Structural basis of sodium-dependent bile salt uptake into the liver

Kapil Goutam, Francesco Ielasi, Els Pardon, Jan Steyaert, Nicolas Reyes

HAL Id: hal-04688088
https://hal.science/hal-04688088

Article

Structural basis of sodium-dependent bile salt uptake into the liver

Kapil Goutam[1,2], Francesco S. Ielasi[2], Els Pardon[3,4], Jan Steyaert[3,4] & Nicolas Reyes[1,2 ✉]

The liver takes up bile salts from blood to generate bile, enabling absorption of lipophilic nutrients and excretion of metabolites and drugs[1]. Human Na+–taurocholate co-transporting polypeptide (NTCP) is the main bile salt uptake system in liver. NTCP is also the cellular entry receptor of human hepatitis B and D viruses[2,3] (HBV/HDV), and has emerged as an important target for antiviral drugs[4]. However, the molecular mechanisms underlying NTCP transport and viral receptor functions remain incompletely understood. Here we present cryo-electron microscopy structures of human NTCP in complexes with nanobodies, revealing key conformations of its transport cycle. NTCP undergoes a conformational transition opening a wide transmembrane pore that serves as the transport pathway for bile salts, and exposes key determinant residues for HBV/HDV binding to the outside of the cell. A nanobody that stabilizes pore closure and inward-facing states impairs recognition of the HBV/HDV receptor-binding domain preS1, demonstrating binding selectivity of the viruses for open-to-outside over inward-facing conformations of the NTCP transport cycle. These results provide molecular insights into NTCP 'gated-pore' transport and HBV/HDV receptor recognition mechanisms, and are expected to help with development of liver disease therapies targeting NTCP.

Bile salts are essential molecules for absorption of lipophilic nutrients and vitamins (vitamin A, D, E and K) in the small intestine, as well as for maintenance of endocrine and cholesterol homeostasis and excretion of toxins[1]. The vast majority—more than 90%—of the body's bile salts pool is recycled daily, shuttling between intestine and liver, where bile salts are used to aid nutrient absorption and generate bile, respectively. Human members of the solute carrier 10 (SLC10) protein family are key bile salt transporters for the maintenance of enterohepatic circulation[5,6]: NTCP[7] (also known as SLC10A1) is mainly expressed in the hepatocyte basolateral membrane, and constitutes the main active transport route of bile salts into the liver from blood, whereas apical sodium-dependent bile acid transporter[8] (ASBT (also known as SLA10A2)) is expressed in ileum enterocytes and takes up bile salts from the intestinal lumen. Both transporters are important pharmacological targets, as they can be used to facilitate oral absorption[9,10] (ASBT) and liver uptake[11,12] (NTCP) of drugs conjugated to bile salts, and are involved in the action mechanism (ASBT)[13] and pharmacokinetics (NTCP)[14,15] of cholesterol-lowering therapies. Moreover, NTCP down-regulation in mouse models is associated with increased cholesterol and phospholipid excretion[16], as well as decreased weight gain with a high-fat diet[17]. Notably, NTCP has a fundamental role in liver pathology, as the human cellular entry receptor for HBV/HDV[2,3]. Chronic HBV infection is a major cause of hepatocellular carcinoma and liver cirrhosis, and affects around 250 million people globally[18,19]. The viruses use the myristoylated and unstructured N-terminal domain in the large envelope protein—the preS1 domain (myr-preS1)—to recognize and bind

human NTCP[20–22], explaining viral hepatotropism and the narrow range of animal hosts. Consistently, myristoylated peptides encompassing the residues 2–48 of myr-preS1 (myr-preS1[48]) act as potent inhibitors of HBV/HDV entry into cells[23–26].

Structural insights into the transport mechanism of NTCP and ASBT have come from early X-ray crystal structures of prokaryotic homologues that revealed a ten-transmembrane-helix topology, arranged into core and panel domains[27,28]. The homologues follow an alternating-access transport mechanism, in which relative movements of the two domains provide alternating access to substrate- and sodium-binding sites on opposite sides of the membrane.

Here we set out to study the structural basis of human NTCP function using cryo-electron microscopy (cryo-EM) in combination with conformation-specific nanobodies to reveal key conformational transitions of the NTCP transport cycle.

Cryo-EM structure determination

Human NTCP is a relatively small (approximately 38 kDa) dynamic membrane protein that lacks soluble folded domains and is biochemically unstable in non-denaturing detergent solution, posing a challenge for single-particle cryo-EM structure determination. To overcome these problems, we first exchanged amino acids in the sequence of wild-type NTCP for consensus residues of representative vertebrate orthologues to confer stability to the protein[29] (Methods). The initial consensus design, NTCP_CO, was more stable than wild-type NTCP in

[1]Membrane Protein Mechanisms Group, European Institute of Chemistry and Biology, University of Bordeaux, CNRS-UMR5234, Pessac, France. [2]Membrane Protein Mechanisms Unit, Institut Pasteur, Paris, France. [3]Structural Biology Brussels, Vrije Universiteit Brussel, VUB, Brussels, Belgium. [4]VIB-VUB Center for Structural Biology, VIB, Brussels, Belgium. ✉e-mail: nicolas.reyes@u-bordeaux.fr
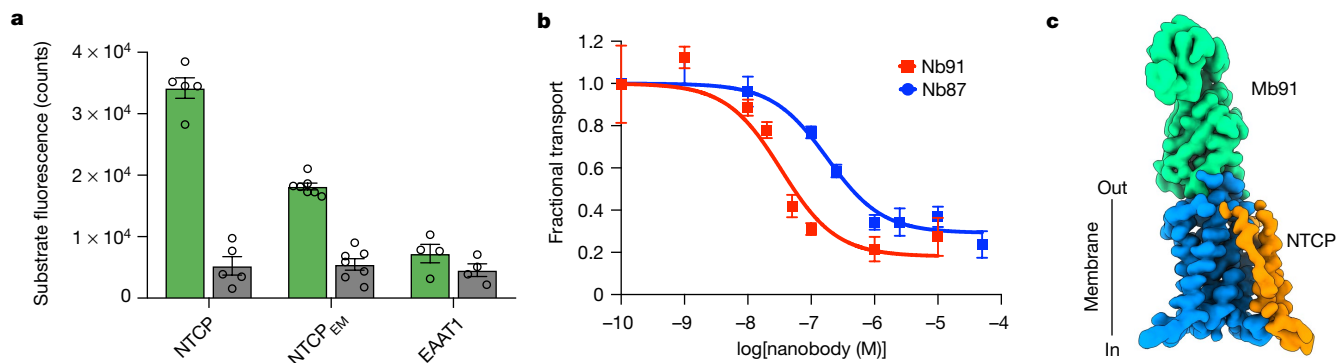
Fig. 1 | Functional and structural analyses of NTCP$_{EM}$–Nb complexes.
a, Uptake of the fluorescent substrate analogue (tauro-nor-THCA-24-DBD) in cells expressing NTCP constructs or negative control (excitatory amino acid transporter EAAT1) using sodium-based (green) and choline-based (grey) buffers, respectively. Bars depict mean of 5 (NTCP), 7 (NTCP$_{EM}$) and 4 (EAAT1) biologically independent experiments, error bars represent s.e.m. and circles show values from individual experiments. **b**, Inhibition by Nb91 (red) and Nb87 (blue) of tauro-nor-THCA-24-DBD transport in cells expressing NTCP$_{EM}$. Solid lines are fits of a single-site binding equation (Methods). Squares depict mean of three biologically independent experiments and error bars represent s.e.m. **c**, Density corresponding to NTCP$_{EM}$ core (blue) and panel (orange) domains, and the nanobody part of Mb91 (green).

detergent solution. To minimize the number of consensus exchanges and maximize stability, we determined the contribution of single exchanges, and retained only those that increased stability, yielding a final construct that shares approximately 98% amino acid identity with wild-type NTCP (Extended Data Fig. 1) and enables purification of monodisperse material in milligram amounts. We refer to this construct as NTCP$_{EM}$. NTCP$_{EM}$ showed robust Na$^+$-dependent uptake of the fluorescent substrate analogue tauro-nor-THCA-24-DBD (4.5 ± 1.3-fold increase in the sodium- over the choline-based condition), similar to that of wild-type NTCP (10.2 ± 3.9), whereas control cells expressing the unrelated Na$^+$-dependent neurotransmitter transporter EAAT1 lacked bile salt uptake (1.6 ± 0.1 sodium-dependent increase) (Fig. 1a). These results show that the transport mechanism is conserved in NTCP$_{EM}$.

Second, to provide molecular features on the NTCP$_{EM}$ surface for cryo-EM analysis, we generated and selected nanobodies that potently bind NTCP$_{EM}$. Nanobody (Nb)87 and Nb91 inhibit Na$^+$-induced fluorescent-substrate uptake by cells expressing NTCP$_{EM}$ with half-maximal inhibitory concentrations (IC$_{50}$) of approximately 180 and 34 nM (Fig. 1b), respectively, showing that they recognize NTCP$_{EM}$ from the extracellular side, and suggesting that they stabilize conformational intermediates of the transport cycle. During cryo-EM sample optimization, we screened NTCP$_{EM}$ complexes with these nanobodies and megabody scaffolds that result in an additional 85 kDa of folded

domains[30] in both detergent solutions, as well as reconstituted in nanodiscs. This yielded final cryo-EM maps of NTCP$_{EM}$–Nb87 in nanodiscs, and NTCP$_{EM}$–megabody (Mb)91 in detergent at overall resolutions of 3.7 and 3.3 Å, respectively, enabling structure determination (Fig. 1c, Extended Data Figs. 2–4, Extended Data Table 1).

NTCP architecture

NTCP$_{EM}$ adopts an SLC10 fold with two structurally distinct domains—core and panel (Fig. 2a, b)—and contains nine transmembrane α-helices (TM1–9) with an unstructured N terminus on the extracellular side. The transmembrane helices are connected by short loops, as well as extracellular α-helices (ECH) and intracellular α-helices (ICH) lying nearly parallel to the membrane. The panel domain is formed by TM1, TM5 and TM6, and has lost pseudo-internal symmetry compared with its equivalent in SLC10 prokaryotic homologues, owing to the evolutionary loss of one transmembrane helix. The NTCP$_{EM}$ core domain is formed by packing of two helix bundles, TM2–4 and TM7–9, which are related by pseudo-two-fold symmetry (Cα root mean squared deviation (r.m.s.d.) ≈ 5 Å). TM3 and TM8 unwind close to the middle of the membrane, and pack against each other to form a characteristic X-shaped structure that displays highly conserved polar residue motifs among vertebrate SLC10 bile salt transporters (Extended Data Figs. 5, 6).
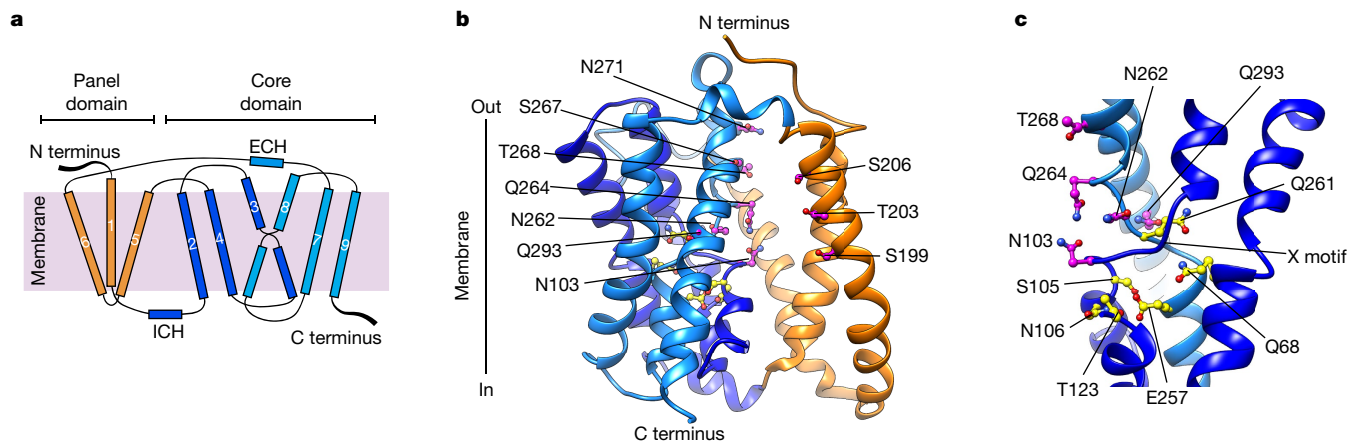

