	58.96 - 2.1 (2.175 - 2.1)*
Space group	C2	$P\ 2_1\ 2_1\ 2_1$
Unit cell: a, b, c (Å),α,β,γ (°)	85.0 89.4 53.1 90 124.4 90	49.5 80.6 86.5 90 90 90
Total reflections	55444 (3034)	146457 (14741)
Unique reflections	14273 (1201)	20744 (2032)
Multiplicity	3.9 (2.5)	7.1 (7.3)
Completeness (%)	97.7 (82.8)	99.4 (99.3)
Mean I/sigma(I)	5.7 (1.7)	6.7 (2.0)
Wilson B-factor	36	30
R-merge	0.1181 (0.4632)	0.1356 (0.8541)
R-meas	0.1368 (0.5779)	0.1467 (0.9198)
R-pim	0.06752 (0.3397)	0.05402 (0.3316)
CC1/2	0.988 (0.804)	0.957 (0.658)
CC*	0.997 (0.944)	0.989 (0.891)
Reflections used in refinement	14259 (1201)	20691 (2027)
Reflections used for R-free	774 (55)	1089 (96)
R-work	0.2284 (0.3488)	0.2115 (0.2933)
R-free	0.2648 (0.4564)	0.2387 (0.2885)
CC(work)	0.937 (0.861)	0.929 (0.846)
CC(free)	0.927 (0.678)	0.835 (0.907)
Number of non-hydrogen atoms	2324	2412
Macromolecules	2276	2282
Ligands	34	135
Solvent	25	34
Protein residues	310	310
RMS(bonds)	0.002	0.002
RMS(angles)	0.44	0.53
Ramachandran favoured (%)	98.05	99.03
Ramachandran allowed (%)	1.95	0.97
Ramachandran outliers (%)	0	0
Rotamer outliers (%)	0.85	0.84
Clashscore	2.52	3.25
Average B-factor	52.2	41.5
Macromolecules	52.3	40.3
Ligands	46.0	69.4
Solvent	47.6	42.8
Number of TLS groups	1	1

* Statistics for the highest resolution shell are shown in parentheses.

Figures

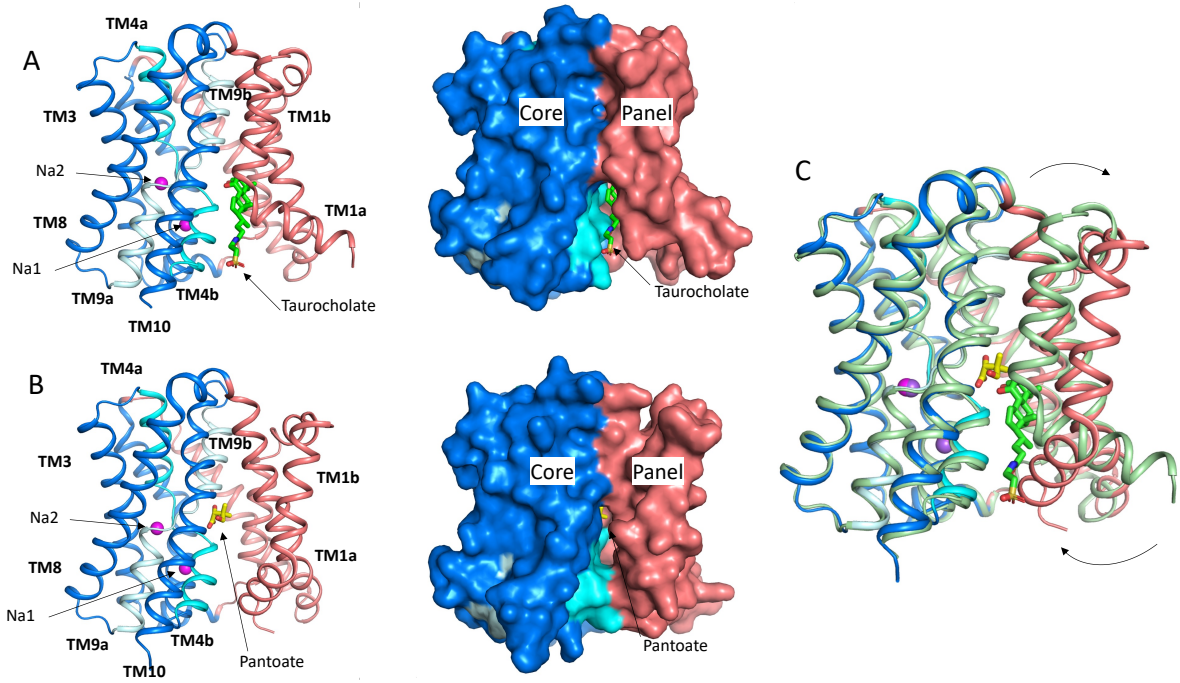


Figure 1: Structure of $ASBT_{NM}$. A) Structure of $ASBT_{NM}$ in complex with Taurocholate $ASBT_{NM(TCH)}$ (Hu et al., 2011). The panel domain is coloured salmon. The core domain is coloured blue with the cross-over helices, TM4 and TM9 in cyan and pale blue respectively. The taurocholate is shown in a stick representation with green carbon atoms and the sodium ions are shown as magenta spheres. Left, cartoon representation, Right, surface representation **B)** As A for the structure in complex with pantoate ($ASBT_{NM(Pan)}$). The pantoate is depicted with yellow carbon atoms. **C)** Superposition of $ASBT_{NM(Pan)}$ (colouring as A) on $ASBT_{NM(TCH)}$ (pale green carbon atoms). The arrows show the movement of TM1.

