SUPPORTING INFORMATION

Structural fold and binding sites of the human Na^+-phosphate cotransporter NaPi-II

Cristina Fenollar-Ferrer[1,4], Monica Patti[2], Thomas Knöpfel[2],
Andreas Werner[3*], Ian C. Forster[2*] and Lucy R. Forrest[1,4*]

Biological Sciences, Biophysics

[1]Computational Structural Biology Group,
Max Planck Institute of Biophysics,
Max-von-Laue-Straße 3,
D-60438 Frankfurt am Main,
Germany

[2]Institute of Physiology and Zurich Center for Integrative Human Physiology,
University of Zurich,
Winterthurerstrasse 190,
CH-8057 Zürich
Switzerland

[3]Epithelial Research Group
Institute for Cell and Molecular Biosciences
Newcastle University
Framlington Place
Newcastle upon Tyne, NE2 4HH
United Kingdom

[4]Present address:
Computational Structural Biology Section,
Porter Neuroscience Research Center,
National Institutes of Neurological Disorders and Stroke,
National Institutes of Health,
Bethesda, MD 20892, USA

Detailed Methods

Sequence analysis of NaPi-IIa

To identify conserved structural regions in NaPi-II transporters, sequence homologs of NaPi-IIa were collected with a 3-iteration HMMER search using the sequence of mouse NaPi-IIa (accession NP_035522) on the HHMER server (1). Note that the mouse and human NaPi-IIa sequences share 91% identical residues overall, and 97% in core regions; for ease of comparison we use the human residue numbering throughout. The NaPi-IIa sequence was scanned against the NCBI nr database dated 20.02.13 and the homologs obtained were clustered using UCLUST (2) with a similarity cutoff of 50%. Representative sequences <400 residues in length were excluded so that every sequence in the final alignment covered >70% of the query sequence. The resultant 300 sequences were multiply aligned using MSAProbs (3). Blocks of conserved residues in the resultant multiple-sequence alignment (see subset in Fig. S1) were identified as belonging to repeated regions in NaPi-IIa (see Results), and given preliminary assignments of region 1 (residues 86-256), and region 2 (residues 335-489). These residue definitions were made as inclusive as possible, so as not to truncate any helices. However, it was not clear whether an additional conserved segment at the C-terminal end of region 2 (region C, residues 504-564) is part of the structural repeat.

To refine the boundaries of the structural repeats, we analyzed hydropathy profiles of the three regions. From the MSAprobs multiple-sequence alignment, three segments were extracted, namely region 1, region 2, and a longer version of the latter including region C, called region 2+C. Hydropathy profiles for these segments, averaged over all sequences in each multiple-sequence alignment, were constructed and aligned using the AlignMe (4) web server (http://www.bioinfo.mpg.de/AlignMe), using the Hessa et al. (5) hydrophobicity scale and a 13-residue long triangular window for smoothing.

To identify more precisely the equivalent residues in the identified structural repeats of NaPi-II at the sequence level, we used hidden Markov Models (HMMs) as descriptors of a given sequence region. First, an HHalign HMM based on NaPi-II sequences was obtained by using the HHblits web server (6) to scan the mouse NaPi-IIa sequence against the NCBI nr20 database dated 22.02.13. From these results, the best 50 sequence hits were selected, short sequences that covered only one of the two repeats were excluded, and from the remaining sequences the final HMM for NaPi-II was constructed. From this full-length profile, shorter HMMs were extracted that corresponded to the different regions mentioned above. Subsequently, the region 1 HMM was aligned with either the region 2 HMM, or the region2+C HMM, using the HHalign server (7) (Fig. S2).

Identification of VcINDY as a suitable template

Putative templates were searched for using PSI-BLAST (8), Phyre2 (9) and COMA (10). We also used the full-length HHalign HMM of NaPi-IIa to scan the pdb70 database (from 4.4.13) using the HHpred server (7). Hits from the HHpred search were compared with NaPi-II by full-length pair-wise alignments using the Profile+Secondary-Structure (AlignMePS) mode on the AlignMe server (http://www.bioinfo.mpg.de/AlignMe).