Fig. 2 | NTCP topology and architecture. a, Cartoon representation of NTCP topology. **b**, Structure of NTCP$_{EM}$ in complex with Mb91. Mb91 is omitted for clarity. TM2–TM4 (dark blue) and TM7–TM9 (light blue) in the core domain are related by pseudo-two-fold symmetry, and the panel domain is formed by TM1 and TM5–TM6 (orange). Polar conserved residues lining the space between the core and panel domain (pink), as well as sidechains contributing to Na1 and Na2 (yellow) are shown. **c**, The X motif is formed by unwinding of TM3 and TM8. Only TM2 and TM3 (dark blue) and TM8 (light blue) are shown.
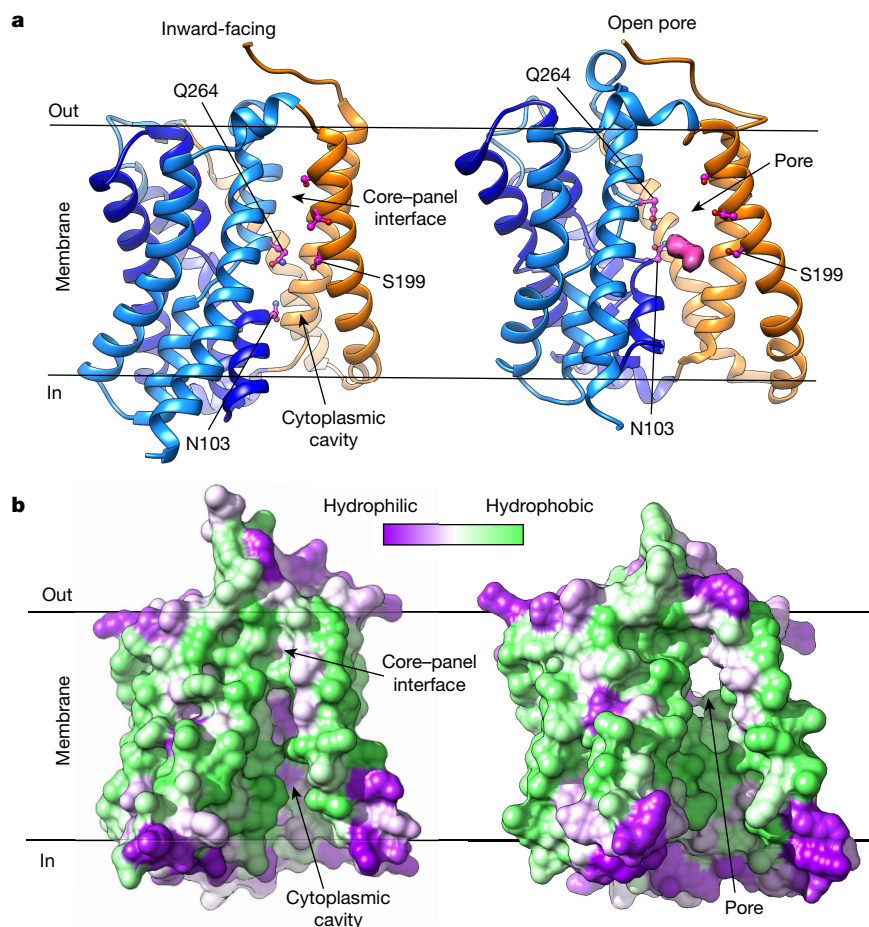
Fig. 3 | Isomerization between open-pore and inward-facing states.
a, NTCP$_{EM}$ structures in complex with Nb87 (left) and Mb91 (right) adopting inward-facing and open-pore conformations, respectively. Nb87 and Mb91 are not shown. Additional cryo-EM density in the open-pore structure is shown (pink surface) and polar residues in close proximity to the density are labelled in both structures for comparison. **b**, Molecular surface representation of NTCP$_{EM}$ inward-facing (left) and open-pore (right) structures. Surfaces are coloured on the basis of a residue hydrophobicity scale, with green representing the most hydrophobic residues and purple representing the most hydrophilic residues. The structures have been tilted by around 20° relative to those in **a** to better display the cytoplasmic cavity (left), and the open pore (right).

Most of the reported residues important for binding of sodium and substrate map to the core domain. Sodium-binding sites 1 (Na1; including S105, N106, T123 and E257 sidechains) and 2 (Na2; including Q68 and Q261), which were first observed in crystallographic studies of prokaryotic SLC10 homologues[27], are structurally conserved in NTCP$_{EM}$ (Fig. 2c). Structural conservation and NTCP mutagenesis[31,32] strongly suggest that the two sodium ions that are thermodynamically coupled to bile salt transport[33,34] bind to these sites. Beyond Na1 and Na2, mutations at residues in the X motif[27] (equivalent to N262) or in close proximity[32] (Q293) impaired transport function, suggesting a role in substrate binding. Consistently, the NTCP-inactivating mutation[35] S267F, which is associated with hypercholanaemia and vitamin D deficiency in humans[36] lays just above the X motif, and A64T[37] is close to the sodium-binding sites.

NTCP inward-facing state

In complex with Nb87, NTCP$_{EM}$ adopts an inward-facing state with core and panel domains tightly packing against each other on the extracellular side of the membrane (Fig. 3a, b). On the intracellular side, the domains separate, uncovering an amphiphilic large cavity (molecular surface volume > 1,500 Å^3) that opens to the cytoplasm, as well as laterally to the hydrophobic core of the membrane through a crevice between TM6 and TM9. On the other side of the transporter,

TM1 and TM5 pack against the core domain, occluding the cavity from the membrane.

Na1 and Na2 face this cavity and localize behind the conserved X motif. In addition, the cavity is lined by several conserved polar residues from the core domain (including N103, N262, Q264 and Q293), some of which have been shown to be important for transport; hydrophobic residues are mostly located in the panel domain. Amino acid conservation, mutagenesis studies and the large volume of the cavity suggest that it is part of the substrate pathway on the cytoplasmic side. Consistently, a molecule of taurocholate has been reported to bind to the equivalent region in the structure of a prokaryotic homologue[27].

Transition to the open-pore conformation

In complex with Mb91, NTCP$_{EM}$ shows a marked conformational change compared with the inward-facing state (Fig. 3a, b, Supplementary Video 1). Core and panel domains rotate around 20° and move approximately 5 Å towards opposite sides of the membrane as nearly rigid bodies. These movements are facilitated by conserved glycine and proline residues that act as hinges in the connecting loops, as well as in the ICH and ECH (Extended Data Fig. 5). As a consequence, the two domains separate from each other on both extracellular and cytoplasmic sides, and open a wide pore through the transporter, exposing Na$^+$-binding sites and X motif residues simultaneously to opposite
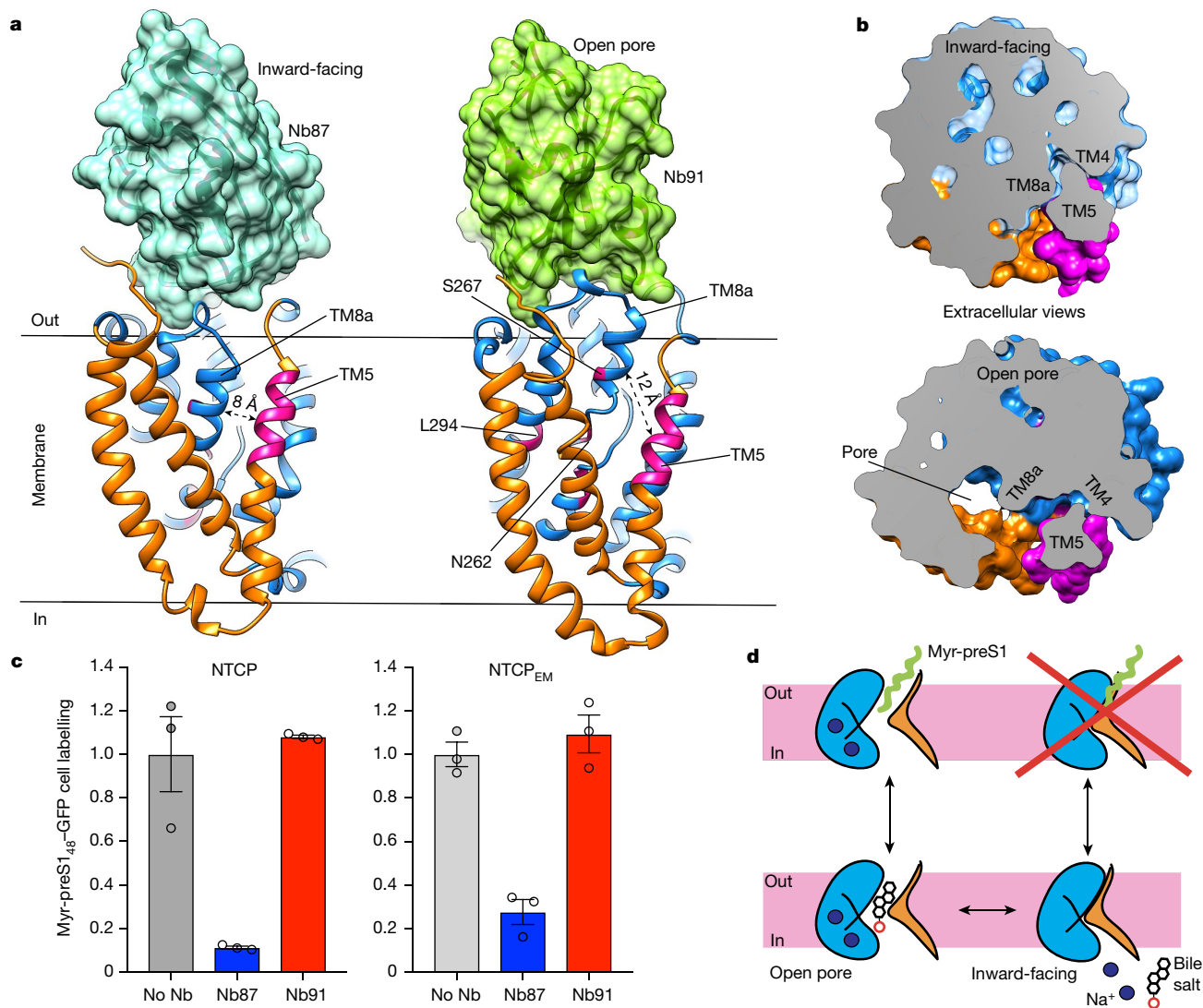
Fig. 4 | Nb87 inhibits myr-preS1 binding. a, Nb87 (left, cyan surface) and Nb91 (right, green surface) bind overlapping 3D epitopes on the extracellular surface of the core domain, distant from the myr-preS1 binding-determinant region in TM5 and residues within the pore (highlighted in pink). In the inward-facing state (left), TM5 is packed against the core domain (blue). In the open-pore state (right), the core domain moves outward and away from TM5, exposing important residues for myr-preS1 binding (pink). **b**, An extracellular view of cross-sections passing through the myr-preS1 binding-determinant region in TM5 (highlighted in pink). Inward-facing (top) and open-pore (bottom)

structures show the TM5–core domain (blue) interfaces. **c**, Myr-preS1$_{48}$–GFP labelling of cells expressing wild-type NTCP (left) and NTCP$_{EM}$ (right), respectively. Pre-incubation with Nb87, but not with Nb91, impaired myr-preS1$_{48}$–GFP labelling. Plots depict the mean of three biologically independent experiments and circles show values from individual experiments. Error bars represent s.e.m. **d**, Cartoon representation of the NTCP gated-pore transport mechanism and the relative movements of the core (blue) and panel (orange) domains. The myr-preS1 domain of HBV/HDV (green) preferentially binds to open-to-outside states of the NTCP transport cycle.