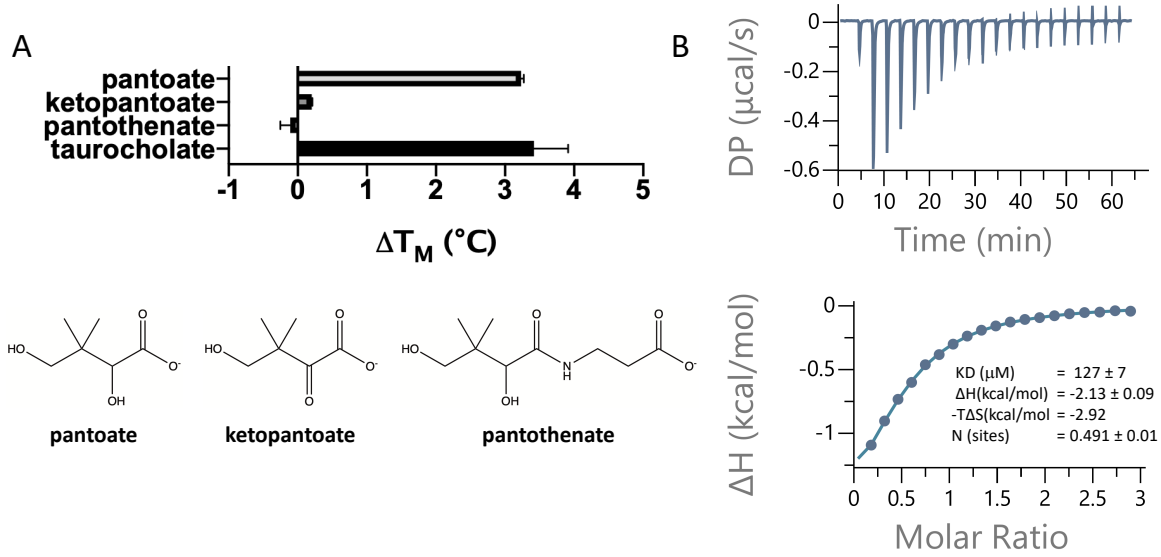


Figure 2: Pantoate binding to $ASBT_{NM}$. A) Results from thermostability assay showing that pantoate stabilises $ASBT_{NM}$ to a similar extent to taurocholate. The compounds are shown below. The mean standard deviation is shown based on three individual experiments. **B)** Pantoate binding to $ASBT_{NM}$ measured by Isothermal calorimetry.

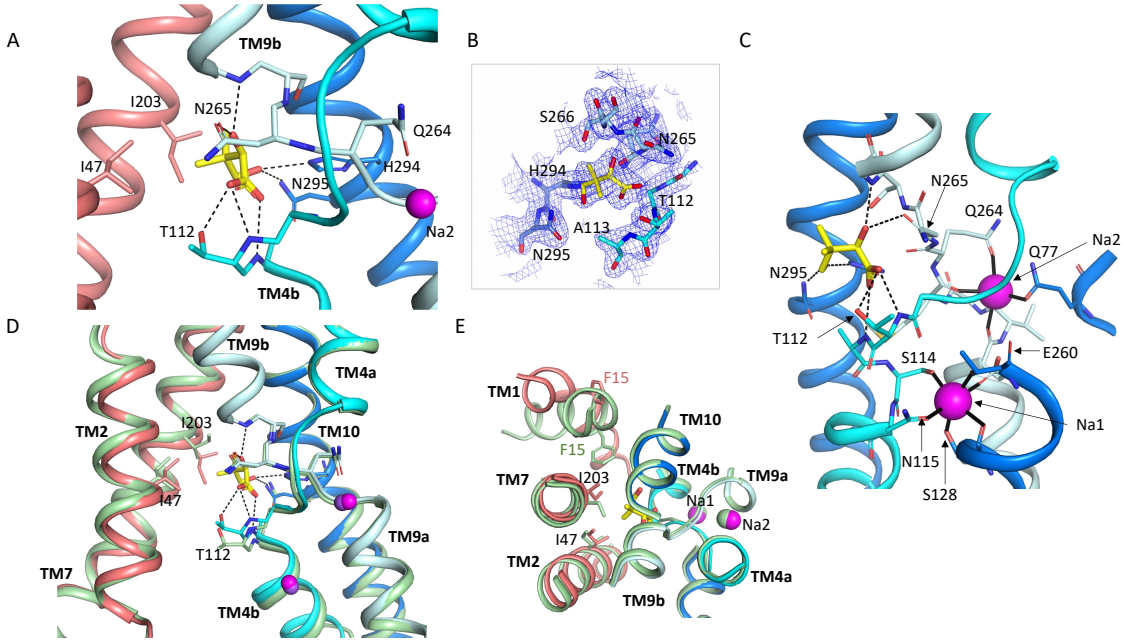


Figure 3: Pantoate binding site. A) The pantoate binding site in the $ASBT_{NM(Pan)}$ structure, coloured as in Fig. 1. Hydrogen bonds are shown as dashed lines. **B)** 2mFo-Fc density for the refined structure. The density is contoured at 1σ. **C)** View of the $ASBT_{NM(Pan)}$ structure highlighting the juxtaposition of the residues interacting with the sodium ions and those interacting with the pantoate. **D)** Superposition of the $ASBT_{NM(TCH)}$ structure (pale green) on the $ASBT_{NM(Pan)}$ structure highlighting the difference in position of TM4b and especially Thr 112 between the two structures. **E)** As D but shown from the extracellular side highlighting the differences in position of Ile 203 and Phe 15.

