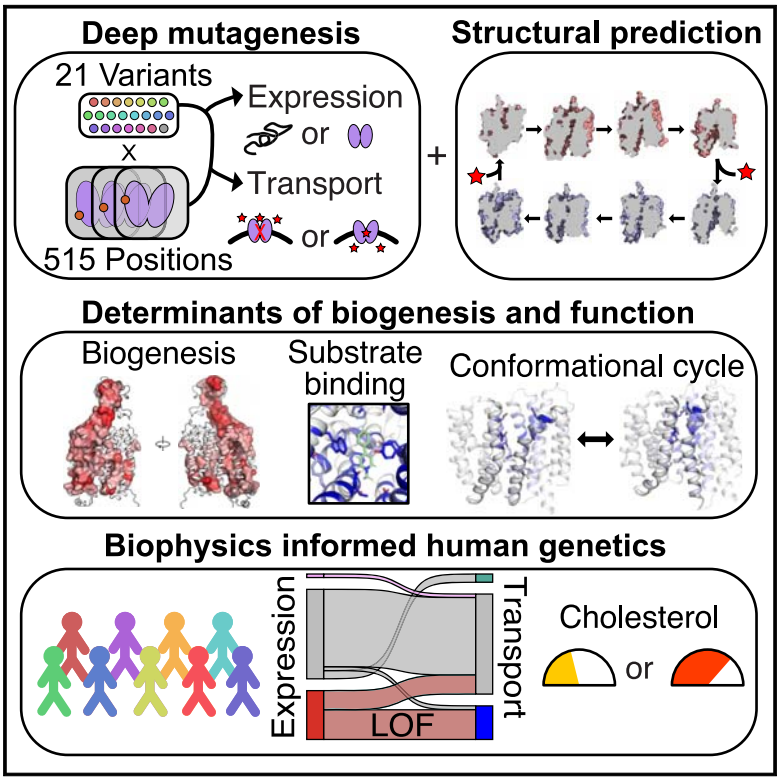
Molecular Cell

Article

The full spectrum of SLC22 OCT1 mutations illuminates the bridge between drug transporter biophysics and pharmacogenomics

Graphical abstract



Highlights

- Deep mutational scan of OCT1 reveals the determinants of biogenesis and substrate uptake
- Discovery of a conserved motif, the stability helix, essential to SLC22 biogenesis
- AI structure prediction and mutational data for accurate structure-function modeling
- Integrating genomic-health records reveals variants effect on human physiology

Authors

Sook Wah Yee,
Christian B. Macdonald,
Darko Mitrovic, ..., Lucie Delemotte,
Kathleen M. Giacomini,
Willow Coyote-Maestas

Correspondence

lucie.delemotte@scilifelab.se (L.D.),
kathy.giacomini@ucsf.edu (K.M.G.),
willow.coyote-maestas@ucsf.edu (W.C.-M.)

In brief

Yee et al. create a complete functional map of how genetic variants in the OCT1 transporter alter biogenesis and substrate uptake. This work provides a framework for biophysically informed precision medicine by integrating functional genomics, mechanistic modeling, and human genetics to predict variant effects on disease and drug efficacy.

Yee et al., 2024, Molecular Cell *84*, 1932–1947
May 16, 2024
https://doi.org/10.1016/j.molcel.2024.04.008

CellPress

CellPress

Molecular Cell

Article

The full spectrum of SLC22 OCT1 mutations illuminates the bridge between drug transporter biophysics and pharmacogenomics

Sook Wah Yee,[1,8] Christian B. Macdonald,[1,8] Darko Mitrovic,[4,8] Xujia Zhou,[1] Megan L. Koleske,[1] Jia Yang,[1] Dina Buitrago Silva,[1] Patrick Rockefeller Grimes,[1] Donovan D. Trinidad,[2] Swati S. More,[1,9] Linda Kachuri,[5,6] John S. Witte,[5,6] Lucie Delemotte,[4,*] Kathleen M. Giacomini,[1,*] and Willow Coyote-Maestas[1,3,7,10,*]

[1]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA 94143, USA
[2]Department of Medicine, Division of Infectious Disease, University of California, San Francisco, San Francisco, CA 94143, USA
[3]Quantitative Biosciences Institute, University of California, San Francisco, San Francisco, CA 94143, USA
[4]Science for Life Laboratory, Department of Applied Physics, KTH Royal Institute of Technology, 12121 Solna, Stockholm, Stockholm County 114 28, Sweden
[5]Department of Epidemiology and Population Health, Stanford University, Stanford, CA 94305, USA
[6]Stanford Cancer Institute, Stanford University, Stanford, CA 94305, USA
[7]Chan Zuckerberg Biohub, San Francisco, CA 94148, USA
[8]These authors contributed equally
[9]Present address: Center for Drug Design (CDD), College of Pharmacy, University of Minnesota, MN 55455, USA
[10]Lead contact
*Correspondence: lucie.delemotte@scilifelab.se (L.D.), kathy.giacomini@ucsf.edu (K.M.G.), willow.coyote-maestas@ucsf.edu (W.C.-M.)
https://doi.org/10.1016/j.molcel.2024.04.008

SUMMARY

Mutations in transporters can impact an individual's response to drugs and cause many diseases. Few variants in transporters have been evaluated for their functional impact. Here, we combine saturation mutagenesis and multi-phenotypic screening to dissect the impact of 11,213 missense single-amino-acid deletions, and synonymous variants across the 554 residues of OCT1, a key liver xenobiotic transporter. By quantifying in parallel expression and substrate uptake, we find that most variants exert their primary effect on protein abundance, a phenotype not commonly measured alongside function. Using our mutagenesis results combined with structure prediction and molecular dynamic simulations, we develop accurate structure-function models of the entire transport cycle, providing biophysical characterization of all known and possible human OCT1 polymorphisms. This work provides a complete functional map of OCT1 variants along with a framework for integrating functional genomics, biophysical modeling, and human genetics to predict variant effects on disease and drug efficacy.

INTRODUCTION

Membrane proteins connect cells to their external environments and play major roles in human health and disease. Transporter defects cause diseases, and transporter-mediated uptake is the primary way therapeutics reach their targets.[1–3] The human genome encodes over 450 solute carrier (SLC) transporters, with over 120 directly linked to diseases and over 40 playing important roles in drug disposition.[4–7] The organic cation transporter subgroup of the SLC22 family (OCTs) are polyspecific transporters expressed in the liver, kidney, and intestine. With their tissue-specific expression and broad substrate selectivity, OCTs play major roles in the uptake of endogenous molecules and xenobiotics.[8–10] OCT1 transports a range of cations (e.g., 1-methyl-4-phenylpyridinium [MPP^+] and tetraethylammonium [TEA]), endogenous molecules (e.g., thiamine[11] and serotonin[12]), and drugs (e.g., metformin[13] and morphine[14]).