sides of the membrane. This is an unexpected conformational transition, as active transporters typically alternate exposure of their ligand binding sites to the extracellular and intracellular milieus, and adopt intermediate states with substrates occluded within the protein[38,39]. We discuss a plausible NTCP transport mechanism including an open-pore state below.

The surface lining the pore is amphiphilic, and most polar residues in this surface come from the core domain, including conserved side-chains in the X motif. Human NTCP mutations S267F and S199R—both of which are associated with hypercholanaemia[35,36,40]—also map to this surface, on opposite sides of the membrane. The pore has a minimum diameter of approximately 5 Å, and contains a large volume (2,400 Å^3), with its long axis oriented at an angle of about 45° to the membrane plane. It displays wide openings on both extracellular and intracellular sides to bulk solutions, as well as hydrophobic membrane leaflets. The amino acid conservation, the architecture and the amphiphilic

nature of the pore strongly suggest that it is the pathway for translocation of a wide range of amphiphilic bulky substrates transported by NTCP, including bile salts[7,41], sulfated steroids[42,43] and statins[14,15]. Consistently, in the cryo-EM map of NTCP$_{EM}$–Mb91, we observed extra density that partially occupies the pore on the cytoplasmic side, wedged in the crevice between TM6 and TM9 and in close proximity to conserved residues in the X motifs (Fig. 3a, Extended Data Fig. 7). Our cryo-EM sample included both Na$^+$ and substrate taurocholate, and the density probably corresponds to a substrate molecule bound to NTCP$_{EM}$. However, the lack of molecular features in the density precluded unambiguous determination of the bound molecule.

It is worth noting that the NTCP$_{EM}$–Mb91 complex structure was determined from samples in detergent solutions, raising the possibility that detergent molecules could have facilitated the open-pore state. To shed light on this question, we determined the cryo-EM structure of NTCP$_{EM}$–Nb91 complex reconstituted in nanodiscs. Despite the limited

resolution of the cryo-EM map (approximately 4.3 Å), we were able to confidently model NTCP_EM in a conformation nearly identical to that observed in detergent solutions (r.m.s.d. ≈ 1.4 Å) (Extended Data Fig. 8), demonstrating that NTCP_EM adopts an open-pore state in a lipid bilayer, which therefore represents a functional state of the transport cycle. We also observed similar extra density localized to the pore in the nanodisc-reconstituted NTCP_EM–Nb91 complex (Extended Data Fig. 7), further supporting the idea that the density corresponds to a substrate molecule, rather than detergent bound to the transporter.

Nb87 impairs myr-preS1 binding

The conformational changes associated with NTCP_EM pore opening have implications for the HBV/HDV receptor-recognition mechanism. A reported critical region for myr-preS1 binding and viral infection[2] (NTCP residues K157–L165) maps to the extracellular half of TM5 in the panel domain, and localizes far (more than 20 Å) from both Nb87- and Nb91- binding interfaces on the surface of the core domain (Fig. 4a). Notably, there is a conformational change around TM5 when comparing NTCP_EM inward-facing and open-pore states (Fig. 4a, b, Supplementary Video 2). In the inward-facing state, TM5 packs tightly against TM4 and TM8b in the core, generating a shallow groove at the interdomain interface lined by hydrophobic residues. By contrast, relative movements of core and panel domains and tilting of TM5 towards the membrane in the open-pore state unpack the extracellular part of this helix away from the core domain (by as much as 4 Å), creating a crevice between TM5 on one side, and TM8b and the X motif on the other. Moreover, pore-lining residues that impair both myr-preS1 binding and bile salt transport[32] (including N262, S267 and L294) are accessible to the outside only in the open-pore state. The changes in accessibility around critical regions for HBV/HDV recognition suggests that myr-preS1 may bind differentially to open-pore and inward-facing states.

To test this hypothesis, we optimized a fluorescence-based myr-preS1 binding assay in cells using a purified myr-preS1_48 lipopeptide fused to GFP (myr-preS1_48–GFP). Indeed, myr-preS1_48–GFP labelling of cells expressing wild-type NTCP or NTCP_EM was greatly decreased in the presence of Nb87, but was not affected by Nb91 (Fig. 4c). Nb87- and Nb91-overlapping epitopes on the surface of the core domain distant from HBV/HDV binding determinants strongly indicate that the inhibitory effect of Nb87 on myr-preS1_48–GFP binding is not owing to direct steric hindrance, but rather to stabilization of the inward-facing state that allosterically buries myr-preS1 binding determinants in the protein core. Overall, structural and functional results indicate that myr-preS1 binds preferentially to the open-pore state and interacts with exposed residues lining the pore at the interface between core and panel domains.

Discussion

Our structural and functional analyses of NTCP_EM in complexes with conformation-specific nanobodies reveal key molecular aspects of NTCP transport and HBV/HDV receptor-recognition mechanisms.

The NTCP_EM open-pore structure is apparently at odds with the alternating-access transport mechanisms observed in most solute carrier families[38,39], including prokaryotic homologues of SLC10[27,28], which involve occluded substrate-bound intermediates of the transport cycle, raising the question of how to reconcile an open-pore intermediate state with thermodynamically active transport. Our structures suggest a plausible mechanism in which the pore is transiently open in the presence of substrate (and thermodynamically coupled Na[+]) and closes upon release of ligands into the cytoplasm in the inward-facing state (Fig. 4d). The presence of an additional cryo-EM density in the pore, probably representing a bile salt molecule bound to the transporter, supports this type of mechanism. Moreover, sodium ions would contribute to gate the pore, avoiding bile salt permeation down its

electrochemical gradient and preferentially enabling bile salt binding at high extracellular sodium concentrations from the outside (under physiological ionic gradients), for instance by inducing outward-facing states that resemble those observed in prokaryotic homologues of SLC10[28]. Consistent with this line of thinking, early ion transport theories considered active carriers as pores whose gates are controlled by the energy source[44,45], challenging the classical distinction between channels and pumps[46]. To our knowledge, the NTCP open-pore state is the first structural demonstration of an active transporter displaying a wide-open-pore transport pathway for a bulky solute. Detailed knowledge of how the NTCP pore is gated on the extracellular side will require further structural and biophysical work.

The NTCP_EM open-pore structure further shows that HBV/HDV-binding determinants line the pore within the membrane plane, accessible to the outside, and overlap with the substrate transport pathway. By sharp contrast, the inward-facing state shows tight packing of core and panel domains on the extracellular side burying virus-binding determinant residues within the protein core, and consistently, Nb87 antagonizes myr-preS1 binding. These results converge to suggest that myr-preS1 interacts with residues in the pore and, hence, that HBV/HDV selectively recognize NTCP conformations with an open-to-outside substrate pathway, while binding to inward-facing states is impaired (Fig. 4d). Such a mechanism explains the reported inhibitory effect of myr-preS1 binding on bile salt transport[32], as bound myr-preS1 would stabilize open-to-outside states and preclude isomerization to inward-facing ones and the antagonism between myr-preS1 and substrate binding[32], as both ligands would interact with overlapping binding sites within the pore.

The inhibitory effect of Nb87 on myr-preS1 binding reveals the therapeutic potential of molecules that stabilize NTCP inward-facing state(s), as allosteric inhibitors of viral cell entry. Such molecules could constitute alternative and/or synergistic therapeutic tools to existing lipopeptides that mimic high-affinity myr-preS1 binding[23,47], as well as neutralizing antibodies against HBV[48,49].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04723-z.

1. Doring, B., Lutteke, T., Geyer, J. & Petzinger, E. The SLC10 carrier family: transport functions and molecular structure. *Curr. Top. Membr.* **70**, 105–168 (2012).
2. Yan, H. et al. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *eLife* **1**, e00049 (2012).
3. Ni, Y. et al. Hepatitis B and D viruses exploit sodium taurocholate co-transporting polypeptide for species-specific entry into hepatocytes. *Gastroenterology* **146**, 1070–1083 (2014).
4. Lempp, F. A. & Urban, S. Inhibitors of hepatitis B virus attachment and entry. *Intervirology* **57**, 151–157 (2014).
5. Hagenbuch, B. & Dawson, P. The sodium bile salt cotransport family SLC10. *Pflugers Arch.* **447**, 566–570 (2004).
6. Claro da Silva, T., Polli, J. E. & Swaan, P. W. The solute carrier family 10 (SLC10): beyond bile acid transport. *Mol. Aspects Med.* **34**, 252–269 (2013).
7. Hagenbuch, B. & Meier, P. J. Molecular cloning, chromosomal localization, and functional characterization of a human liver Na[+]/bile acid cotransporter. *J. Clin. Invest.* **93**, 1326–1331 (1994).
8. Wong, M. H., Oelkers, P., Craddock, A. L. & Dawson, P. A. Expression cloning and characterization of the hamster ileal sodium-dependent bile acid transporter. *J. Biol. Chem.* **269**, 1340–1347 (1994).
9. Kramer, W. Transporters, Trojan horses and therapeutics: suitability of bile acid and peptide transporters for drug delivery. *Biol. Chem.* **392**, 77–94 (2011).
10. Tolle-Sander, S., Lentz, K. A., Maeda, D. Y., Coop, A. & Polli, J. E. Increased acyclovir oral bioavailability via a bile acid conjugate. *Mol. Pharm.* **1**, 40–48 (2004).
11. Kullak-Ublick, G. A. et al. Chlorambucil-taurocholate is transported by bile acid carriers expressed in human hepatocellular carcinomas. *Gastroenterology* **113**, 1295–1305 (1997).
12. Bhat, L., Jandeleit, B., Dias, T. M., Moors, T. L. & Gallop, M. A. Synthesis and biological evaluation of novel steroidal pyrazoles as substrates for bile acid transporters. *Bioorg. Med. Chem. Lett.* **15**, 85–87 (2005).