Sequence homologs of a putative template, VcINDY (see Results) were identified as above for NaPi-IIa. Sequences too short to cover an entire structural repeat of VcINDY (<300 residues) were excluded. The resultant HHalign HMM of full-length VcINDY contained 729 sequences. From this, fragment HMMs were

extracted for RU1 and RU2 of VcINDY, i.e., residues 40-242 covering TMs 2-6, and residues 253-462 covering TMs 7-11, respectively.

Structural Modelling of NaPi-IIa

For a given template, the sequence alignment to the target protein (here, human NaPi-IIa) is the most critical determinant of model accuracy. However, attempts to align full-length NaPi-IIa with full-length VcINDY using several approaches, such as HHalign, produced unsuitable alignments replete with gaps, particularly in the C-terminal half. Thus, we aligned the repeat units independently. First, to obtain an overview of the equivalent TM helices, AlignMe was used to align family-averaged hydropathy profiles of a given repeat unit of the template and target proteins. A more precise alignment of each repeat unit for VcINDY and NaPi-IIa was then obtained using the AlignMe by providing the server with a position-specific substitution matrix (generated for the HHalign homologues) for each repeat unit using the Profile+Secondary-Structure+Transmembrane (AlignMePST) mode. These alignments of RU1 and RU2 were combined into a single alignment of the core of the two proteins.

The alignment of the core was used to build a preliminary model of human NaPi-IIa with Modeller 9v5 (11), using the X-ray structure of VcINDY (PDB identifier 4F35) as a template, in particular TMs 4-6 (residues 83-242) and TMs 9-11 (residues 310-462) for RU1 and RU2, respectively (see Fig S3). Following an iterative procedure, the alignment was refined in order to: a) remove gaps within secondary structure elements; b) orient less conserved residues towards the protein surface; c) ensure that more conserved residues participate in the packing of the protein or in the binding site and; d) be as consistent as possible with a local (fragment) alignment obtained by HHalign for each repeat (Figs. S4C, D). Optimization of the position of conserved and variable residues was aided by conservation scores for human NaPi-IIa obtained using the ConSurf server (12) In the final alignment the percentage of identical residues between the template and model was ~15% in RU1 and 11% in RU2.

The final alignment (Fig. 2B) was used to model the NaPi-IIa core domain, which included residues 97-249 of RU1 and 341-505 of RU2, i.e., those residues for which a template was present. In a final modeling stage, two Na^+ ions and a phosphate substrate were included. Harmonic upper bound distance restraints of 3.3 Å were imposed between: the Na^+ ion at Na2 and side chain O atoms of residues S164, T195, S196, N199 and substrate P_i; the Na^+ ion at Na3 and side chain O atoms of residues Q417, S419; and P_i and side chain O atom of residue S164, or side chain N atom of N199. The final model was that with the lowest Modeller probability distribution function score out of a set of 2000 models. The ProQM score (13) of this model was 0.545 (improved from 0.465 for the preliminary model) compared to a ProQM score of 0.675 for the template; and only four residues were found in disallowed regions of the Ramachandran plot according to PROCHECK (14), all of which are located in loops. The model is available upon request from the authors.

Immunostaining

Oocyte sections were blocked in 1% BSA/PBS for 15 min at room temperature. NaPi-IIa antibody raised against a COOH-terminal peptide (15) diluted 1/400 in 0.02% Na-azide/PBS and incubated on the slides overnight at 4°C. Sections were washed twice with hypertonic PBS (PBS with additional 18 g NaCl/L) and once in PBS for 5 min. Sections were incubated with secondary Donkey anti-rabbit Alexa-488 antibody (Invitrogen) diluted 1/1000 in 0.02% Na-azide/PBS for 1 hour. Fluorescence was detected using a fluorescence microscope (Leica CTR600).