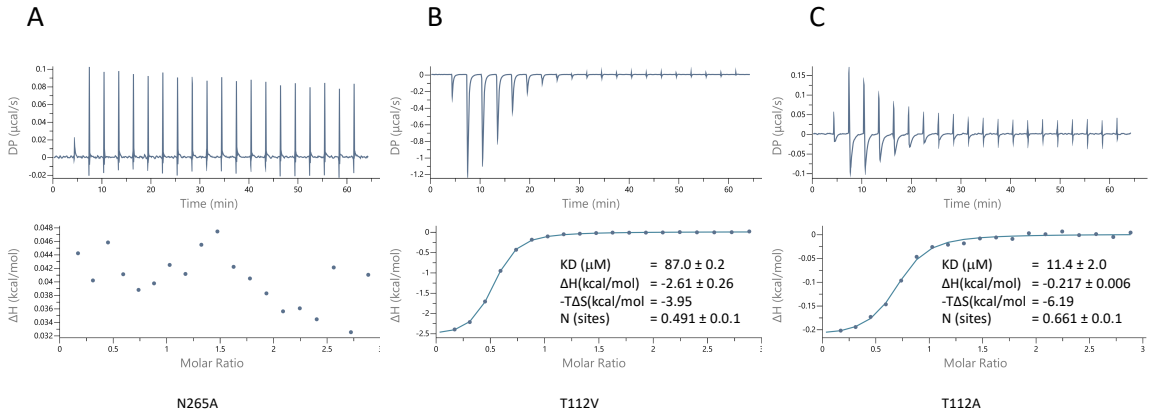


Figure 4: Characterisation of pantoate binding to mutants of $ASBT_{NM}$. Pantoate binding to $ASBT_{NM}$ mutants measured by Isothermal calorimetry.

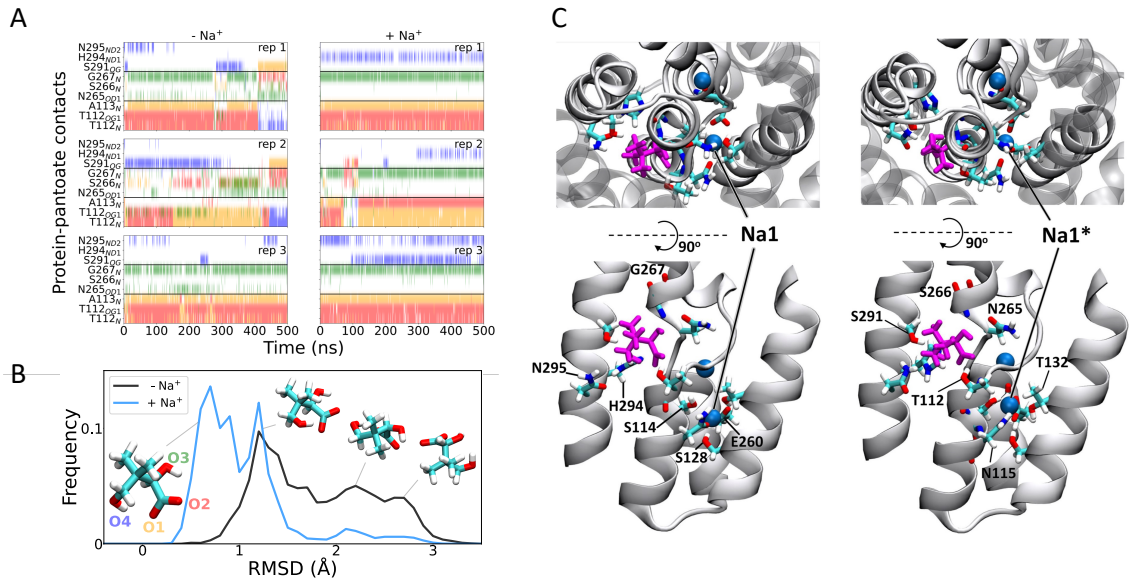


Figure 5: Molecular dynamics simulations. A) Hydrogen bonds between the pantoate and protein followed over the course of the simulations starting without i) or with ii) sodium bound. Red, yellow, green and blue indicate a contact with the O1, O2, O3 and O4 atoms of pantoate, respectively (as shown in B). Contacts are shown for all residues with contacts in greater than 10% of any simulation. **B**) Histograms of pantoate heavy-atom RMSD over 3 simulations starting with (blue) or without (black) bound sodium, calculated following Cα alignment of the protein around the binding site (residues 108-117 (TM4), 199-207 (TM 9) and 287-296 (TM 10)). Representative snapshots of pantoate, relative to the starting position on the left are shown. **C)** Representative snapshots showing bound pantoate (magenta) and sodium (blue spheres), showing the location of the canonical sodium binding site 1 and the alternate site 1*. ASBT is shown with cartoon representation; for clarity, only helices contributing to the binding sites (TMs 4, 5, 9, 10) are shown in the lower (side view) panels. Residues making up the pantoate and Na1 binding sites are shown in stick representation.