Article

13. Balakrishnan, A. & Polli, J. E. Apical sodium dependent bile acid transporter (ASBT, SLC10A2): a potential prodrug target. *Mol. Pharm.* **3**, 223–230 (2006).
14. Ho, R. H. et al. Drug and bile acid transporters in rosuvastatin hepatic uptake: function, expression, and pharmacogenetics. *Gastroenterology* **130**, 1793–1806 (2006).
15. Fujino, H., Saito, T., Ogawa, S. & Kojima, J. Transporter-mediated influx and efflux mechanisms of pitavastatin, a new inhibitor of HMG-CoA reductase. *J. Pharm. Pharmacol.* **57**, 1305–1311 (2005).
16. Roscam Abbing, R. L. P. et al. Blocking sodium-taurocholate cotransporting polypeptide stimulates biliary cholesterol and phospholipid secretion in mice. *Hepatology* **71**, 247–258 (2020).
17. Donkers, J. M. et al. Inhibition of hepatic bile acid uptake by myrcludex B promotes glucagon-like peptide-1 release and reduces obesity. *Cell. Mol. Gastroenterol. Hepatol.* **10**, 451–466 (2020).
18. Levrero, M., Testoni, B. & Zoulim, F. HBV cure: why, how, when? *Curr. Opin. Virol.* **18**, 135–143 (2016).
19. Zeisel, M. B. et al. Towards an HBV cure: state-of-the-art and unresolved questions-report of the ANRS workshop on HBV cure. *Gut* **64**, 1314–1326 (2015).
20. Engelke, M. et al. Characterization of a hepatitis B and hepatitis delta virus receptor binding site. *Hepatology* **43**, 750–760 (2006).
21. Le Seyec, J., Chouteau, P., Cannie, I., Guguen-Guillouzo, C. & Gripon, P. Infection process of the hepatitis B virus depends on the presence of a defined sequence in the pre-S1 domain. *J. Virol.* **73**, 2052–2057 (1999).
22. Blanchet, M. & Sureau, C. Infectivity determinants of the hepatitis B virus pre-S domain are confined to the N-terminal 75 amino acid residues. *J. Virol.* **81**, 5841–5849 (2007).
23. Gripon, P., Cannie, I. & Urban, S. Efficient inhibition of hepatitis B virus infection by acylated peptides derived from the large viral surface protein. *J. Virol.* **79**, 1613–1622 (2005).
24. Petersen, J. et al. Prevention of hepatitis B virus infection in vivo by entry inhibitors derived from the large envelope protein. *Nat. Biotechnol.* **26**, 335–341 (2008).
25. Schulze, A., Schieck, A., Ni, Y., Mier, W. & Urban, S. Fine mapping of pre-S sequence requirements for hepatitis B virus large envelope protein-mediated receptor interaction. *J. Virol.* **84**, 1989–2000 (2010).
26. Glebe, D. et al. Mapping of the hepatitis B virus attachment site by use of infection-inhibiting preS1 lipopeptides and tupaia hepatocytes. *Gastroenterology* **129**, 234–245 (2005).
27. Hu, N. J., Iwata, S., Cameron, A. D. & Drew, D. Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT. *Nature* **478**, 408–411 (2011).
28. Zhou, X. et al. Structural basis of the alternating-access mechanism in a bile acid transporter. *Nature* **505**, 569–573 (2014).
29. Cirri, E. et al. Consensus designs and thermal stability determinants of a human glutamate transporter. *eLife* **7**, e40110 (2018).
30. Uchanski, T. et al. Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. *Nat. Methods* **18**, 60–68 (2021).
31. Zahner, D., Eckhardt, U. & Petzinger, E. Transport of taurocholate by mutants of negatively charged amino acids, cysteines, and threonines of the rat liver sodium-dependent taurocholate cotransporting polypeptide NTCP. *Eur. J. Biochem.* **270**, 1117–1127 (2003).
32. Yan, H. et al. Viral entry of hepatitis B and D viruses and bile salts transportation share common molecular determinants on sodium taurocholate cotransporting polypeptide. *J. Virol.* **88**, 3273–3284 (2014).
33. Weinman, S. A., Carruth, M. W. & Dawson, P. A. Bile acid uptake via the human apical sodium-bile acid cotransporter is electrogenic. *J. Biol. Chem.* **273**, 34691–34695 (1998).
34. Hagenbuch, B. & Meier, P. J. Sinusoidal (basolateral) bile salt uptake systems of hepatocytes. *Semin. Liver Dis.* **16**, 129–136 (1996).
35. Ho, R. H., Leake, B. F., Roberts, R. L., Lee, W. & Kim, R. B. Ethnicity-dependent polymorphism in Na⁺-taurocholate cotransporting polypeptide (SLC10A1) reveals a domain critical for bile acid substrate recognition. *J. Biol. Chem.* **279**, 7213–7222 (2004).
36. Liu, R. et al. Homozygous p.Ser267Phe in SLC10A1 is associated with a new type of hypercholanemia and implications for personalized medicine. *Sci. Rep.* **7**, 9214 (2017).
37. Pan, W. et al. Genetic polymorphisms in Na⁺-taurocholate co-transporting polypeptide (NTCP) and ileal apical sodium-dependent bile acid transporter (ASBT) and ethnic comparisons of functional variants of NTCP among Asian populations. *Xenobiotica* **41**, 501–510 (2011).
38. 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 (2016).
39. Drew, D. & Boudker, O. Shared molecular mechanisms of membrane transporters. *Annu. Rev. Biochem.* **85**, 543–572 (2016).
40. Li, H. et al. Clinical and molecular characterization of four patients with NTCP deficiency from two unrelated families harboring the novel SLC10A1 variant c.595A>C (p.Ser199Arg). *Mol. Med. Rep.* **20**, 4915–4924 (2019).
41. Novak, D. A., Ryckman, F. C. & Suchy, F. J. Taurocholate transport by basolateral plasma membrane vesicles isolated from human liver. *Hepatology* **10**, 447–453 (1989).
42. Craddock, A. L. et al. Expression and transport properties of the human ileal and renal sodium-dependent bile acid transporter. *Am. J. Physiol.* **274**, G157–G169 (1998).
43. Han, Y. H. et al. Transporter studies with the 3-O-sulfate conjugate of 17α-ethinylestradiol: assessment of human kidney drug transporters. *Drug Metab. Dispos.* **38**, 1064–1071 (2010).
44. Lauger, P. Kinetic properties of ion carriers and channels. *J. Membr. Biol.* **57**, 163–178 (1980).
45. West, I. C. Ligand conduction and the gated-pore mechanism of transmembrane transport. *Biochim. Biophys. Acta* **1331**, 213–234 (1997).
46. Ashcroft, F., Gadsby, D. & Miller, C. The blurred boundary between channels and transporters. *Philos. Trans. R. Soc. B* **364**, 145–147 (2009).
47. Cheng, D., Han, B., Zhang, W. & Wu, W. Clinical effects of NTCP-inhibitor myrcludex B. *J. Viral Hepat.* **28**, 852–858 (2021).
48. Hehle, V. et al. Potent human broadly neutralizing antibodies to hepatitis B virus from natural controllers. *J. Exp. Med.* **217**, e20200840 (2020).
49. Corti, D., Benigni, F. & Shouval, D. Viral envelope-specific antibodies in chronic hepatitis B virus infection. *Curr. Opin. Virol.* **30**, 48–57 (2018).

Methods

Thermostable NTCP constructs

Consensus amino acids were calculated using JALVIEW[50] and reported criteria[29] from sequences of representative NTCP vertebrate orthologues (Extended Data Fig. 1), aligned using Muscle[51]. Consensus amino acid exchanges were simultaneously introduced into wild-type NTCP sequence background with *N*-glycosylation mutations N5T and N11T, improving protein stability. Deletions of N-terminal residue E2, and the unstructured C terminus (residues T329–A349) in the consensus non-glycosylated construct further improved homogeneity of the sample, yielding the so-called $NTCP_{CO}$.

In general, the consensus approach generates protein samples with overall improved stability, but it is expected that by simultaneously introducing all consensus mutations, some destabilizing exchanges are included. To minimize the latter, we probed thermal stability of single-point $NTCP_{CO}$ mutants, in which we reverted consensus amino acids to the wild-type residues, using fluorescence-detection size-exclusion chromatography[52] (SEC). Removal of destabilizing consensus exchanges in $NTCP_{CO}$, yielded a consensus design, $NTCP_{EM}$, which is nearly identical to wild-type NTCP (approximately 98% identity) (Extended Data Figs. 1, 6), while preserving Na^+-dependent bile salt transport as well as myr-preS1 recognition mechanisms.

Protein expression and purification

cDNAs encoding NTCP constructs were synthesized (GenScript) and subcloned into a pcDNA3.1(+) vector encompassing a C-terminal PreScission site, followed by GFP, and two Strep-tags in tandem for affinity purification. Protein expression was done in HEK 293F cells (Thermo Fisher; cells were not authenticated or tested for mycoplasma contamination) by transient transfection, as described[53] with small variations. In brief, cells grown in FreeStyle 293 medium (Thermo Scientific) were transfected with linear 25K polyethyleneimine (PEI) (Polysciences) at a cell density of 2.5×10^6 cells per ml using 3 µg ml⁻¹ DNA. Valproic acid (VPA) was added to the culture at a final concentration of 2.2 mM 6–12 h after transfection and cells were grown for additional 48 h before collection.

Cell pellets were resuspended and lysed in buffer containing 50 mM HEPES pH 7.4, 200 mM NaCl, 5% v/v glycerol, 1 mM EDTA, 1 mM TCEP, 0.5 mM sodium taurocholate and supplemented with protease inhibitors (1 mM PMSF and protease inhibitor cocktail from Sigma), 1% dodecyl-β-D-maltopyranoside (DDM) (Anatrace) and 0.2% cholesteryl hemi-succinate tris salt (CHS) (Anatrace), and incubated for 1 h. Cell debris was removed by ultracentrifugation. Detergent-solubilized transporters were purified by affinity chromatography using streptactin sepharose resin (Cytiva Life Sciences). Resin was pre-equilibrated in buffer A containing 50 mM HEPES pH 7.4, 200 mM NaCl, 5% v/v glycerol, 0.017% DDM, 0.0034% CHS, and 0.2 mM sodium taurocholate, and incubated with transporters for 1 h under rotation. Resin was extensively washed with buffer A, and protein was eluted in buffer B containing 50 mM HEPES pH 7.4, 200 mM NaCl, 5% v/v glycerol, 0.017% DDM, 0.0034% CHS, 0.2 mM sodium taurocholate, and 2.5 mM desthiobiotin. The eluted protein was digested with PreScission protease overnight, concentrated to several mg per ml using 100 kDa MWCO concentrator (Corning Spin-X UF concentrators) and injected in a Superose 6 column (GE Healthcare Life Sciences) using SEC buffer containing 20 mM HEPES pH 7.4, 100 mM NaCl, 0.017% DDM, 0.0034% CHS, and 0.2 mM sodium taurocholate. Purified transporters were used immediately or flash frozen and stored at −80 °C. All purification steps were done at 4 °C.