References

1. Finn RD, Clements J, & Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39(Web Server):W29-W37.
2. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19):2460-2461.
3. Liu Y, Schmidt B, & Maskell DL (2010) MSAProbs: multiple sequence alignment based on pair hidden Markov models and partition function posterior probabilities. *Bioinformatics* 26(16):1958-1964.
4. Stamm M, Staritzbichler R, Khafizov K, & Forrest LR (2013) Alignment of helical membrane protein sequences using AlignMe. *PLoS One* 8(3):e57731.
5. Hessa T, *et al.* (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* 433(7024):377-381.
6. Remmert M, Biegert A, Hauser A, & Soding J (2011) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods*.
7. Söding J (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21(7):951-960.
8. Altschul SF, *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25(17):3389-3402.
9. Kelley LA & Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 4(3):363-371.
10. Margelevicius M, Laganeckas M, & Venclovas C (2010) COMA server for protein distant homology search. *Bioinformatics* 26(15):1905-1906.
11. Šali A & Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234:779-815.
12. Ashkenazy H, Erez E, Martz E, Pupko T, & Ben-Tal N (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 38(Web Server issue):W529-533.
13. Ray A, Lindahl E, & Wallner B (2010) Model quality assessment for membrane proteins. *Bioinformatics* 26(24):3067-3074.
14. Laskowski RA, Macarthur MW, Moss DS, & Thornton JM (1993) PROCHECK - A program to check the stereochemical quality of protein structures. *J. Appl. Crystall.* 26:283-291.
15. Custer M, Lotscher M, Biber J, Murer H, & Kaissling B (1994) Expression of Na-P(i) cotransport in rat kidney: localization by RT-PCR and immunohistochemistry. *Am. J. Physiol.* 266(5 Pt 2):F767-774.

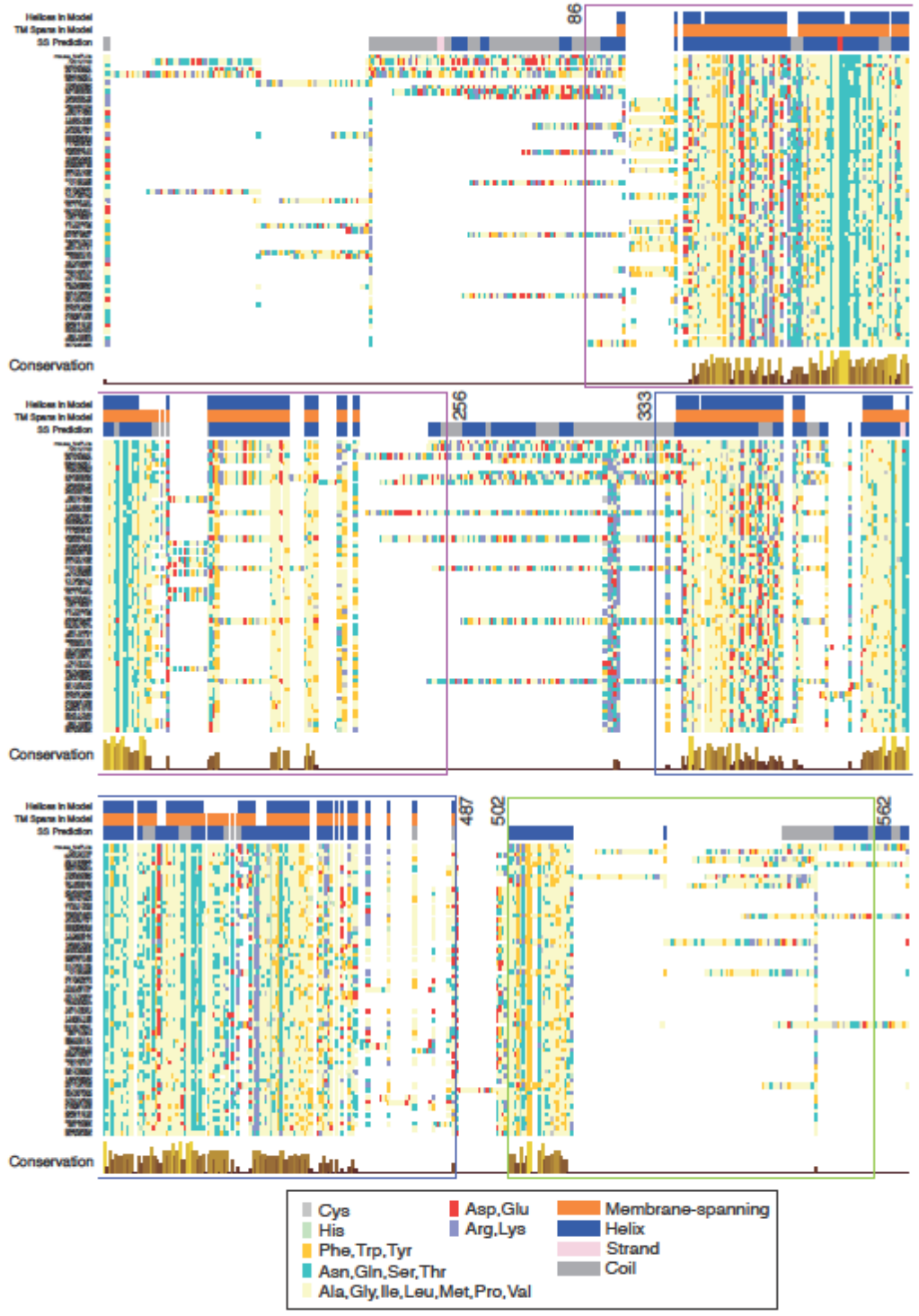
Helices in Model
TM Spans in Model
SS Prediction
Conservation
86
256
333
487
502
562
Cys
His
Phe,Trp,Tyr
Asn,Gln,Ser,Thr
Ala,Gly,Ile,Leu,Met,Pro,Val
Asp,Glu
Arg,Lys
Membrane-spanning
Helix
Strand
Coil