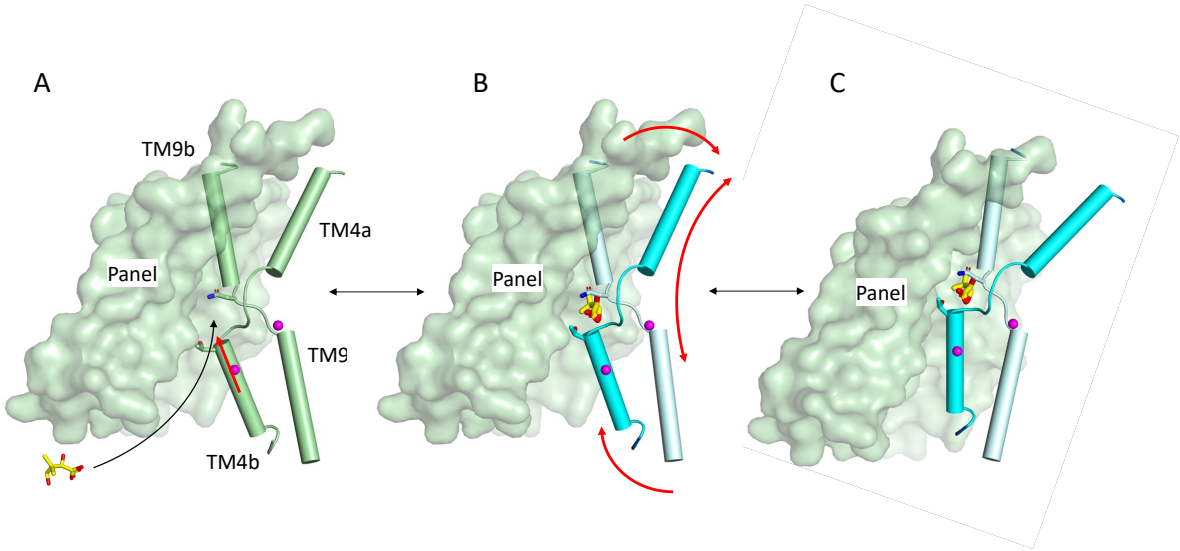


Figure 6: Schematic of mechanism. Pantoate binding to the cross-over region between TM4b and TM9b of the substrate-free structure **(A)** elicits a conformational change in TM4b (red arrow) **(B)**. The change in conformation of the core region of the protein may allow greater freedom of movement of the panel domain relative to the core enabling it to swing upwards in an elevator movement **(C)** (red arrows in B). The position of the core relative to the panel domain in **C** was based on the relative positions of the two domains in the outward-facing structure of $ASBT_{YF}$.