Impaired OCT1 function causes various clinical phenotypes, including increases in plasma drug concentration, shifts in therapeutic windows,[15–17] and alterations in endogenous metabolite and lipid levels.[18–21] While ample evidence of OCT1 polymorphisms modulating drug responses exists, only 39 out of over 800 common and rare human genetic variants have been experimentally characterized.[22] There are well over 10,000 possible single-codon OCT1 variants that could contribute to variation in therapeutic and adverse drug response. Similar issues plague our understanding of other transporters with genetic variation and poorly understood functional consequences. To understand the diverse human responses to therapeutics, we need comprehensive models of how genetic variation impacts drug transporters. Crucially, such models must be *mechanistic* to distinguish between variants with broad effects across substrates, such as disrupting folding, and those with substrate-specific effects. These integrative models must incorporate the biophysics of folding, the structural biology of uptake, and the genomics of human physiology. Conventional variant interpretation methods are laborious, challenging, and provide no structural insight, making progress slow.

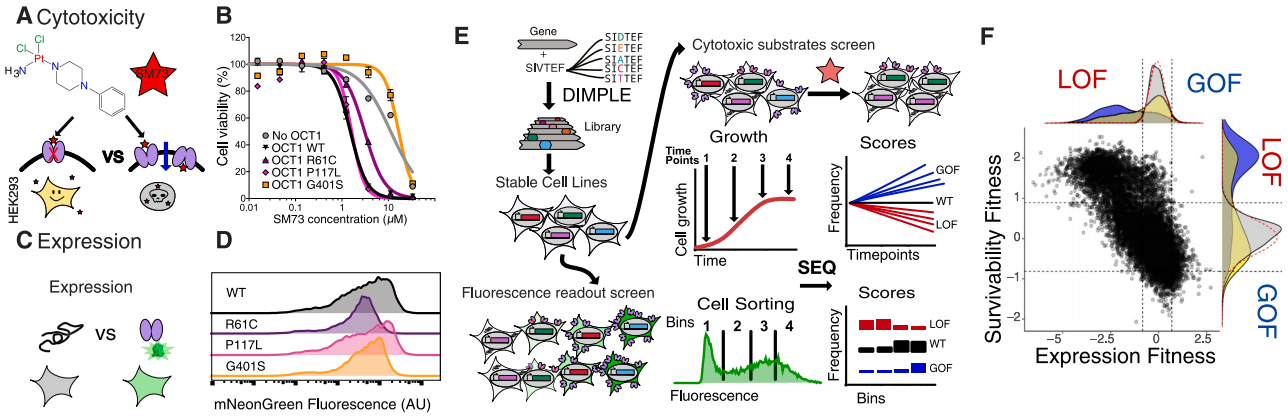


Figure 1. Workflow for multiparametric deep mutational scan of OCT1

(A) Negative selection screen schematic for OCT1 function using the cytotoxic OCT1 substrate SM73.

(B) Cell viability determined for variants with SM73.

(C) Schematic of VAMP-seq assay using split mNeonGreen fluorescent protein: protein of interest is tagged with 15-aa mNG-11 fragment and complementing mNG1-10 is transiently transfected cytosolically.[34]

(D) Flow cytometric detection of variant effects on mNeonGreen fluorescence (x axis).

(E) We generated and integrated an OCT1 deep mutational scanning library into stable landing pad HEK293T cell lines and determined the impact of variants on expression and survivability.[36,37]

(F) Scatterplot of survivability (y axis) and expression (x axis) effects for 11,213 OCT1 variants. Above and to the right are density plots for synonymous (gray), missense (yellow), and single-codon deletion (purple) classes of mutations. Cutoffs for loss and gain of function for are indicated by a dotted line based on 2 standard deviations from a normal distribution fit to synonymous variants (dashed red lines).

Deep mutational scanning (DMS) is coming to prominence as a method for comprehensively assaying genetic variation.[23] However, the majority of scans measure a single phenotype, such as a growth-derived "fitness" value. Recent machine-learning-based variant effect predictors also provide a single "pathogenicity" score for missense mutations across the proteome.[24–26] These single values cannot capture phenomena such as heterozygosity,[27] dominant-negative mutations,[28] or pleiotropy.[29] A mechanistic understanding of variant effects would allow us not only to predict whether a mutation alters drug response or causes disease but also to go further to develop actionable individual guidelines for dosage and therapeutic development.

Here, we take the first step toward this goal by integrating mutational scanning, structure prediction, and population genetics databases to learn how all >11,213 variants alter OCT1 biogenesis, function, and human physiology. This work generates an immediately useful comprehensive mutational dataset for an important drug transporter while providing insights into how mutations in this transporter and other closely related transporters alter drug efficacy. This integrative approach could bridge molecular biophysics to clinical genetics, enabling a new generation of precision medicine.

RESULTS

OCT1 variant expression primarily determines functional effects

To understand how OCT1 polymorphisms alter function, we employed DMS of 11,213 variants across the 554 residues of OCT1. In DMS, the individual effects of an entire library of variants are measured in pooled genetic screens.[23] Here, we measured multiple phenotypes to reduce variant effects to the more fundamental metrics of expression and specific activity.[30–33] Our first screen uses a cisplatin analog transported by OCT1, SM73, with variant-specific cytotoxicity (Figures 1A and 1B). In this negative selection setup, variants with decreased effective transport will increase in frequency in a pooled screen compared with variants with wild-type (WT)-like or increased activity. In our second screen, we use a variation on VAMP-seq to measure how mutations affect OCT1 expression.[34] Here, we tag OCT1 with a fragment of a split-fluorescent protein (mNeonGreen) and co-express the larger remaining fragment of a fluorescent protein such that only cells with OCT1 expression will have complementation and yield fluorescence (Figure 1C).[34,35] Mutations affecting expression cause corresponding decreases in fluorescence, which are distinguishable by flow cytometry (Figure 1D).