Fig. S1 - Multiple sequence alignment of NaPi-II homologs indicating the conserved region 1 (*purple box*, residues 86-256), region 2 (human residues 335-489, *blue box*) and region C (*green box*, human residues 504-564). The top sequence line corresponds to mouse NaPi-IIa. Only homologues that cover 90% of the sequence of mouse NaPi-II are shown. For clarity, the sequences of the cytosolic C-terminal domain are not shown, and residues are represented as squares colored by residue type and not as letters (see legend). The fraction conservation of each column is also indicated. The predicted secondary structure ("SS prediction") is shown above the alignment, along with the TM spanning residues (orange, "TM spans in model") and SS elements ("Helices in model") in the final model. Helical (*dark blue*), coiled (*gray*) and stranded (*pink*) SS elements are indicated.

Fig. S2 - HHalign sequence alignment of human NaPi-IIa repeats to each other. Region 1 containing residues 86-256 was aligned to region 2+C consisting of residues 335-564, resulting in a local (i.e., not full-length) alignment that indicates the locations of the conserved repeated elements. The predicted secondary structure for each residue is indicated above and below the respective sequences, with helices in blue, coil in gray and strand in pink.

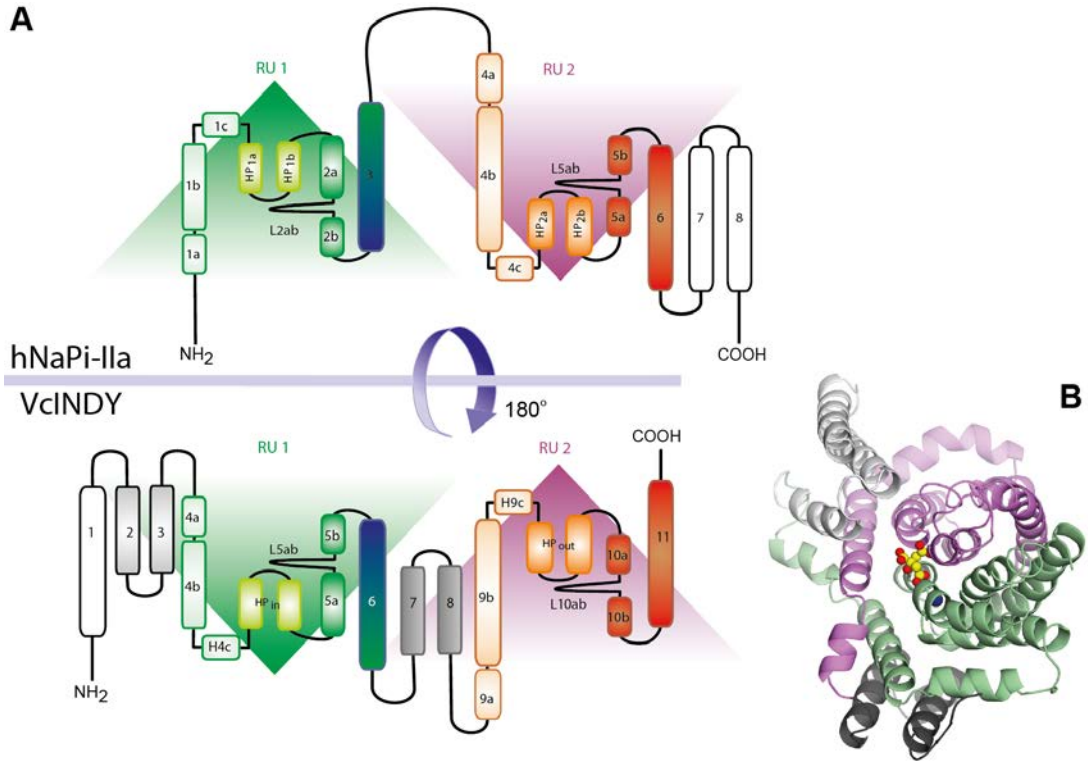

Fig. S3 – Overview of the topology and fold of the template VcINDY X-ray structure (PDB identifier 4F35) compared to the predicted topology of NaPi-II.
(*A*) Schematics of the transmembrane topologies of hNaPi-IIa (*top*) and VcINDY (*bottom*), indicating: the locations of repeat unit 1 (*green triangle*) and repeat unit 2 (*purple triangle*); the two peripheral transmembrane helices in hNaPi-IIa that have no template in VcINDY, namely TMs 7-8 (*white*); and the five peripheral transmembrane helices in VcINDY that have no equivalent in NaPi-II, namely TMs 1-3 (*white/light gray*) and TMs 7-8 (*gray*).
(*B*) Fold of VcINDY viewed from the periplasm shown as cartoon helices, and colored to highlight the repeats (*green and purple*), and the peripheral helices (*white and gray*). Citrate (*yellow and red*) and one sodium (*blue*) observable in the structure are shown as spheres.