Supplementary Figures

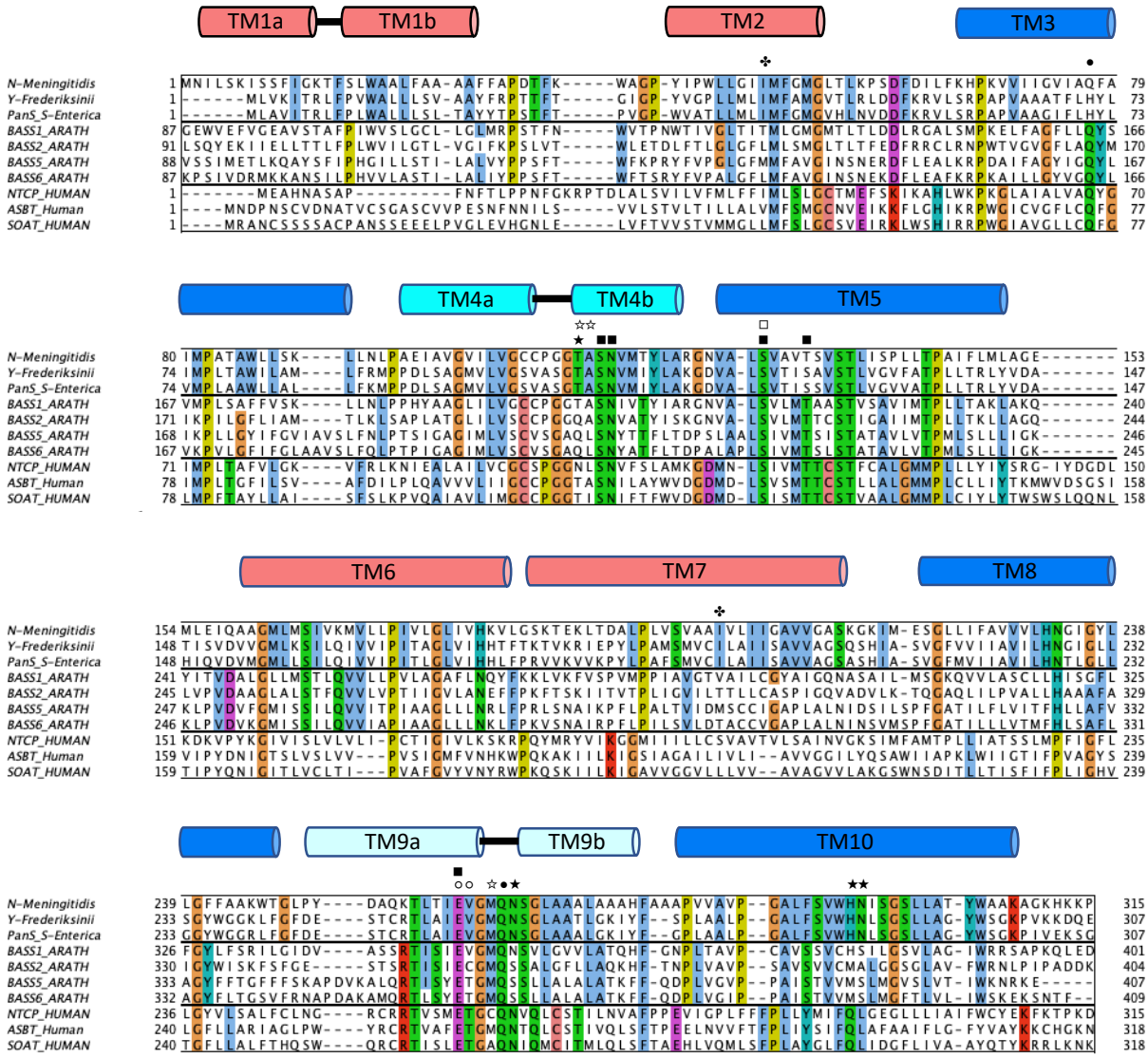


Figure S1: Sequence alignment. Sequence alignment of $ASBT_{NM}$ (Q9K0A9) against the bacterial transporters $ASBT_{YF}$ (4N7W) and PanS Q8ZKL0, and with selected BASS transporters from humans (NTCP Q14973, ASBT Q12908 SOAT Q3KNW5) and *Arabidopsis Thaliana* (BASS 1 Q93YR2, BASS 2 Q1EBV7, BASS 5 F4JPW1 BASS 6 Q8VYY4, . The plant and human BASS transporters chosen are the best characterised in the family and are most likely to be sodium-coupled transporters based on the conservation of the sodium-binding residues. The secondary structure in $ASBT_{NM}$ is shown with the transmembrane helices coloured as in Fig. 1. Bacterial, plant and human transporters are grouped separately with residues completely

conserved within each group coloured according to the ClustalW colouring scheme in Jalview (Waterhouse et al., 2009). Starred residues denote hydrogen bonding to the pantoate with a solid (★) for side chain interactions and an open star (☆) for main chain. The symbol ✤ denotes a residue within van der Waals distance of the pantoate. Interactions with Na1 are denoted by squares with solid squares (■) denoting interactions with the side chain and open squares (□) for interactions through the main-chain carbonyl oxygen. Interactions with Na2 are denoted by circles with filled circles (●) denoting interactions with the side chain and open circles (○) for interactions through the main-chain carbonyl oxygen.

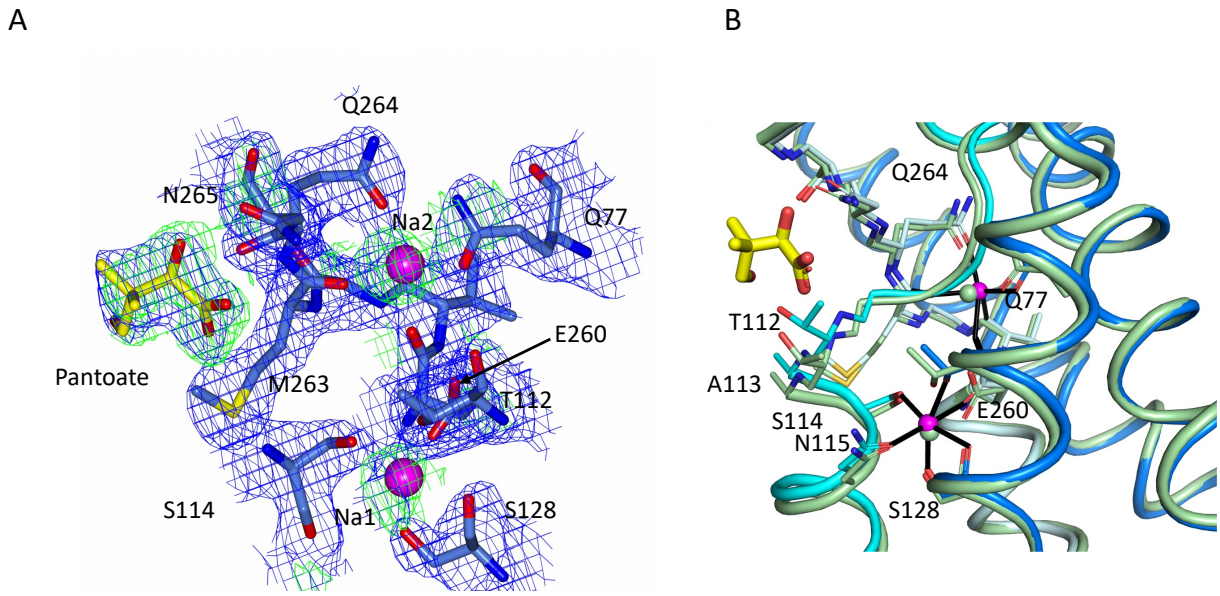


Figure S2: Sodium site for the ASBT$_{NM(Pan)}$ structure. A) Electron density associated with the sodium ions and pantoate. The 2mFo-DFc density (blue) was calculated based on phases from the refined structure and was contoured at 1σ. The mFo-DFc density (green), contoured at 2.5σ was calculated from a map where the structure had been refined with simulated annealing after omission of the sodium ions and pantoate. The methyl-propanol moiety is less well-defined than the hydroxyacetate. **B)** Superposition of the ASBT$_{NM(Pan)}$ structure (coloured as Fig. 1) on the ASBT$_{NM(TCH)}$ structure (pale green carbon atoms).