$NTCP_{EM}$ complexes with nanobodies and megabodies, respectively, were formed by mixing purified protein samples at 1:1.2 (transporter:nanobody, or megabody) molar ratio, and incubated for 2h at 4 °C. Excess nanobody or megabody was removed by SEC using SEC buffer. MSP1D1 nanodisc-scaffold protein was expressed and purified using published protocols[54]. Reconstitution was done by mixing purified $NTCP_{EM}$–Nb and $NTCP_{EM}$–Mb complexes, respectively, with MSP1D1 and liver total lipid extract (Avanti Polar Lipids) at 0.1:1:15 molar ratio, and incubated with methanol-activated biobeads for 2 h. Biobeads were exchanged once, and the mixture was further incubated overnight. Nanodisc-reconstituted sample was purified in a Superdex 200 increase column (GE Healthcare Life Sciences) in buffer containing 20 mM HEPES pH 7.4, 100 mM NaCl, and 0.2 mM sodium taurocholate. Samples were concentrated as described above, and immediately used for cryo-EM grid preparation.

Nanobody generation, expression and purification

Nanobodies against $NTCP_{CO}$ were generated using published protocols[55]. In brief, one llama (*Lama glama*) was six times immunized with a total 0.9 mg of $NTCP_{CO}$ reconstituted in proteoliposomes. Four days after the final boost, blood was taken from the llama to isolate peripheral blood lymphocytes. RNA was purified from these lymphocytes and reverse transcribed by PCR to obtain the cDNA of the open reading frames coding for the nanobodies. The resulting library was cloned into the phage display vector pMESy4 bearing a C-terminal His₆ tag and a CaptureSelect sequence tag (Glu-Pro-Glu-Ala). Different nanobody families, as defined by the difference in the CDR3, were selected by biopanning. For this, $NTCP_{CO}$ reconstituted in proteoliposomes was solid phase coated directly on plates. $NTCP_{CO}$ specific phage were recovered by limited trypsinization, and after two rounds of selection, periplasmic extracts were made and analysed using ELISA screens. Nb87 and Nb91 were expressed in *Escherichia coli* for subsequent purification from the bacterial periplasm. After Ni-NTA (Sigma) affinity purification, nanobodies were further purified by SEC in buffer: 10 mM HEPES pH 7.4, and 110 mM NaCl.

Nb91 was enlarged by fusion to the circular permutated glucosidase of *E. coli* K12 (YgjK, 86 kDa) to build the megabody referred to as Mb91. Mb91 was generated and purified using previously described protocols[30].

Fluorescent substrate analogue transport assay

Sodium-dependent substrate uptake was measured in HEK 293F cells transfected with 2 µg µl⁻¹ cDNA using the above-mentioned protocol with small modifications. Forty-eight h after transfection, around 1 million cells were pelleted, washed, and resuspended in 500 µl of transport buffer (110 mM NaCl, 4 mM KCl, 1 mM $MgSO_4$, 1 mM $CaCl_2$, 45 mM mannitol, 5 mM glucose and 10 mM HEPES pH 7.4), or control buffer in which NaCl was substituted with choline chloride (ChCl). To probe the effect of nanobodies on bile salt transport, cells were incubated with nanobodies for 1.5 h, followed by addition of the fluorescent substrate analogue tauro-nor-THCA-24-DBD[56,57] (tebu-bio) to a final concentration of 10 µM for 30 min at 37 °C. Excess fluorescent analogue was removed by centrifugation (13,000*g* for 30 s), and 1 wash with the above-mentioned control buffer. Then, cells were resuspended and lysed using Pierce IP lysis buffer (Thermo Fisher). Finally, lysates were centrifuged (13,000*g* for 10 min), and transferred to black 96-well flat-bottom plates (Grenier), and quantified by fluorescence in a micro-plate reader (CLARIOstar-Plus) using excitation at 454 nm and emission of 570 nm. Three biologically independent experiments were quantified in triplicate samples. Nb titrations data were fitted in Prism 8.0.1 (GraphPad) to the following dose-respond curve:

$$\text{Fractional transport} = 1 + \frac{Y_{\min} - 1}{1 + 10^{\log IC_{50} - x}}$$

Where $Y_{\min}$ corresponds to fraction of transport at saturating Nb concentrations, IC_{50} is the half-maximal inhibitory concentration, and x is log[Nb].

Myr-preS1 purification and binding assay

cDNA encoding the N-terminal myristoylated consensus residues 2–48 of human HBV myr-preS1 domain (myr-GTNLSVPNPLGFFPDHQL DPAFRANSNNPDWDFNPNKDHWPEANKVG) was synthesized (GenScript)

Article

and subcloned into a pcDNA3.1(+) vector encompassing a C-terminal GFP, and poly Histidine-tag (namely, myr-preS1$_{48}$–GFP). Myr-preS1$_{48}$–GFP was expressed in HEK 293F cells (Thermo Fisher) by transient transfection, as described for expression of NTCP$_{EM}$ purification. Cells were lysed by 3–5 passes through a homogenizer (EmulsiFlex-C5, Avestin) and membrane fraction was collected by ultracentrifugation. Membranes were resuspended in a buffer containing 50 mM HEPES pH 7.4, 200 mM NaCl, 5% v/v glycerol, 1 mM EDTA, protease inhibitors (1 mM PMSF and protease-inhibitor cocktail from Sigma), 1% dodecyl-β-D-maltopyranoside (Anatrace), and 0.2% cholesteryl hemi-succinate tris salt (Anatrace), and incubated for 1 h. Solubilized myr-preS1$_{48}$–GFP was subjected to ultracentrifugation and then purified by affinity chromatography using anti-His affinity resin (Sigma). Resin was pre-equilibrated in a buffer containing 50 mM HEPES pH 7.4, 200 mM NaCl, 5% v/v glycerol, 0.013% DDM, 0.0027% CHS, and incubated with detergent-solubilized myr-preS1$_{48}$–GFP for 1 h under rotation. Resin was extensively washed with buffer containing 50 mM HEPES pH 7.4, 200 mM NaCl, 5% v/v glycerol, 0.013% DDM, 0.0027% CHS and 50 mM imidazole. Myr-preS1$_{48}$–GFP was eluted in buffer containing 50 mM HEPES pH 7.4, 200 mM NaCl, 5% v/v glycerol, 0.013% DDM, 0.0027% CHS, and 250 mM imidazole. The eluted protein was concentrated to several mg ml^{-1} using a 30 kDa MWCO concentrator (Corning Spin-X UF concentrators) and injected into a Superose 6 column (GE Healthcare Life Sciences) using a SEC buffer containing 20 mM HEPES pH 7.4, 200 mM NaCl, 0.013% DDM, and 0.0027% CHS. Myristoylation of the sample was confirmed by mass spectrometry. Purified myr-preS1$_{48}$–GFP was flash frozen and stored at −80 °C. All purification steps were done at 4 °C.

Myr-preS1$_{48}$–GFP binding to NTCP constructs was assayed in HEK 293F cells, grown and transfected with 1 μg ml^{-1} DNA using the protocol described above. Forty-eight hours after transfection, cells were washed with pre-warmed PBS, and ~1 million cells were pelleted and resuspended in 1 ml of PBS. To probe the effect of nanobodies, cells expressing NTCP constructs were pre-incubated with 10 μM nanobodies for 1.5 h. They were then labelled with 10 nM (wild-type NTCP) or 50 nM (NTCP$_{EM}$) purified myr-preS1$_{48}$–GFP for 30 min. Excess fluorescent-probe was removed by centrifugation (13,000g for 30 s), and one wash with PBS. Cells were then re-suspended in PBS and GFP fluorescence was recorded in a micro-plate reader (CLARIOstar-Plus) using excitation at 470 nm and emission at 508 nm.

Electron microscopy sample preparation and data acquisition

Purified NTCP$_{EM}$–Nb or –Mb complexes were applied to glow-discharged Au 300 mesh Quantifoil R1.2/1.3. Typically, 4 μl of sample at 3–4 mg ml^{-1} was applied to the grids, and the Vitrobot chamber was maintained at 100% humidity and 4 °C. Grids were screened in 200 kV Talos Arctica microscope (ThermoFisher) at the IECB cryo-EM imaging facility. Final data collection was performed in 300 kV Titan Krios microscope (ThermoFisher) at EMBL-Heidelberg Cryo-Electron Microscopy Service Platform, equipped with K3 direct electron detector (Gatan). Final images were recorded with SerialEM[58] at a pixel size of 0.504 Å. Dose rate was 15–20 e⁻ pixel s^{-1}.

For NTCP$_{EM}$–Nb87 complexes, 21,792 images were recorded with −0.5 to −1.5 μm defocus range. Images were collected with 0.7-s subframes (total 40 subframes), corresponding to a total dose of 57.8 e⁻ Å^{-2}. For NTCP$_{EM}$–Mb91 complexes, 21,390 images were recorded with −0.6 to −1.75-μm defocus range. Images were collected with 0.7 s subframes (total 40 subframes), corresponding to a total dose of 56.5 e⁻ Å^{-2}.

Cryo-EM data processing, model building and structure analysis

All datasets were processed with cryoSPARC v2 and v3[59]. Movies were gain corrected, and aligned using in-built patch-motion correction routine. Contrast transfer function (CTF) parameters were estimated using the in-built patch-CTF routine in cryoSPARC. Low-quality images were discarded manually upon visual inspection.

For the NTCP$_{EM}$–Mb91 complex, 5,796,802 particles were template-picked from 21,390 micrographs, and selected through several rounds of 2D, as well as 3D ab initio classifications. Particles from 3D ab initio classes displaying interpretable density for transmembrane helices were pooled, and used for homogenous refinement (Extended Data Fig. 2). Cryo-EM density corresponding to both detergent micelle and megabody scaffold were masked out, and particles were further subjected to local refinement using a fulcrum that localized to center of NTCP$_{EM}$ transmembrane region. Focused refinement yielded a final map at an overall resolution of 3.3 Å, based on the gold-standard 0.143 Fourier shell correlation (FSC) cut-off.

For the NTCP$_{EM}$–Nb87 complex, 6,535,687 particles were template-picked from 21,792 micrographs, and classified through several rounds of 2D and 3D ab initio classifications (Extended Data Fig. 3). Around 220,000 selected particles were further classified by heterogenous refinement, yielding a final set of 61,053 particles that were processed by non-uniform refinement[60]. Further focused refinement excluding nanodisc scaffold yielded a final map at an overall resolution of 3.7 Å, based on the gold-standard 0.143 FSC cut-off. Maps were visualized using UCSF Chimera[61] and ChimeraX[62].

The cryo-EM map of the NTCP$_{EM}$–Mb91 complex showed clear density for most sidechains in the transmembrane helices, although TM1 and TM6 in the panel domain displayed fewer molecular features, and was used to build an atomic model of NTCP$_{EM}$ using Coot[63,64]. Secondary structure predictions using Psipred[65] and bacterial homologue structure (Protein Data Bank ID 3ZUY) were used to help initial sequence assignment. Initial Nb models were created with I-TASSER[66], and then fit as rigid bodies into the density, followed by manual building and modification in Coot[63,64]. The inward-facing conformation in the NTCP$_{EM}$–Nb87 complex was built by fitting core and panel domains from NTCP$_{EM}$–Mb91 structure as separate rigid bodies into the density, followed by manual modification in Coot. All atomic models were refined using PHENIX[67].

Structural analyses were carried out as follows: protein cavity calculations with CASTp 3.0[68], pore calculations MOLEonline 2.5[69], protein–protein interfaces with PISA[70], and amino acid conservation surface mapping with ConSurf[71].