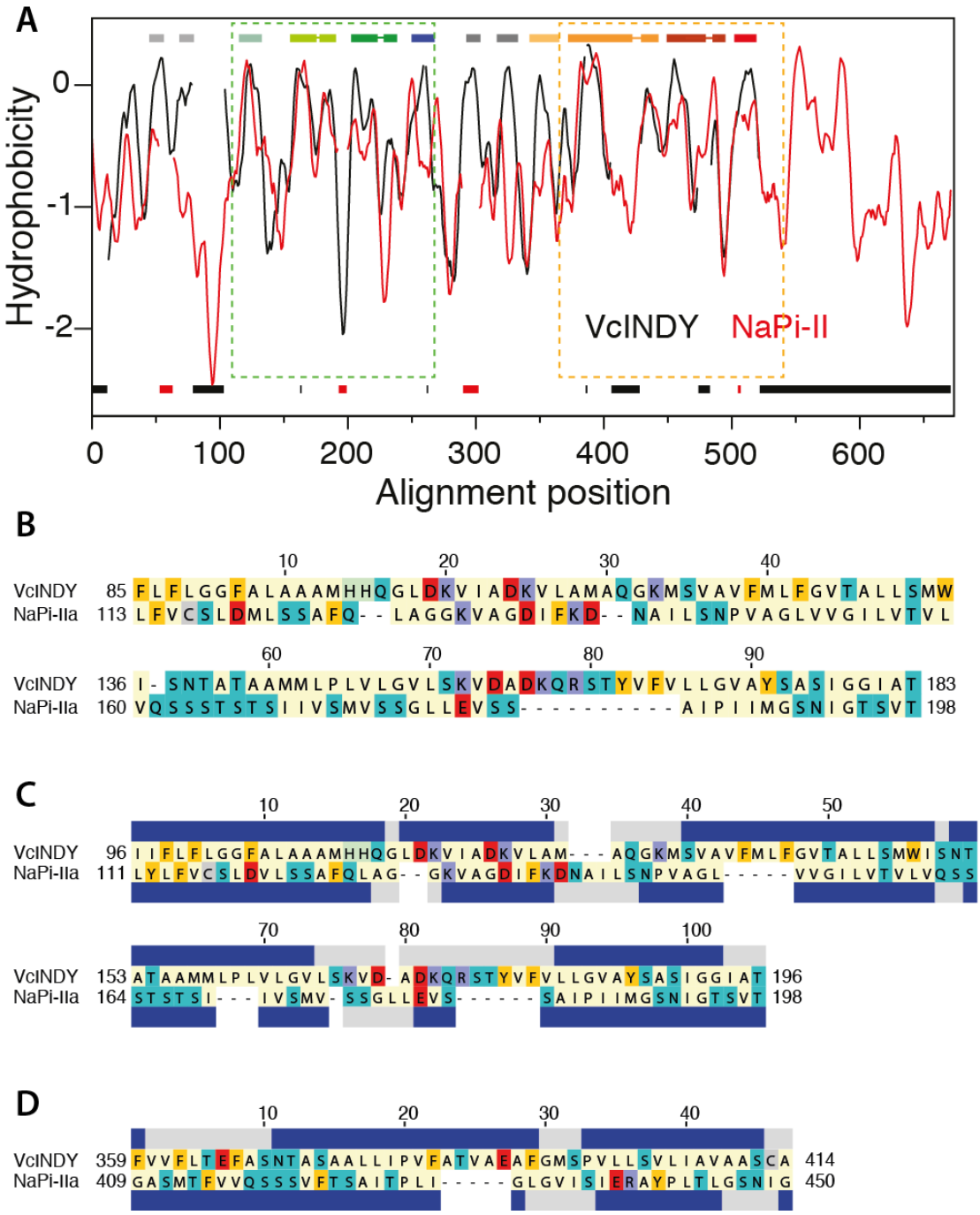

Fig. S4 – Comparison of human NaPi-IIa with VcINDY, a putative structural homologue.
(*A*) Alignment of family-averaged hydropathy profiles of full-length NaPi-IIa (*red*) and VcINDY (*black*). The extents of VcINDY TM helices are shown above the alignment, colored according to Figure 3. NaPi-IIa repeats are indicated with dashed boxes. Insertions in the alignment are indicated as thick bars below the alignment.
(*B*) Local alignment output from HHpred template search, showing that the first repeat units of each protein can be aligned.
(*C, D*) HHalign sequence alignments of VcINDY and human NaPi-IIa RU1 (*C*) and RU2 (*D*). Note that even though in each case only a fragment (local) alignment is output by HHalign, the QSSS motif of the NaPi-II sequence is matched to the region of a conserved xSNT motif in VcINDY in both repeats.