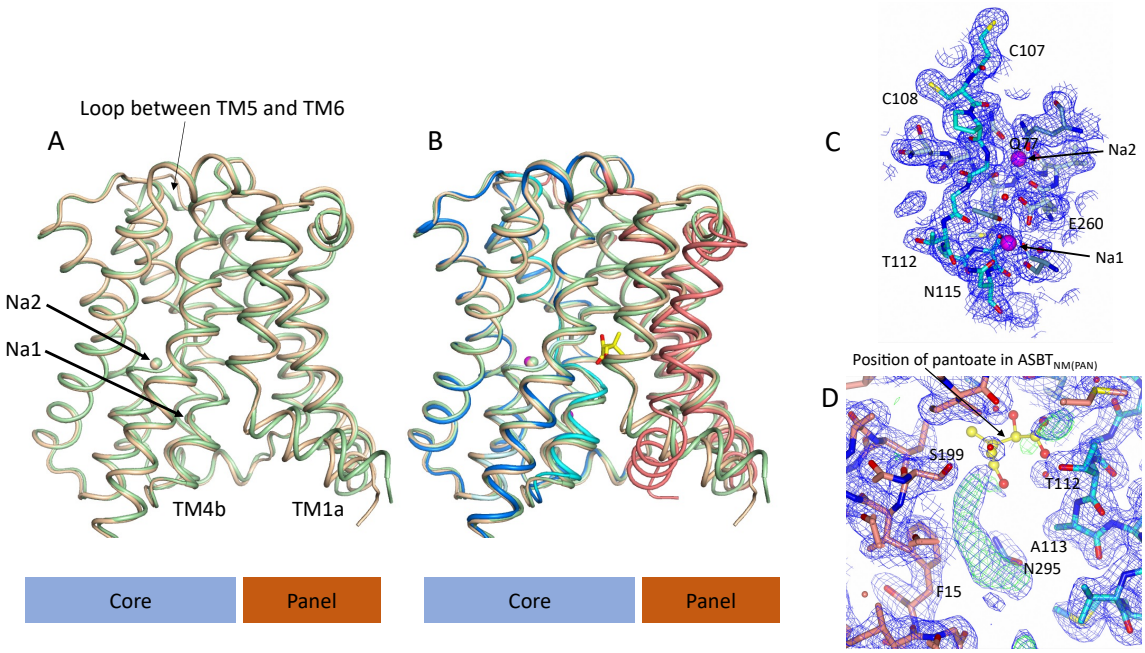
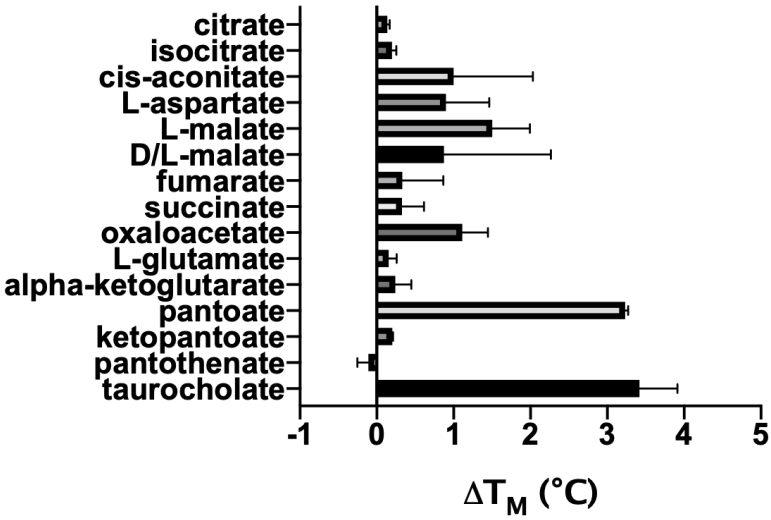


Figure S3: Structure of $ASBT_{NM}$ without pantoate or taurocholate. A) Superposition of $ASBT_{NM(ns)}$ (wheat) on $ASBT_{NM(TCH)}$ (pale green) highlighting the similarity of the two structures. The main differences are in the position of TM1, where TM1a adopts a slightly different angle with respect to TM1b, and in the loop between TMs 5 and 6, which links the core domain to the panel domain. **B)** As A with the addition of the $ASBT_{NM(Pan)}$ structure (coloured as in Fig. 1). The difference in the position of TM1 and TM4b, with respect to the two structures without pantoate is evident. **C)** Electron density associated with the $ASBT_{NM(ns)}$ structure in the vicinity of the sodium ions. The 2mFo-DFc density was calculated based on phases from the refined structure and was contoured at 1σ. **D)** There is additional electron density in the open cleft of the $ASBT_{NM(ns)}$ structure, which has been modelled as the hydrophobic tail of monoolein. For reference the position of the pantoate from the $ASBT_{NM(Pan)}$ structure has been inserted with yellow carbon atoms. The 2mFo-DFc density (blue) is contoured at 1σ and the mFo-DFc density (green) at 3σ.