We used our deep insertion, deletion, and missense mutation libraries for exploring protein variation in evolution, disease, and biology (DIMPLE) approach to generate libraries of OCT1 with all single-codon missense substitutions and deletions.[36] We found that the pre-screen library had 97% of possible variants evenly distributed across 99.5% of possible sites (Figure S1A). We integrated this library into stable landing pad HEK293T cell lines and applied the above cytotoxicity survival and VAMP-seq (variant abundance by massively parallel sequencing) expression screens (Figures 1E and S2).[37] In the cytotoxicity assay, we measured how variant frequency changes over time, relative to WT, by sequencing. For the expression screen, we sorted library-integrated cells into four

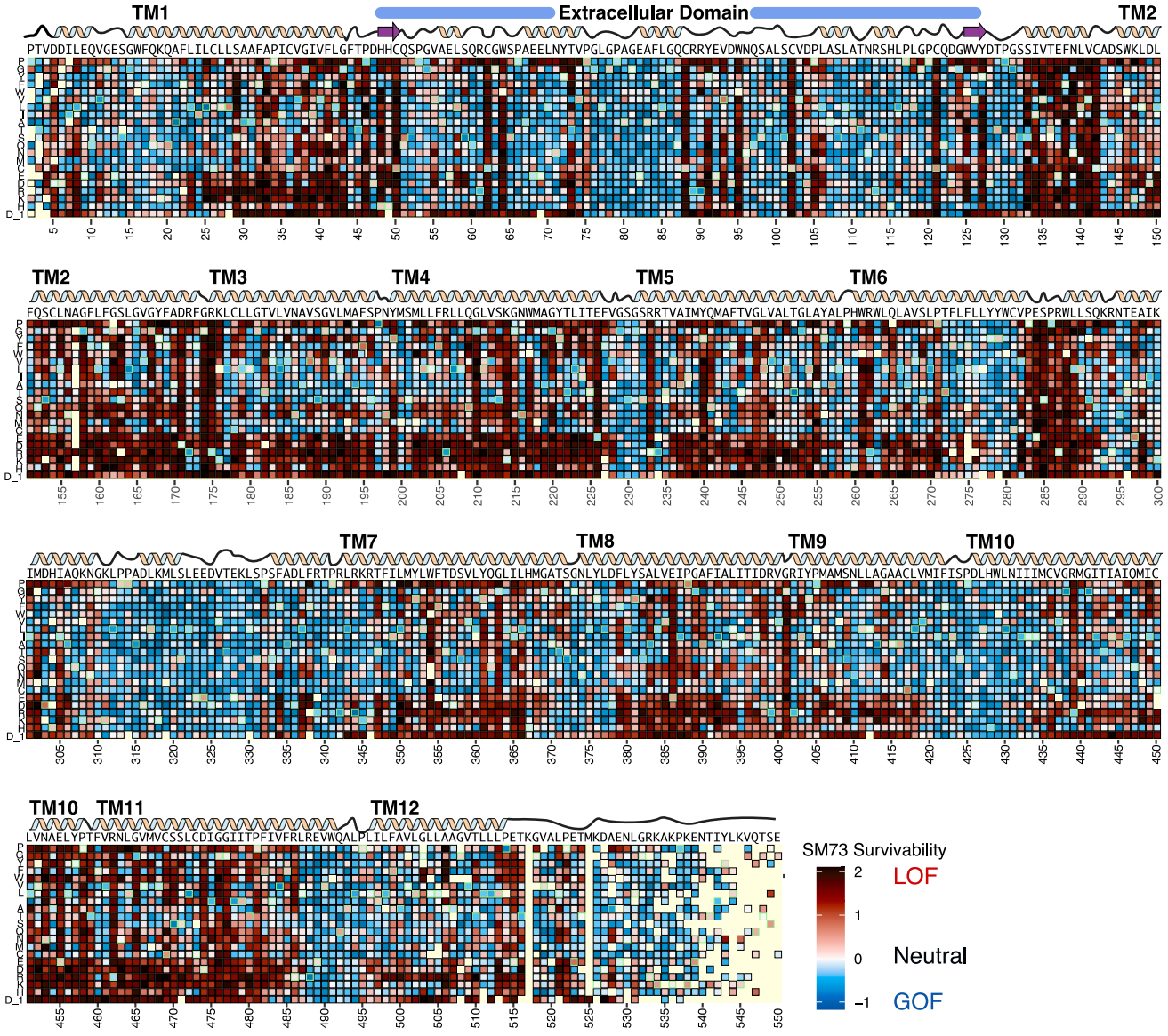


Figure 2. Heatmap of OCT1 survivability deep mutational scan

SM73 survivability screen fitness overview: residue position (x axis) vs. variant identity (y axis) grouped by physical chemistry. Synonymous variants are boxed in green and missing data are in light yellow. Above, wild-type sequence, domain organization, and cartoon secondary structure representation of OCT1.

fluorescence bins and determined the distribution of variants across them by deep sequencing. This gave two scores for each position: a survivability score and an expression score (Figures 1F, 2, and S3A). We found the expression and survivability screens reproducible across triplicates (Pearson correlation coefficients 0.781–0.878, survivability in Figure S1B; 0.852–0.871, expression in Figure S1C) with low error (Figures S3B and S3C).

To understand how variation drives phenotypic change, we compared scores between assays. We found that survivability correlates strongly with expression (−0.76 Spearman coefficient, Figure S1D). The simplest explanation is most mutations alter function primarily via expression levels, which has been seen in other membrane proteins.[33,38–42] To rule out a spurious correlation caused by baseline library bias, we examined whether unselected variant frequency was a good predictor of expression or survivability but found weak correlation (0.029 and −0.085 Spearman coefficients, respectively; Figures S1E and S1F). To test which of the screens best corresponds to evolutionary pressure, we compared conservation scores from ConSurf with mean positional scores.[43] Conservation is slightly more correlated to mean positional survivability than expression (0.53 vs. −0.4 Spearman rho, Figures S1G and S1H). This increased correlation suggests survivability (i.e., effective activity) is closer to the relevant phenotype under selection than expression.