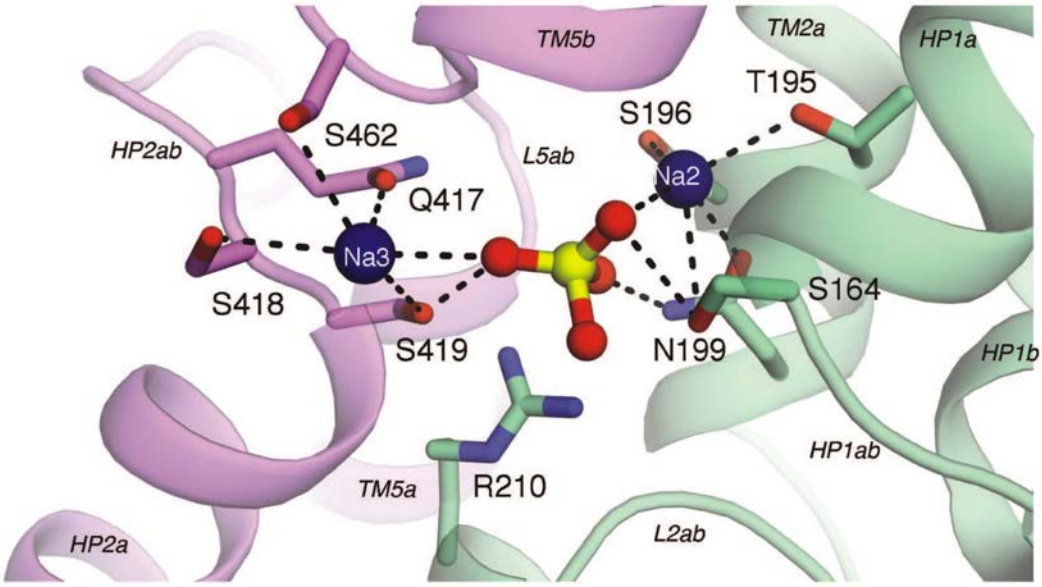

Fig S5 Close-up of the predicted substrate binding sites in NaPi-IIa, using the same orientation as in Fig 4*A*. Residues close to the binding site are highlighted (*sticks*), as are bound P_i (*ball and stick*), and sodium ions (*spheres*). Residues N199 and S462 have previously been implicated in substrate binding. Residues T195, S196, N199, R210, Q417, S418, S419 were investigated in the present study.