Reporting summary

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

Data availability

Structural models of NTCP$_{EM}$–Nb87 and NTCP$_{EM}$–Mb91 complexes have been deposited in the Protein Data Bank (PDB) with accession codes 7PQG and 7PQQ, respectively, and the corresponding cryo-EM maps were deposited in the Electron Microscopy Data Bank (EMDB) under accession numbers EMD-13593 and EMD-13596. Materials are available upon reasonable request and signing of a Material Transfer Agreement.

50. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
51. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
52. Hattori, M., Hibbs, R. E. & Gouaux, E. A fluorescence-detection size-exclusion chromatography-based thermostability assay for membrane protein precrystallization screening. *Structure* **20**, 1293–1299 (2012).
53. Canul-Tec, J. C. et al. Structure and allosteric inhibition of excitatory amino acid transporter 1. *Nature* **544**, 446–451 (2017).
54. Ritchie, T. K. et al. Reconstitution of membrane proteins in phospholipid bilayer nanodiscs. *Methods Enzymol.* **464**, 211–231 (2009).
55. Pardon, E. et al. A general protocol for the generation of nanobodies for structural biology. *Nat. Protoc.* **9**, 674–693 (2014).
56. De Bruyn, T. et al. Confocal imaging with a fluorescent bile acid analogue closely mimicking hepatic taurocholate disposition. *J. Pharm. Sci.* **103**, 1872–1881 (2014).
57. Yamaguchi, K., Murai, T., Yabuuchi, H., Hui, S. P. & Kurosawa, T. Measurement of bile salt export pump transport activities using a fluorescent bile acid derivative. *Drug Metab. Pharmacokinet.* **25**, 214–219 (2010).
58. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

59. 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 (2017).
60. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).
61. Pettersen, E. F. et al. UCSF Chimera-a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
62. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
63. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
64. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
65. Buchan, D. W. A. & Jones, D. T. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019).
66. Yang, J. et al. The I-TASSER suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
67. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
68. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* **46**, W363–W367 (2018).
69. Pravda, L. et al. MOLEonline: a web-based tool for analyzing channels, tunnels and pores (2018 update). *Nucleic Acids Res.* **46**, W368–W373 (2018).
70. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
71. Glaser, F. et al. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **19**, 163–164 (2003).
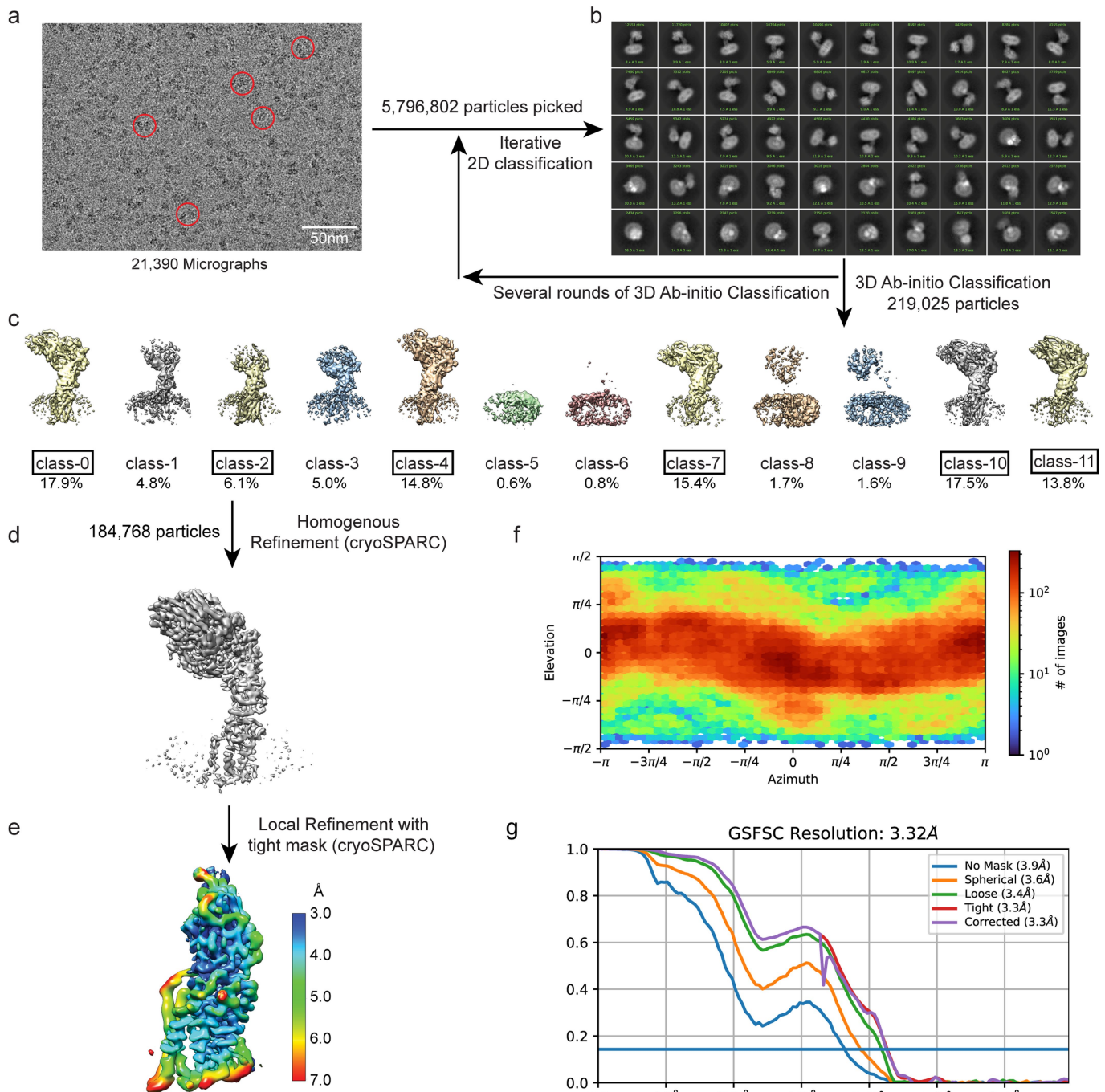
a

b

Extended Data Fig. 1 | NTCP consensus designs. a, Phylogenetic tree of NTCP vertebrate orthologs used to determine consensus amino acids. NCBI protein sequence IDs are given in parenthesis. **b**, Amino acid sequence alignment of human NTCP (residues 3–328) and consensus designs $NTCP_{CO}$ and $NTCP_{EM}$.

Extended Data Fig. 2 | Cryo-EM data processing pipeline of NTCP$_{EM}$-Mb91 complex. a, Representative EM micrograph with examples of individual particles (red circles). 21,390 micrographs were collected. **b**, Gallery of representative 2D class-averages. **c**, 3D classes from *ab initio* classification.

d, Homogenous refined map **e**, Local-refinement map after micelle removal, and color coded according to local resolution estimation. **f**, Viewing direction distribution plot. **g**, Fourier shell correlation (FSC) plot of local refinement with FSC threshold at 0.143.
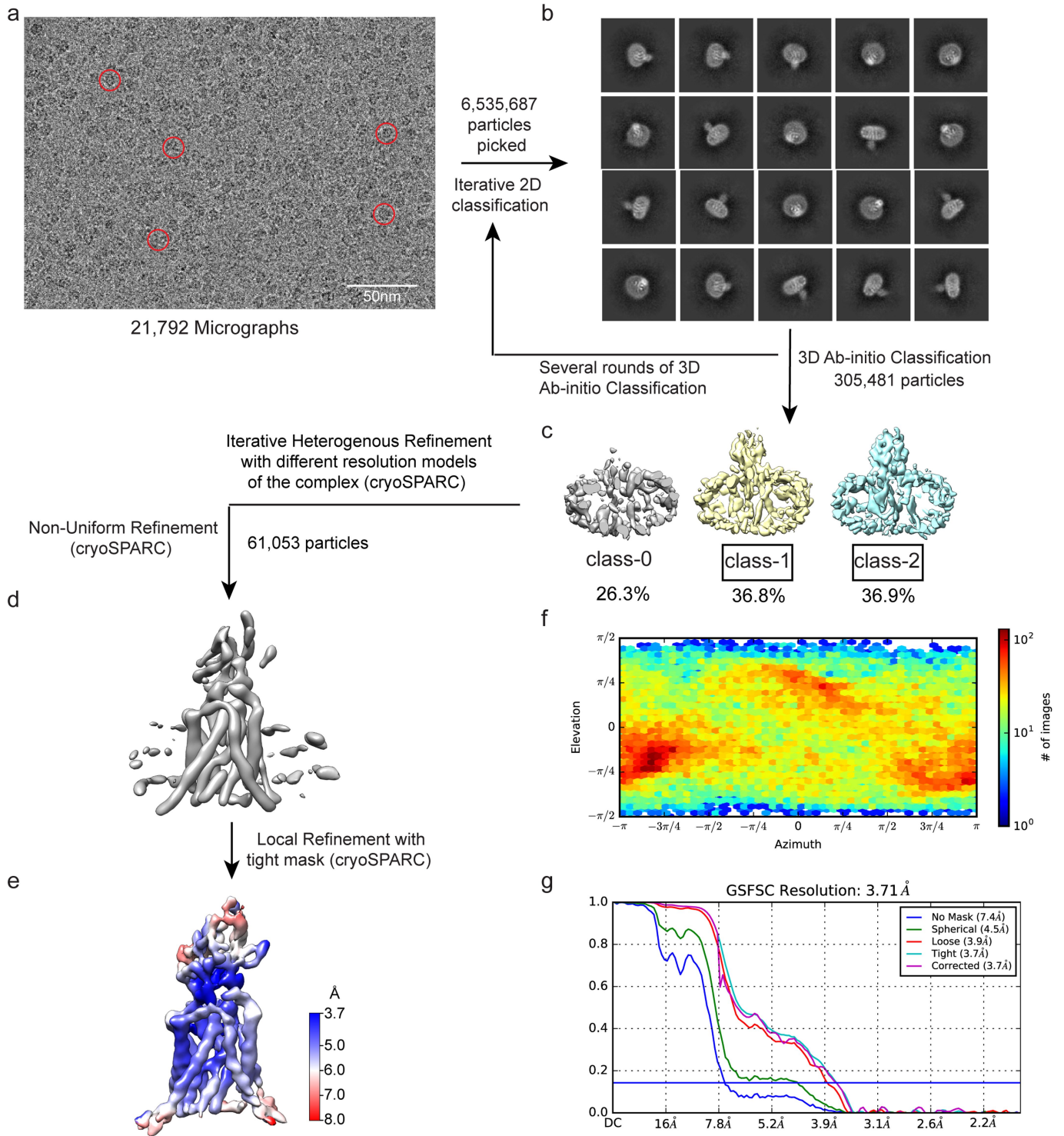
Extended Data Fig. 3 | Cryo-EM data processing pipeline of NTCP$_{EM}$-Nb87 complex. a, Representative EM micrograph with examples of individual particles (red circles). 21,792 micrographs were collected. **b**, Gallery of representative 2D class-averages. **c**, 3D classes from *ab initio* classification. **d**, Non-uniform refinement 3D map. **e**, Local-refinement map after nanodisc removal, and color coded according to local resolution estimation. **f**, Viewing direction distribution plot. **g**, Fourier shell correlation (FSC) plot of local refinement with FSC threshold at 0.143.
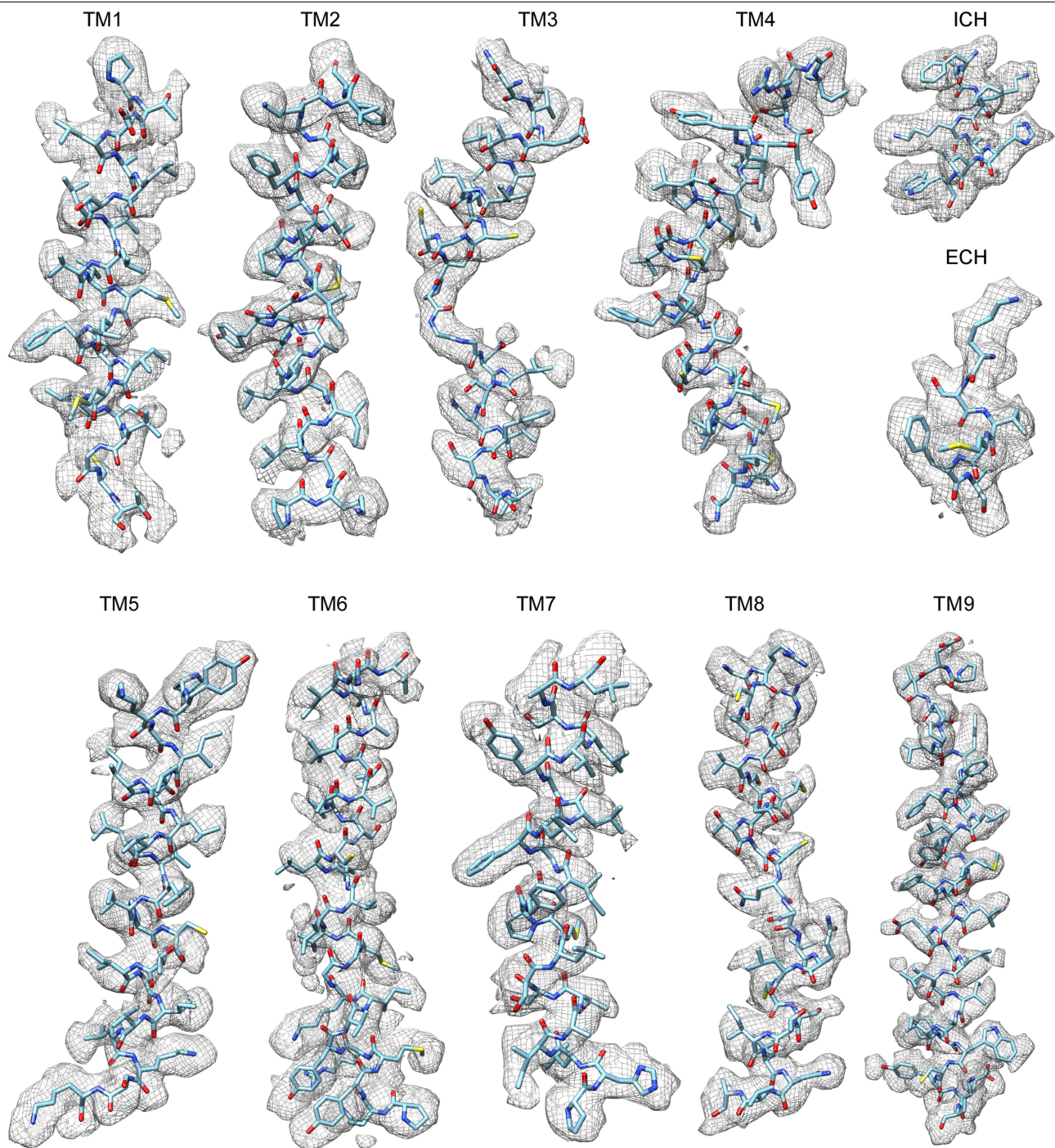
Extended Data Fig. 4 | Cryo-EM density in NTCP$_{EM}$-Mb91 structure. Cryo-EM density corresponding to individual NTCP$_{EM}$ transmembrane helices in complex with Mb91.