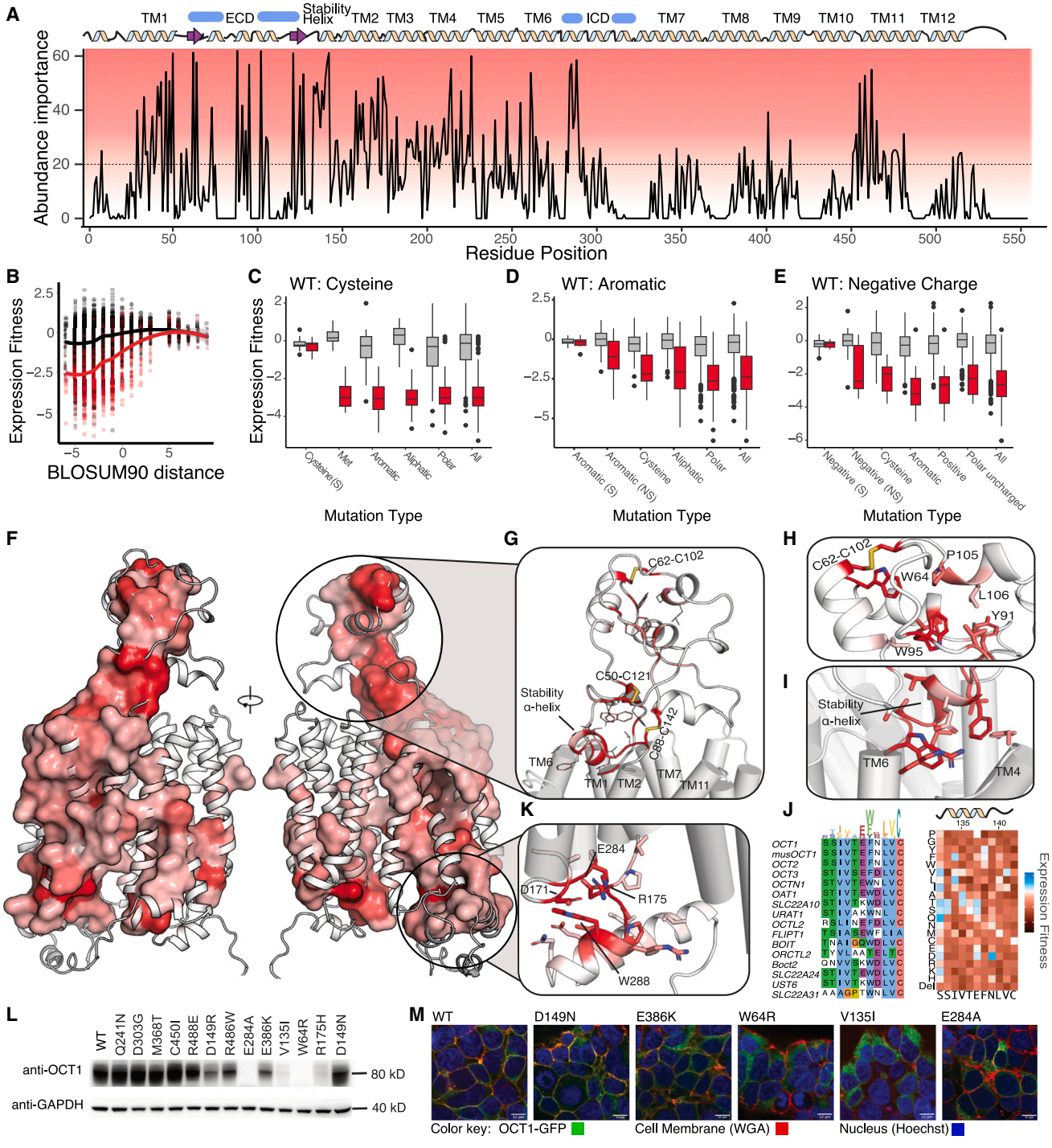


Figure 3. OCT1 abundance determinants revealed by mutational scan

(A) Abundance importance (y axis) vs. position (x axis) of OCT1. Dotted line indicates cutoff of 20.

(B) Comparison of chemical similarity of variant type (using BLOSUM90) with expression impact (y axis). Variants are grouped by positional abundance importance score, with high in red and low in black. LOESS (locally estimated scatterplot smoothing) smoothed curve plotted for each group.

(C–E) Physical-chemical effects at positions with high abundance importance score. Distributions of expression fitness scores of variants changing (C) cysteine, (D) aromatic, or (E) negatively charged residues to other chemistries, between variants at positions with high (red) vs. low (gray) abundance importance scores. S, synonymous; NS, non-synonymous. Box plots show mean (inner line), first and third quartiles (hinges), and 1.5 times interquartile range from hinges (whiskers), with points beyond as dots.

(legend continued on next page)

Transmembrane helices 1–6 and the extracellular domain drive abundance changes

Our observation that expression is the primary mediator of OCT1 variant function raises the question of how mutations alter expression. To identify positions with the largest impact on expression, we calculated a positional "abundance score," defined as the summed absolute value of expression scores at each position (see STAR Methods). Plotting this reveals that high abundance score is concentrated in the first 300 residues of OCT1, and thus the critical abundance determinants as well (Figures 3A and S3A).

Specific residue interactions required for membrane insertion, folding, and stability are often major biogenesis determinants. Such position should show increased sensitivity of abundance to side-chain chemistry. To test this, we used the BLOSUM90 matrix to calculate side-chain chemical similarity between WT and variants (Figure 3B). Here, mutations are assigned a score based on the similarity between the WT and mutant identities.[44] Mutations with similar physicochemistries to WT show little impact on expression across all residues. In contrast, as mutant physicochemistry becomes more dissimilar, positions of high abundance scores diverge from low abundance score, suggesting the abundance score can distinguish positions requiring chemically specific interactions.

Next, we investigated mutational tolerance at residues associated with stability and folding, including cysteines, aromatic, and negatively charged residues (Figures 3C–3E). We again find mutations that change physical chemistry at positions with high abundance score are extremely deleterious, with aromatic residues particularly sensitive. By comparison, hydrophobic or aliphatic residues are less sensitive to physiochemical changes, while cationic residues are sensitive to charge swaps but less so for aliphatic substitutions (Figures S4A–S4C). We conclude that amino acids enabling disulfide formation, electrostatic interactions, and aromatic packing appear to be particularly sensitive in abundance-important positions, as compared with other types of residues, which supports our hypothesis that the residues involved in biogenesis will display high abundance scores.

We were curious why residues involved in biogenesis are enriched in OCT1's first half. As with other major facilitator superfamily (MFS) transporters, OCT1 is made up of two 6-transmembrane (TM) domain bundles, termed the NTD and CTD bundles, but distinct to OCT1 and other SLC22s is an extracellular domain (ECD) between TM helices 1 and 2. We mapped our abundance scores onto the AlphaFold2 model, which matches experimental structures (Table S1). This revealed that abundance importance is primarily located in the N-terminal TMs, the ECD, and a small cluster of interacting residues at the bottom of the second 6-TM bundle (TM7 to TM12) (Figures 3A, 3F, and S4D).

The ECD is composed of a mixture of α helices and β sheets with large loops (Figure 3G). The distal pair of helices interact with a loop, forming a hydrophobic core rich in abundance-critical residues, including aromatics (W64, Y91), hydrophobics (V93, L106), and a disulfide bond (C62–C102) (Figure 3H). All mutations within this motif are sensitive to changes of physicochemistry. Previous studies identified glycosylation sites within ECD residues (N71, N96, S98, and N112)[45,46] that when mutated cause mild loss of expression.

An adjacent network of residues connects the distal structured region to the TM domain of OCT1, comprising a pair of double-stranded antiparallel β sheets among a pair of disulfide bonds (C50–C121 and C88–C142) abutting a pair of α helices alternatingly interacting with TM1-TM6 or TM7-TM12 (Figure 3I). These positions are sensitive to mutation, and the disulfide bonds appear essential (Figure 3J, right). We term this the α helix and following residues (133-SIVTEFNLVC) the "stability helix" because this region makes key interactions between TMs 1, 3, and 6 and the ECD, is enriched in stability-determining positions, and is extremely sensitive to physicochemistry changes. A glutamate (E137) in this helix forms an ionic interaction with R262 within TM 6, an aromatic residue juts into a hydrophobic core with other aromatics, and C142 forms a disulfide bond within the ECD. A sequence alignment of the stability helix (133-SIVTEFNLVC) across human SLC22s revealed high conservation (Figure 3J, left). This combination of conservation and sensitivity to mutation implies that the stability helix plays a key role in mediating biogenesis across the ECD and TM domains in the SLC22s.