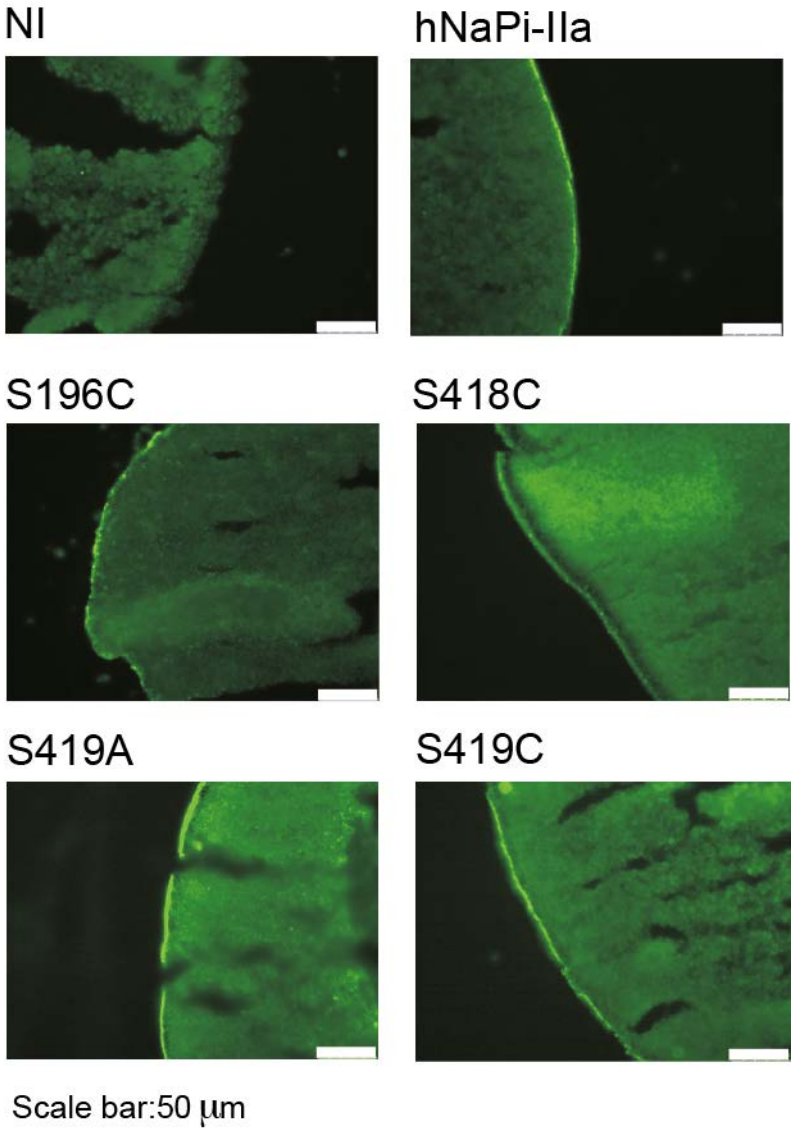

Fig. S6 – Immunostaining of *Xenopus* oocytes confirms the surface localization of wild-type (WT) NaPi-IIa and selected mutants (S196C, S418C and S419C) that showed no functional behavior. NI: non-injected control oocyte. (see SI Detailed Methods for details).

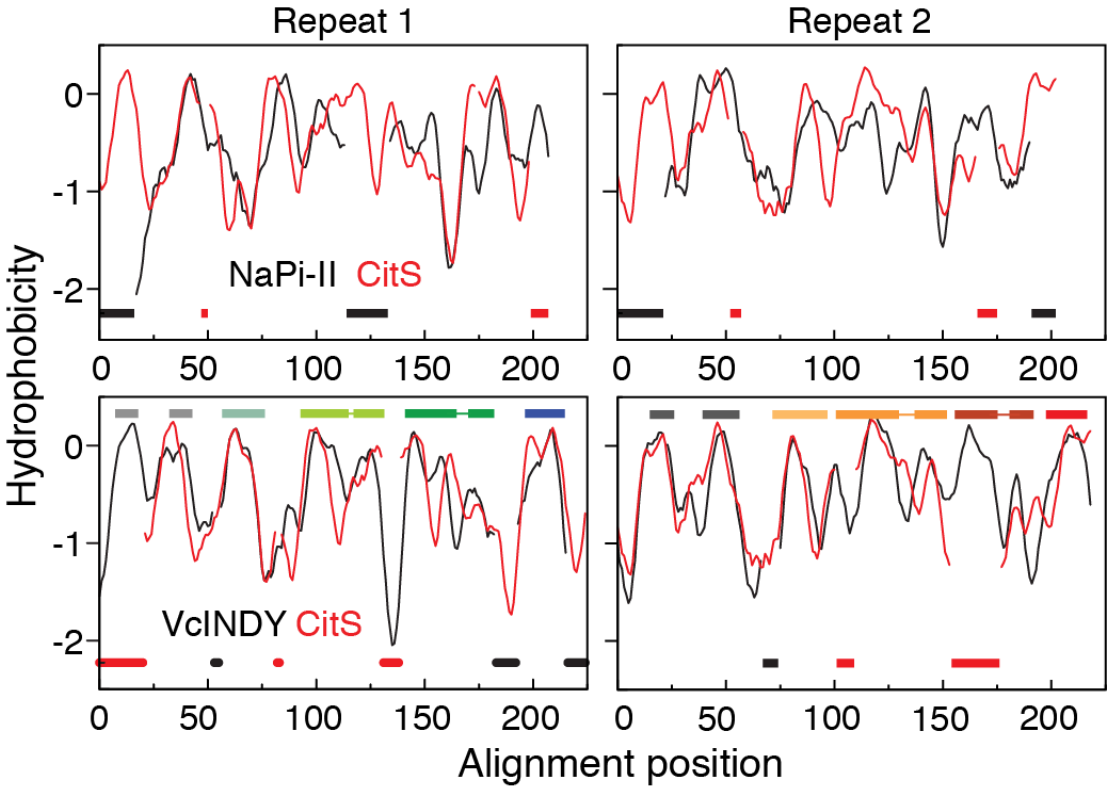

Fig. S7 – Hydropathy profile alignments of an ST[3] class transporter CitS (*red*) with NaPi-IIa (*black*, *top*) or with VcINDY (*black*, *bottom*). Hydropathy profiles of repeats were aligned separately. Approximate extents of the TM segments in VcINDY are shown above the lower plot and colored according to Fig. 3. Insertions are indicated using thick bars below each alignment.

1.