A



B

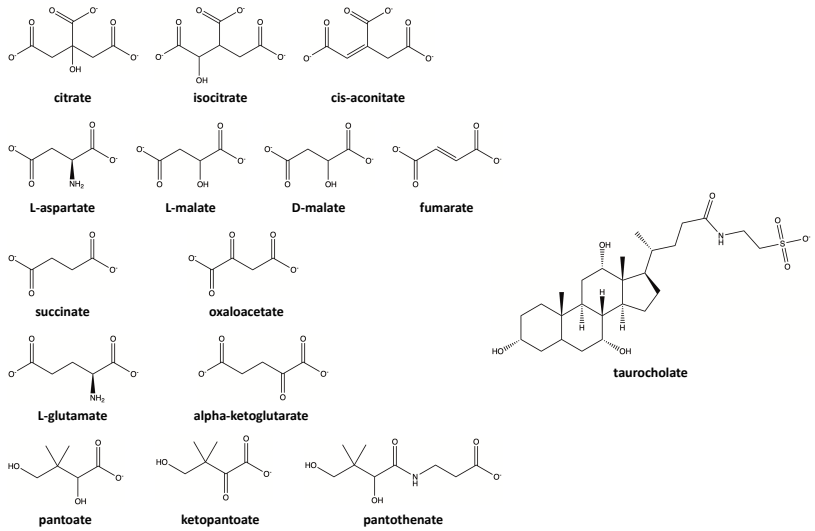


Figure S4: Testing a panel of compounds for potential binding to $ASBT_{NM}$. A) Compounds were subjected to the stability assay. These compounds include citrate, which has been observed in two different crystal structures of $ASBT_{YF}$ (Wang et al., 2021; Zhou et al., 2014), other compounds from the citric acid cycle, all of which have a carboxylic acid group in common with pantoate and citrate, aspartate given that a gene encoding aspartate kinase is located next to the gene encoding $ASBT_{YF}$ in *Yersinia Frederiksinii*, and glutamate. **B)** Chemical formula of the compounds tested.

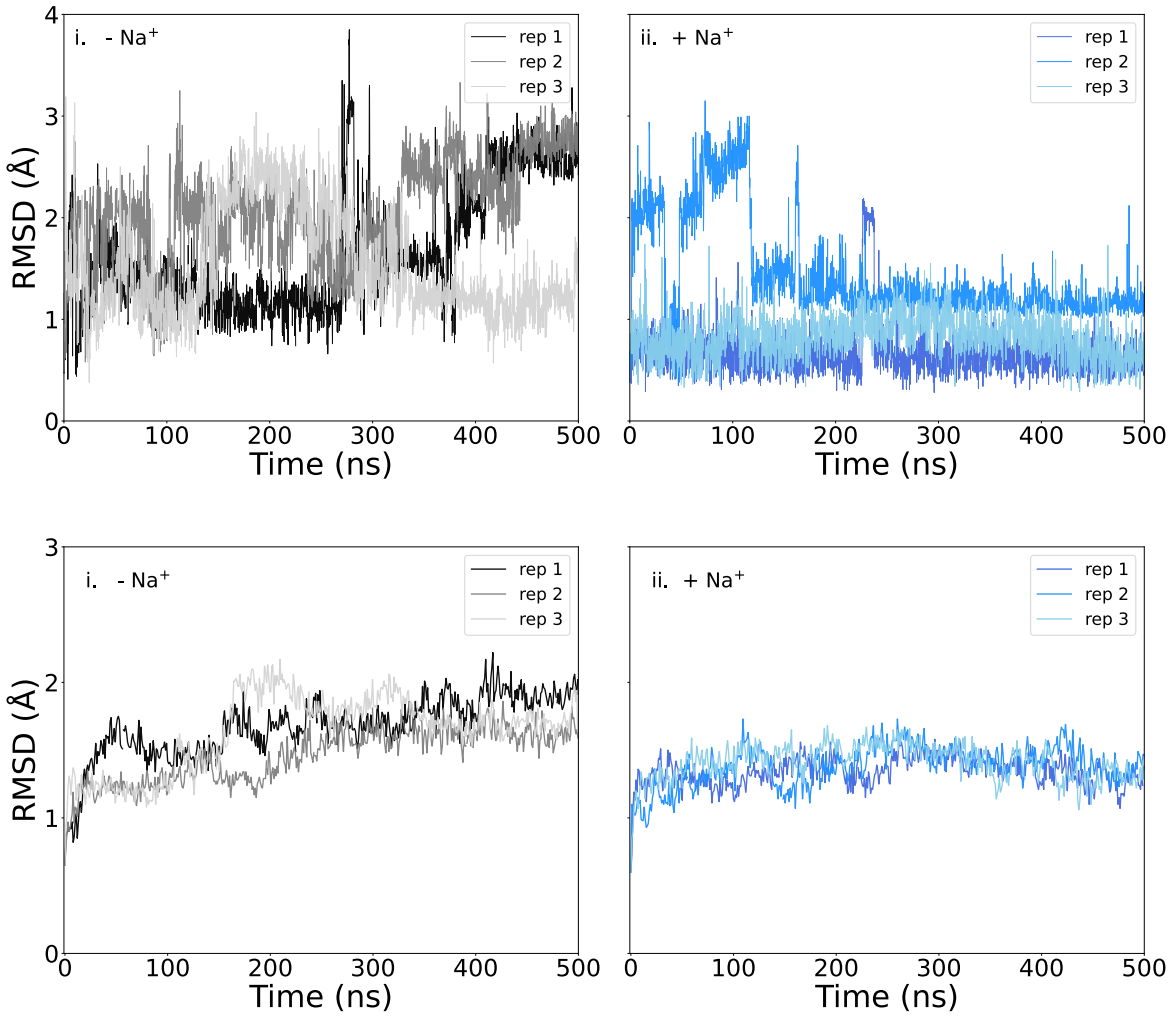


Figure S5: RMSD of protein and pantoate during simulations. A) Pantoate heavy-atom RMSD relative to the initial structure for each repeat simulation starting without i) or with ii) sodium bound. The RMSD was calculated following the alignment of protein structures, using the Cα atoms of residues surrounding the pantoate binding site (residues 108-117 (TM4), 199-207 (TM 9) and 287-296 (TM 10)). **B)** Protein Cα RMSD over all simulations for simulations starting without i) or with ii) sodium bound, relative to the initial structure.