Residues between interfaces in TMs 1–6 tend to be enriched in mutationally sensitive hydrophobic amino acids, suggesting a hydrophobic network connecting the ECD to the intracellular face of the N-terminal TM. On the intracellular bundle face, TMs 2–4, residues in the large intracellular TMs 6–7 loop, and numerous charged residues seem important in determining expression, including D171, R175, K176, E226, and E284 (Figure 3K). In agreement with these residues being key in biogenesis, they are extremely sensitive to mutations, with charge swaps being particularly deleterious. Surprisingly, these charged mutationally intolerant residues are not interacting in this model, perhaps suggesting alternative conformations.

To validate our screens, we chose a subset of 13 variants across a range of effects, including cases where the mechanism seemed to be mediated by abundance or uptake. We measured total protein abundance by western blotting (Figures 3L and S5A–S5D) and found that all low-expression mutants from our

(F) Abundance importance scores mapped white-to-red onto AlphaFold2 OCT1 model (AF-O15245-F1), with surfaces shown for positions with scores >20.
(G) Interactions between ECD, N-terminal TM, and stability α helix.
(H) A hydrophobic core in the extracellular domain.
(I) The stability α helix caps the N-terminal TM and interacts with TM4 and TM6.
(J) Left: sequence alignment of Human SLC22 stability helix colored by amino acid class. Above: sequence logo. Right: heatmap of expression fitness scores for stability helix colored red to blue.
(K) Cytosolic N-terminal TM showing cluster of charged residues with high abundance scores.
(L) Western blots using anti-OCT1 antibody on OCT1 variants identified in the screens with a range of phenotypes transfected in HEK293T cells. Full gels in Figures S5A–S5D.
(M) Confocal microscopy of OCT1 variants. Blue: Hoechst (nuclear); red: cell surface wheat germ agglutinin (WGA); green: OCT1-GFP.

screen (W64R, V135I, R175H, and E284A) had reduced apparent protein abundance, implying that the defects were mediated by protein biogenesis or upstream (e.g., including transcript abundance and stability). In contrast, mutants demonstrating substrate but not expression deficits (D149N, D149R, E386K, and R486W) showed WT -like abundance, pointing toward specific functional effects.

A limitation of VAMP-seq for membrane proteins is that it measures overall OCT1 expression but not specifically trafficking. Despite this, we expect variants with high expression in our screen to have high surface localization, barring changes to specific trafficking or targeting motifs. To test this hypothesis, we compared the localization of individual variants with confocal microscopy (Figures 3M and S5E–S5M). WT OCT1 was observed at the plasma membrane, and two loss-of-function variants (D149N and E386K) with WT-like abundance in the screen show similar surface localization, implying that these variants specifically alter transporter function. In contrast, three variants (W64R in the ECD hydrophobic core, V135I in the stabilizing α helix, and E284A in the intracellular stabilizing region) with loss of function in both screens had low surface localization. Although not exhaustive, this suggests that mutational impacts from our high-throughput screens likely represent their impacts including trafficking and that loss-of-function phenotypes in transporter function are primarily via biogenesis disruption.

OCT1-ligand-dependent conformational ensembles through coevolution, MD, and machine learning

While most loss-of-function variants act by altering expression, a portion (3.6%) have WT-like expression yet substantial effects on uptake (Figure 1F). To understand specific mechanisms for transport changes, we developed a structure-function model. In SLC transporters, uptake requires the protein to cycle from outward-facing states, which can bind substrate, to inward-facing states, where a substrate disassociates. Despite a recent deluge of SLC22 superfamily structures, there are still gaps in the transport cycle that must be filled for a complete structure-function model.[46–48] We therefore sought to describe the entire substrate coupling mechanism of OCT1 by reconstructing the entire transport cycle through a hybrid approach combining coevolutionary information, molecular dynamics (MD) simulations, and deep learning (see Figure S6 and STAR Methods). This approach has been extensively validated in well-studied systems such as sugar transporters and class A GPCRs.[49,50] Starting from a putatively inward-facing AlphaFold2 model, we find five stable states, spanning outward- to inward-open states (Figure 4A). We repeated this process with the substrate MPP^+ because it is a canonical cationic OCT1 substrate and chemically like the substrate we used in our cytotoxicity screen, SM73. The MPP^+ simulations yielded four models spanning outward- to inward-open states (Figure 4B).

To understand how substrate binding couples to OCT1 conformational changes, we analyzed the energetic response to MPP^+. Our results suggest that the resting conformation is the outward-facing one. Upon substrate binding, the thermodynamic equilibrium shifts toward inward-facing states (Figure 4C), where the substrate can be released, which then shifts the equilibrium back to the outward-facing face to repeat the cycle. Viewed in collective variable (CV) space, the substrate stabilizes the transition state, effectively "oiling" the transporter rocking alternating access motion (Figures 4A and 4B).[51]

We next sought to validate the predicted conformational states. Because there were no experimental SLC22 structures available when we started the MD simulations or when AlphaFold2 was trained, our fully computationally predicted models are independent of the recently published experimental structures. Out of the 12 experimental OCT1/2/3 structures, there are six outward-open, five inward-open, and one intermediate outward-occluded state.[46–48] To compare our models with solved structures, we calculated the root-mean-square deviation (RMSD) between each experimental structure and each predicted structure (Figure 4D), finding them quite similar (Figure 4E). Across homologs, outward (or inward) structures were closest to outward (inward) models, and OCT1 structures closer than homologs. The outward-occluded state, 8ET9, is of OCT2 bound to MPP^+, which is the same substrate used in our simulations, allowing us to directly compare how our computational method predicts the placement of MPP^+ and interaction with OCT1. The MPP^+ placement and orientation of side chains from the predicted and experimental models are strikingly similar (Figures 4F and 4G). Our predictions reveal additional states, namely A_{apo}, a more extreme outward-open state, as well as intermediate states along the minimum free energy pathway that represents (B_{mpp}, C_{apo}) and inward (C_{mpp}, D_{apo}) occluded states. Our approach thus accurately predicted OCT1 conformational states absent experimental data, identified the binding site and orientation of MPP^+, and described the complete conformational ensemble of OCT1.