Article
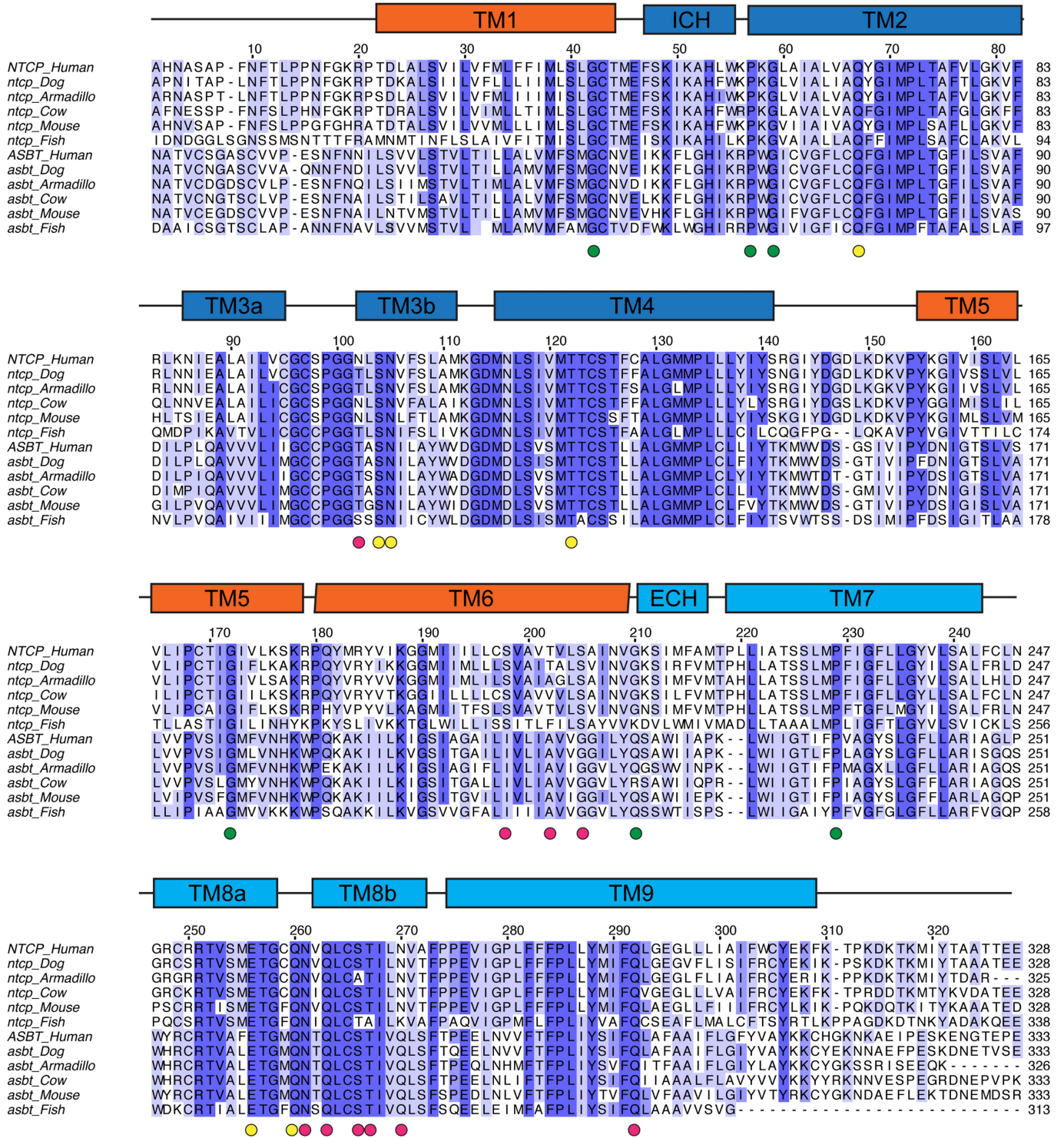
**Extended Data Fig. 5 | Amino acid sequence alignment of NTCP and ASBT vertebrate orthologs.** Yellow circles indicate residues that contribute sidechains to Na1 and Na2, pink circles conserved polar residues in NTCP orthologs that line the space between core and panel domains, and green circles proline and glycine residues that act as hinges during the isomerization between open-pore and inward-facing states.
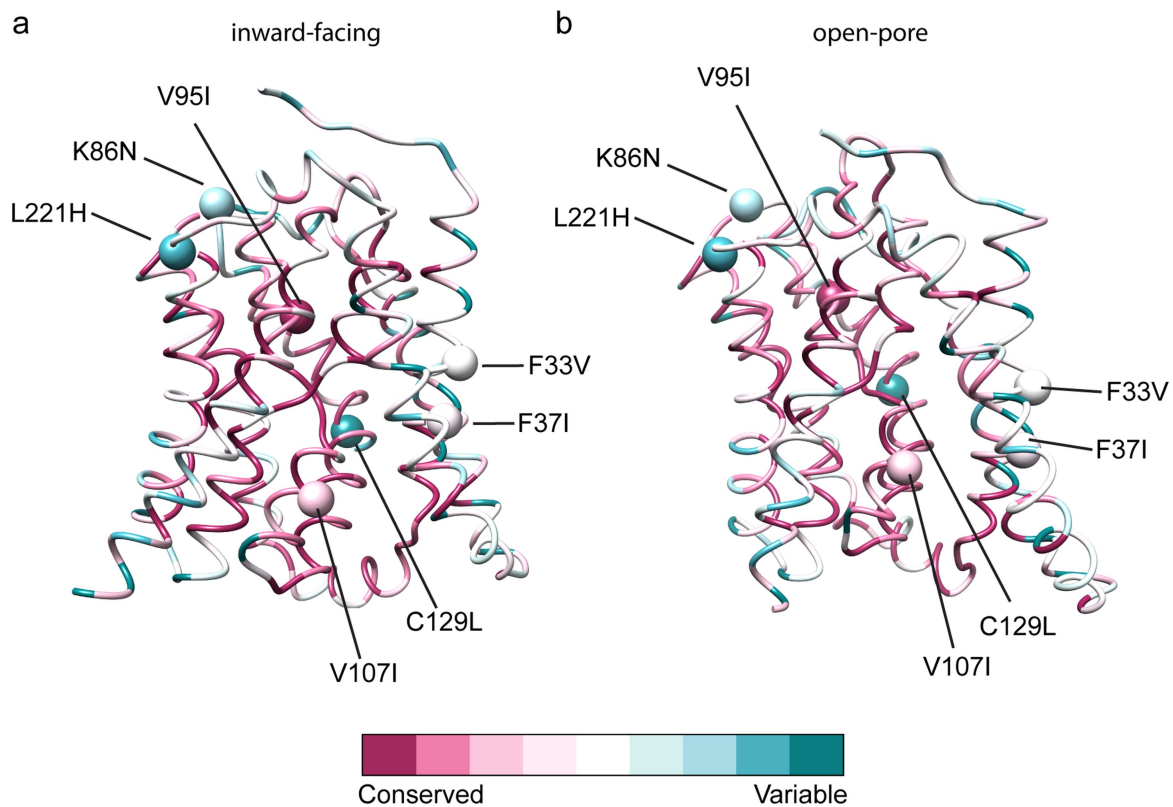
Extended Data Fig. 6 | NTCP amino acid conservation surface mapping. Amino acid conservation across NTCP vertebrate orthologs is mapped into NTCP$_{EM}$ inward-facing (left) and open-pore (right) structures. Spheres represent alpha-carbon atoms of 7 consensus mutations to improve thermal stability in the construct used for cryo-EM analysis (NTCP$_{EM}$).