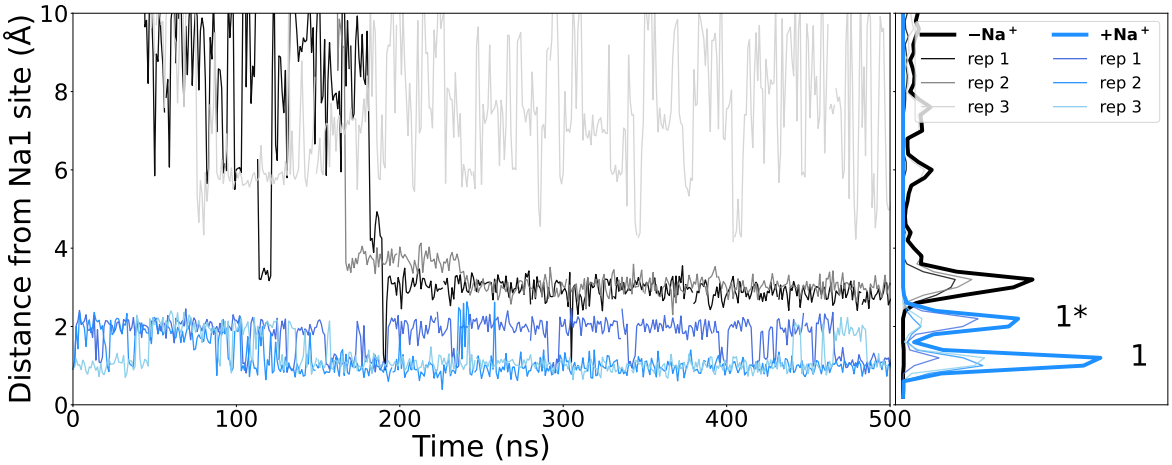


Figure S6: Sodium binding during simulations. Distance of closest sodium ion to the Na1 binding site (measured as the centre of mass of Cα atoms of residues making the Na1 binding site – S114, N115, S128, T132 and E260) throughout simulations starting with (blue) or without (black) sodium bound. The locations of ions while in the canonical Na1 site or the alternate 1* site are indicated.

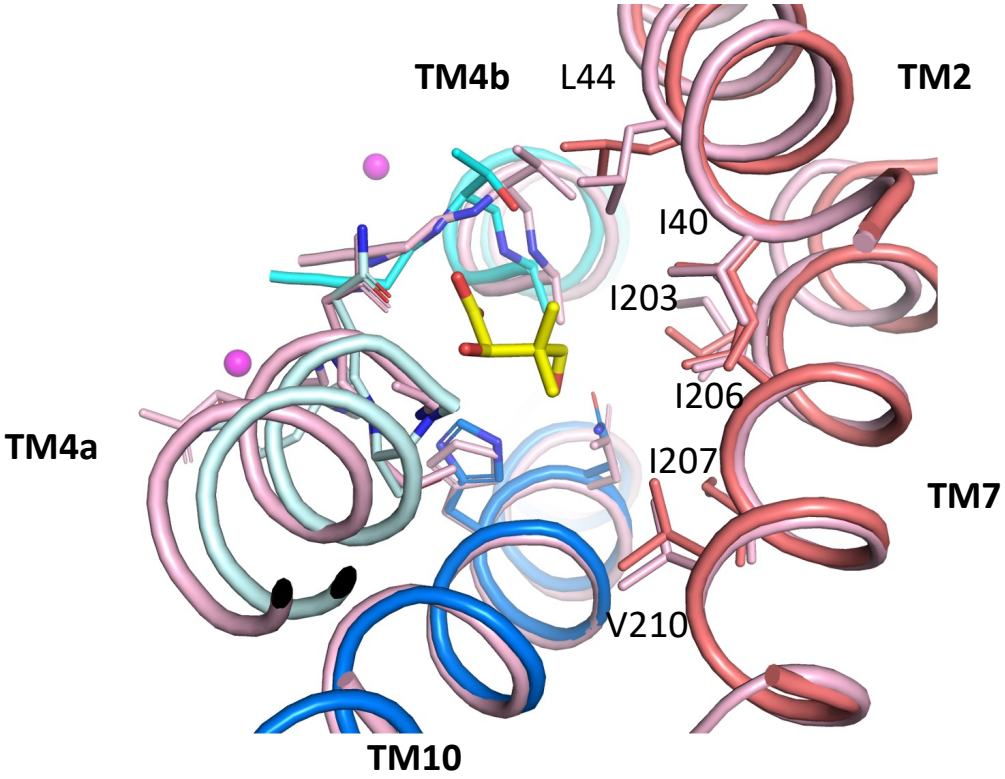


Figure S7: Pantoate binding to an outward-facing state model. Pantoate binding region of $ASBT_{NM(Pan)}$ and panel domain of $ASBT_{NM(TCH)}$ superposed separately on the core and panel regions of the outward-facing structure of $ASBT_{YF}$ (4N7X; pink). Pantoate would easily be accommodated in the outward-facing structure. Numbering is shown for $ASBT_{NM}$.

Movie S1: Morph between $ASBT_{NM(Pan)}$ and $ASBT_{NM(TCH)}$. The colouring is shown as in Fig. 1. The sodium ions and pantoate from the $ASBT_{NM(Pan)}$ structure are held rigid during the morph of the protein atoms.

Movie S2: Morph between $ASBT_{NM(Pan)}$ and $ASBT_{NM(ns)}$ focussed on TM4b. The colouring is shown as in Fig. 1. The sodium ions and pantoate from the $ASBT_{NM(Pan)}$ structure are held rigid during the morph of the protein atoms.