Building a comprehensive structure-function model by integrating functional effects with thermodynamic ensembles

With confidence in the accuracy of the computational states, we built a structure-function model by integrating the DMS results. To directly compare how each residue contributes to the conformational cycle, we quantified interaction networks between contacting residues across all states. This yielded a score for each residue's importance to the conformational ensemble (Figure 5A, black). To identify which residues are crucial for substrate uptake from the unbiased mutational scan, we calculated a residue-wise functional importance score by taking the summed absolute value of survivability scores at each position across all mutations at least neutral within the expression screen (Figure 5A, blue). A comparison between the residue-wise conformational and functional scores has a qualitative similarity (Figure 5A), which a confusion matrix finds statistically significant (chi-square with Yates correction p value <0.1E−5, Figure 5B) implying that the residues that drive the conformational cycle are critical for function.

Using this integrated data, we first attempted to address the functional role of the ECD, a long-standing question in the field. From functional and conformational importance scores, we see the ECD (residues 45–145, Figures 5A, 5C, and 5D) is not enriched in interactions that drive the conformational cycle or are essential for substrate uptake. This gives unambiguous evidence that the ECD does not directly contribute to the conformational

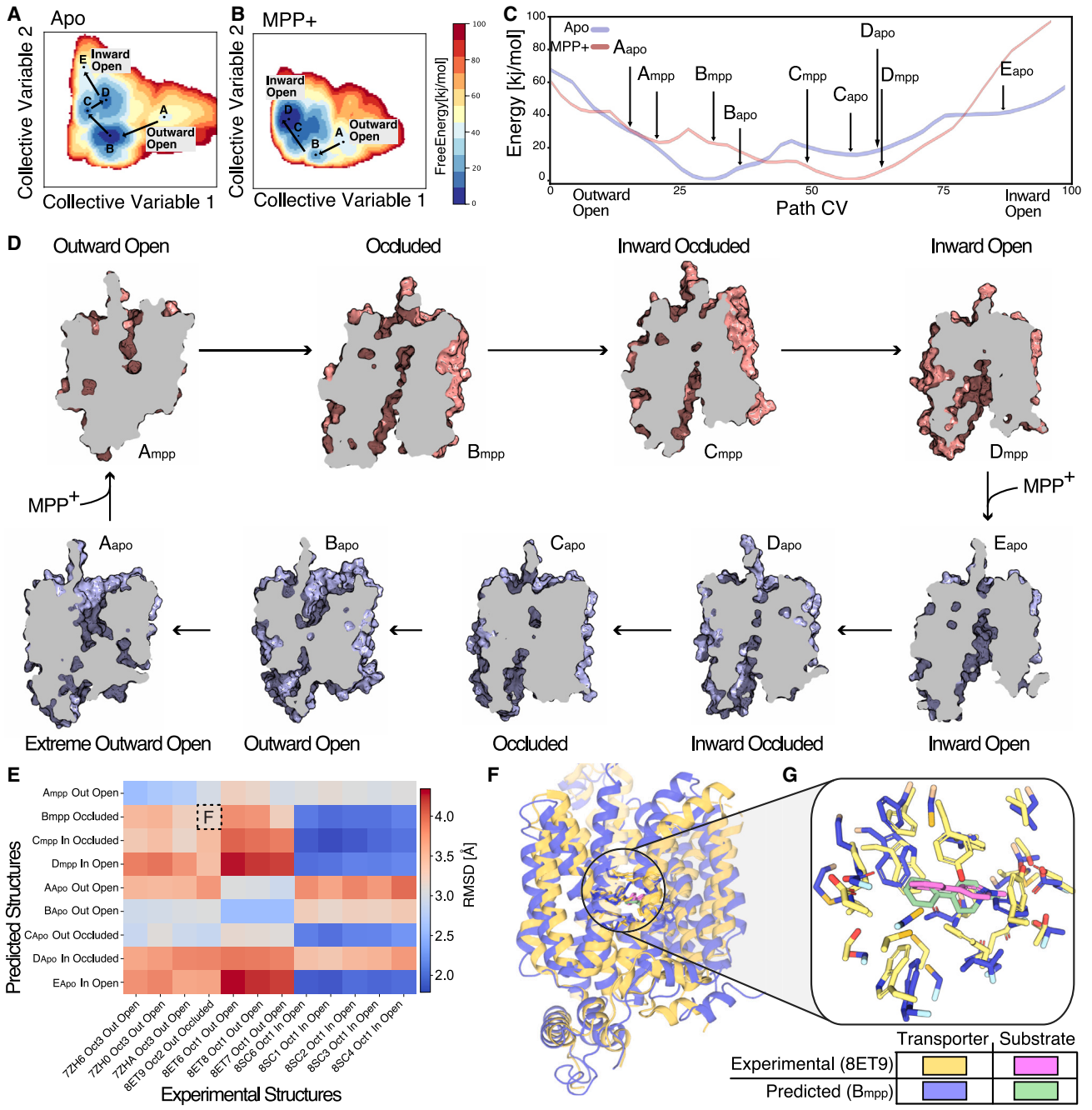


Figure 4. Accurate prediction of the thermodynamic OCT1 conformational ensemble with and without substrate

(A and B) 2D conformational landscapes of (A) apo and (B) MPP$^+$-bound OCT1 determined using enhanced-sampling MD simulations. Collective variables (x and y axes) derived from coevolutionary-based neural networks. Letters are local minima in the landscape. Colors represent free energy (kJ/mol), increasing from blue to red.

(C) 1D projection of the free energy landscapes from (A) and (B) with x axis representing the outward-open to inward-open path for Apo (blue) and MPP (red) simulations. y axis: free energy (kJ/mol).

(D) Sliced volume-filling models generated from meta-stable low energy states in Figures 5A and 5B with Apo (blue) and MPP+ (red).

(E) Comparison between experimental OCT structures (x axis) and predicted structures (y axis) from this work, colored by RMSD between complete structures. x axis labels note PDB accession, transporter consensus, and state of the structure.

(F and G) Comparison transporter backbone, substrate-binding residues, and substrate poses between experimental OCT2-MPP+ (8ET9, yellow-transporter, magenta-substrate) outward-occluded state and predicted B-MPP+ occluded state (blue-transporter, green-substrate).