Article
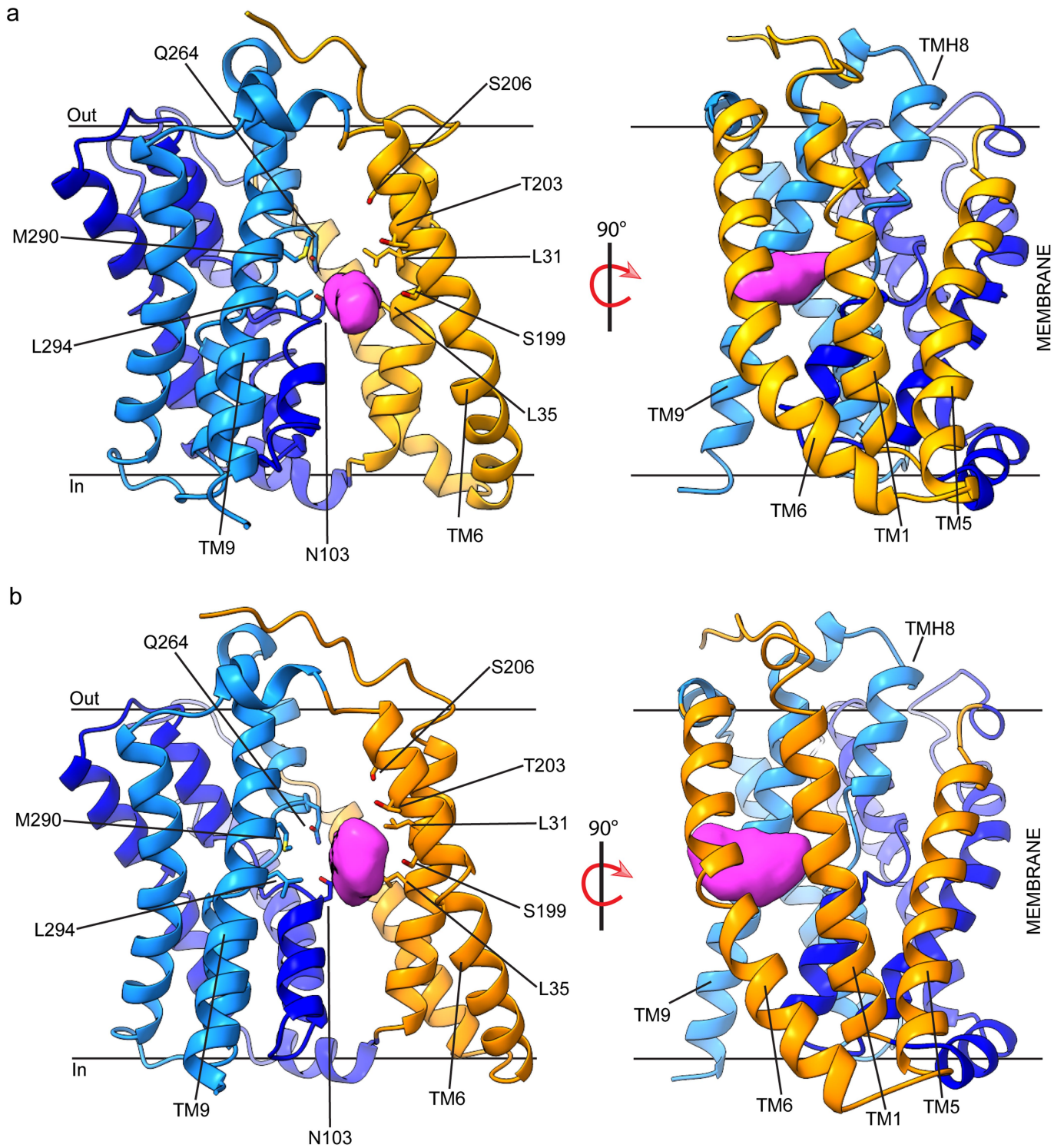
Extended Data Fig. 7 | Extra-density in NTCP$_{EM}$ open-pore structures. **a**, Membrane views of NTCP$_{EM}$-Mb91 in detergent solutions highlighting extra cryo-EM density in the pore (purple surface), and residues in proximity of the density. **b**, Membrane views of NTCP$_{EM}$-Nb91 in nanodisc highlighting similar extra cryo-EM density in the pore, and residues in proximity.

a

b
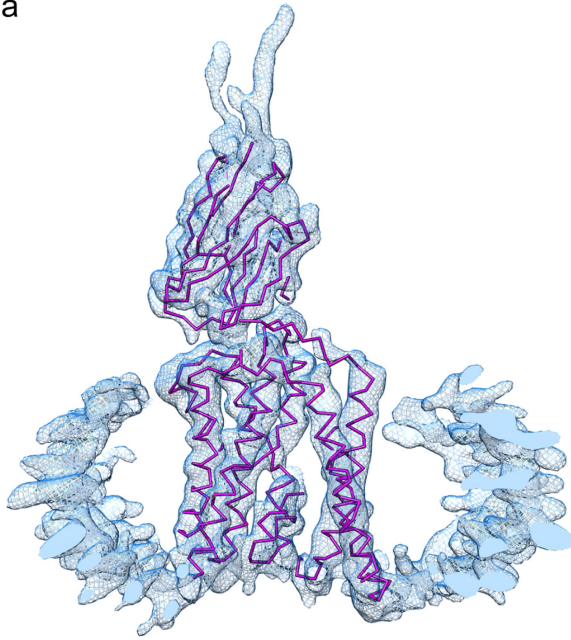
Extended Data Fig. 8 | Cryo-EM density and structure of NTCP$_{EM}$-Nb91 reconstituted in nanodisc. a, Structure of NTCP$_{EM}$-Nb91 complex reconstituted in nanodisc and the corresponding cryo-EM density.
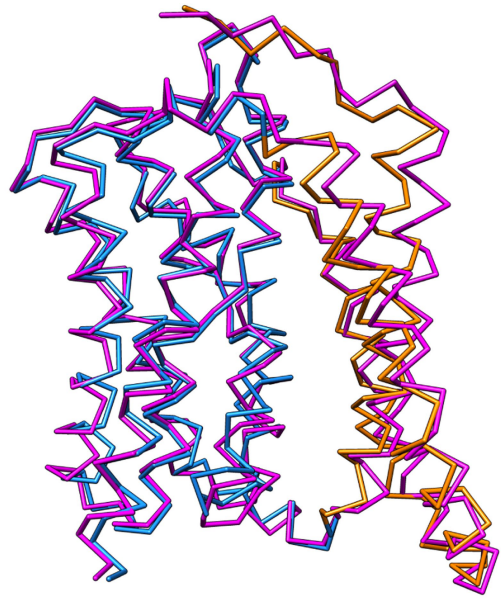
b, Superimposition of NTCP$_{EM}$ structures in complex with Nb91 in nanodisc (dark pink) and that in complex with Mb91 in detergent (core domain, blue; panel domain, orange) (RMSD ~1.4 Å over all atoms).

Article

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

	#1 NTCP$_{EM}$-Mb91 (EMDB-13596) (PDB 7PQQ)	#2 NTCP$_{EM}$-Nb87 (EMDB-13593) (PDB 7PQG)
Sample preparation, Data collection and processing		
Medium for sample preparation	DDM	Nanodisc
Microscope	Titan Krios 2	Titan Krios 2
TEM mode	EFTEM Nanoprobe	EFTEM Nanoprobe
Imaging Mode	Counting	Counting
Camera	Gatan Quantum-K3	K3
Magnification	165,000x	165,000x
Voltage (kV)	300	300
Electron exposure (e–/Å^2)	56.5	57.8
Defocus range (µm)	-0.6 to -1.75	-0.5 to -1.5
Pixel size (Å)	0.504	0.504
Symmetry imposed	C1	C1
Initial particle images (no.)	5,796,802	6,535,687
Final particle images (no.)	184,768	61,053
Map resolution (Å)	3.3	3.7
FSC threshold	0.143	0.143
Map resolution range (Å)	22.3-2.8	19.9-3.4
Refinement		
Initial model used (PDB code)	none	7PQQ
Model resolution (Å)	3.2	3.7
FSC threshold	0.143	0.143
Model resolution range (Å)	22.3-2.8	19.9-3.4
Map sharpening B factor (Å^2)	-60.3	-18.1
Model composition		
Non-hydrogen atoms	3,207	3,251
Ligands	none	none
B factors (Å^2)		
Protein	43.57	156.22
Ligand	-	-
R.m.s. deviations		
Bond lengths (Å)	0.004 (0)	0.004 (0)
Bond angles (°)	0.678 (0)	0.877 (0)
Validation		
MolProbity score	1.92	2.39
Clashscore	12.2	27.61
Poor rotamers (%)	0	0.28
Ramachandran plot		
Favored (%)	95.38	92.6
Allowed (%)	4.62	7.40
Disallowed (%)	0	0

nature portfolio

Corresponding author(s): Nicolas Reyes

Last updated by author(s): Jan 2, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	SerialEM 3.8.16
Data analysis	cryoSPARC v2 and V3, USF Chimera 1.14, ChimeraX 1.2, COOT 0.8.9.2, Prism 8.0.1, Jalview 2.10.4b1, Phenix 1.18.2, CASTp 3.0, PISA 1.52, Consurf 2016, I-TASSER 5.1, MOLEonline 2.5, Muscle 3.8.31, Psipred 4.02

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Structural models of NTCPEM-Np87 and NTCPEM-Nb91 complexes have been deposited in Protein Data Bank (PDB) with accession codes 7PQG and 7PQQ, respectively, and the corresponding cryo-EM maps were deposited in the Electron Microscopy Data Bank (EMDB) under accession numbers EMD-13593, and EMD-13596.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. Functional experiments were done in at least n=3 biologically independent experiments. Sizes of the cryo-EM data sets were predetermined based on microscope availability, and data collected was sufficient to obtain maps with the reported resolutions that enable structure determination.
Data exclusions	No functional data was excluded from the analyses. Poor-quality cryo-EM micrographs and particles were removed using standard analysis software and protocols
Replication	Functional studies were repeated during biologically experiments as stated in Figure legends and Methods, and results were replicated reasonably well within the experimental error reported. For structural studies, the quality of purified protein and EM specimen was fully reproducible using at least three independent biological replicates
Randomization	Randomization is not relevant to this study
Blinding	Blinding was not performed, because grouping was not applicable to this study

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	llama (Lama glama) antibody fragments against NTCPCO (nanobodies or VHH) were created and used in this study
Validation	Nanobody 87 and 91 binding to NTCPEM was assayed by size-exclusion chromatography in detergent solutions, as well as by fluorescence-based methods in HEK293 cells expressing the transporters as shown and described in Fig. 1b, and Fig. 4c of this study. Binding of Nb87 to human NTCPWT was also confirmed in HEK293 cells as shown and described in Fig. 4c

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293F (Thermo Fisher)
Authentication	No cell-line authentication was performed
Mycoplasma contamination	No mycoplasma contamination tests were performed
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used