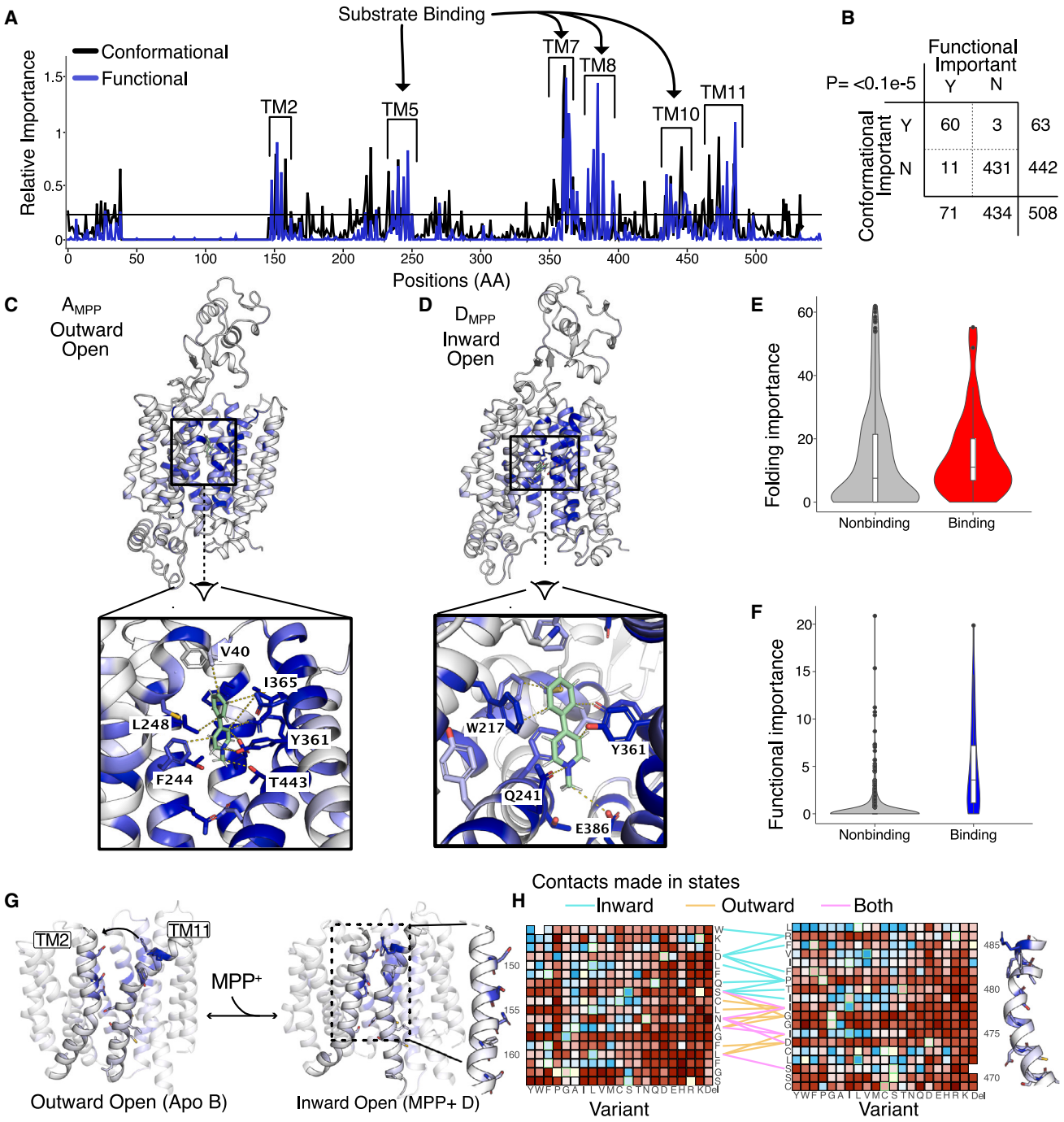


Figure 5. The integrated structure-function data reveal function residues that drive conformational cycle

(A) Relative (y axis) DMS-based functional (blue) and MD-based conformational (black) importance scores plotted across the OCT1 sequence (x axis). Regions with high functional and conformational scores are labeled, with substrate binding indicated by arrows.

(B) A confusion matrix between DMS and MD scores, calculated by setting a threshold at 0.2 for the functional and conformational importance scores and a chi-square test with Yates correction (p value < 0.1E−5, chi-square statistic 384).

(C and D) Functional importance scores mapped onto the outward-open (A_{mpp}, C) and inward-open (D_{mpp}, D) states colored from low (white) to high (blue) scores, with MPP green. Below each model inset of substrate-binding region with side-chain-MPP$^+$ interactions within 4 Å shown as lines.

(E and F) Violin plots comparing the score distributions of binding (<4 Å of ligand in any state) and other residues. Abundance importance scores (E, gray and red) and functional importance scores (F, blue and gray) on the y axes. Interacting residues appear to be enriched among functionally important (F) but not (E) abundance-important positions. Box plots are overlayed, indicating means (inner line), first and third quartiles (hinges), and 1.5 times the interquartile range from hinges (whiskers), and points beyond as dots.

(legend continued on next page)

cycle. Previous studies found that exchanging ECDs between homologs alter substrate specificity,[52] though, hinting at an indirect role in substrate uptake.

We next set out to determine which regions were most important for substrate uptake. We identified six TMs with high conformational and functional importance scores (TM2, TM5, TM7, TM8, TM10, and TM11, Figure 5A). Of these, TM5, TM7, TM8, and TM10 harbor residues that interact specifically with the substrate within one of the states (Figures 5C and 5D). We find two distinct state-specific binding modes, with the pose of the outward-open state not previously described (Figure 5E) and the occluded, inward-occluded, and inward-open states all having similar binding modes to those seen in the solved OCT2-MPP+ 8ET9 outward-occluded structure (Figures S5E–S5J). In the previously observed lower binding site, MPP+ is coordinated by a nexus of mainly aromatic residues (F159, W217, F244, Y361, and more distantly Y240), and stabilized by a long-range electrostatic interaction with E386 (Figures 5C and 5D). These interactions are in agreement with the recently published experimentally determined structures.[46–48] The upper site, which is distinct within our models, features fewer stabilizing interactions between OCT1 and MPP+ with the substrate-bound closer to the open extracellular gate residues, specifically V40 and I365, as well as lower aromatic residues that are close to the higher occupancy binding site (F244 and Y361). As would be expected, the substrate-binding residues (<5 angstroms from substrate in any state) are enriched in functionally important residues and not those important for abundance (Figures 5E and 5F). These residues were also implicated in substrate recognition and extracellular gate closing by Zeng et al.[46] Within our mutational scan, substitutions within these aromatic residues are deleterious, with mutations to non-aromatic residues particularly disruptive (Figure S5K). As OCT1 transports cationic substrates, it is not surprising that negatively charged residues coordinate the substrate, with E386 having the strongest phenotypes. More broadly, these negative residues are quite sensitive to mutations with a particularly drastic effect for charge swaps (Figure S5L). Given that the thermodynamic shift is induced by substrate presence, the alternative binding mode we observe could enforce the closing of the extracellular gate when the substrate first binds, starting the transition to the outward-occluded state.

Of the conformation-function important regions, we found TM5, TM7, TM8, and TM10 directly interact with the substrate to aid in transport, leaving TM2 and TM11 without a clear role. In the inward-open facing state, TM2 and TM11 make contacts near the extracellular face (Figure 5G). In contrast, further down the TM2-TM11 interface, some interactions are maintained across states, which we interpret as forming a pivot, while other contacts are state specific. D149 and R486 appear to stabilize the inward state, and P479 forms a kink likely important in enabling the geometries necessary for these motions, with all being sensitive to mutation. Further away from the extracellular face, a series of glycines on TM2 (G158 and G162) and TM11 (G474-475) likely provide the flexibility necessary for the conformational changes. In addition, a key interaction F158-D474 located immediately next to the flexible glycine hinges could stabilize the outward state. Therefore, we find conformationally important interactions between TM11 and TM2 likely stabilize the transition through the conformational cycle. Mutations at all these residues at the TM2-TM11 interface are disruptive to substrate uptake, implying a critical role in function likely by stabilizing the inward-facing conformational state (Figure 5H). The conformational ensembles complemented by multi-phenotypic DMS data thus allow us to propose an integrated and comprehensive structure-function model for OCT1 and discover the role of the TM2-TM11 interface in substrate uptake.

Inferring the biophysical mechanisms by which mutations alter OCT1 function

Our understanding of OCT1 function led us toward a biophysically motivated model for how OCT1 variants alter function in humans. We initially focused on a subset of 20 mutations that span both loss-of-function and gain-of-function phenotypes. These mutations are in regions crucial for function or abundance and 12 (V135I, R175H, E284A, D303E, D303G, M368T, A370T, S372G, A383S, E386K, I449T, and R486W) seen in humans in the genome aggregation database (gnomAD). To ensure that variants are indeed loss of function, we performed radioligand uptake experiments in HEK293T cells for all 20 across a range of substrates, including two canonical cationic OCT1 substrates MPP^+ and TEA^+ and pharmaceuticals sumatriptan and metformin (Figures 6A, 6B, and S6E–S6H). These assays recapitulate the results of our screens. MPP^+, which is chemically most like SM73, has the highest correlation between radioligand uptake experiments and the cytotoxicity screen. Two variants, R486E and R488E, which were not loss of function in our screen, also were not with MPP^+, did show strongly decreased transport with metformin and sumatriptan (Figures S6E–S6H), implying a change to substrate specificity.

Some of the mechanisms underlying the effects of a mutation can be directly intuited, such as charge repulsion the substrate-binding E386K mutation or biogenesis disrupting variants (W64R, V135I, R175H, and E284A). In other cases, where mutations that alter substrate uptake are likely changing the conformational ensemble, it is harder to interpret the effects. To develop a mechanistic understanding of how all mutations within OCT1 alter substrate uptake, we predicted the conformational landscape with and without the substrate MPP^+ for all mutants using the same MD pipeline as before (Figures 6B–6D, S6I, and S6J). Using fitness scores, conformational ensembles, and comparing the thermodynamic stability of inward- and outward-facing states, we developed a biophysically grounded prediction for how each mutation disrupts OCT1 (Figure 6E). From the MD-derived free energy landscapes, we find examples in

(G) Interface between TM helices 2 and 11 during conformational transitions involved in uptake: outward-open state (Apo A in 5D); right, inward open (MPP+ D in 5D). Functional scores are mapped in blue.

(H) Comparison of structural context and functional screen heatmap for TM helices 2 and 11. Left, TM 2, with corresponding section of survivability screen, and right, TM 11 with corresponding region of screen with ordering reversed to match TM 11 direction. Middle indicates contacts between residue Cα within 8 Å in inward-open (cyan), outward-open (wheat), or both (magenta) states are indicated. Heatmap is colored red to blue with increasing survivability scores.

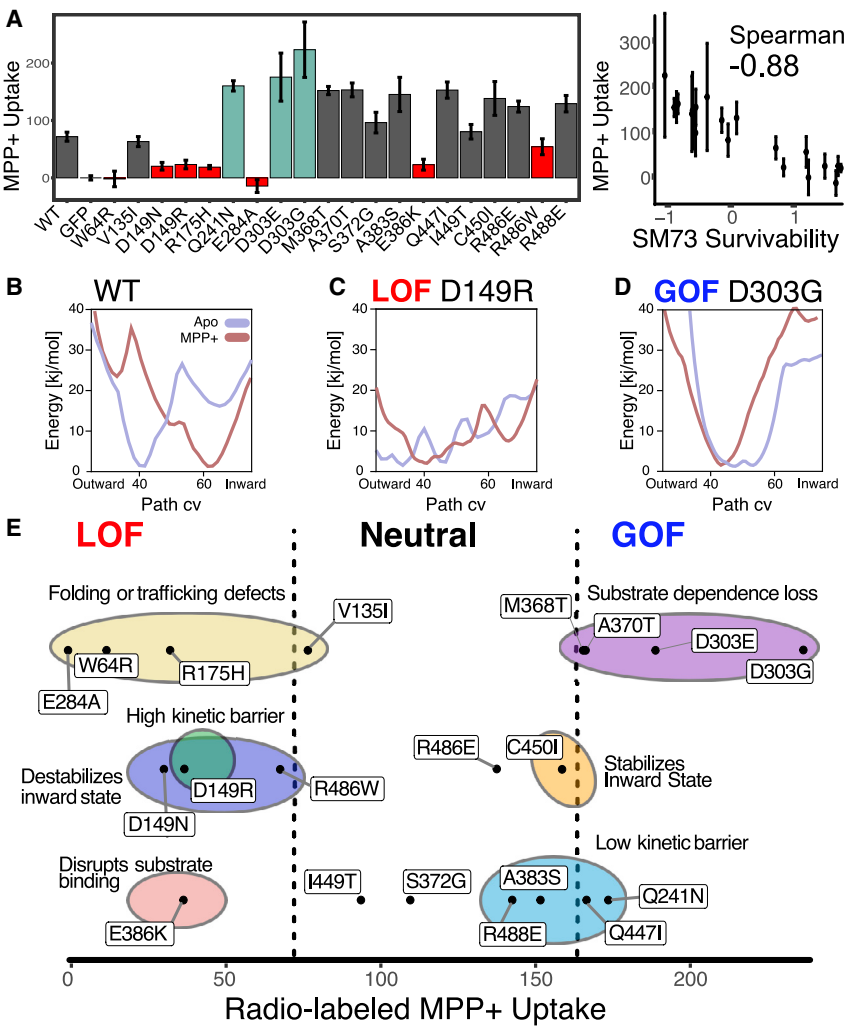


Figure 6. Biophysical means by which mutations alter OCT1 function

(A) Radio-labeled MPP$^+$ uptake experiments were conducted for a selection of mutations. Left: mean experimental uptake (n = 3; error bars: SEM). Right: comparison between SM73 DMS cytotoxicity scores and radio-labeled MPP+ uptake scores, with SEM error bars (Spearman rho: −0.88)

(B–D) 1D projection of free energy landscapes of Apo and liganded OCT1 WT, D149R, D303G, either Apo (blue) or MPP-bound (red). x axis indicates the path collective variable and y axis is free energy in kJ/mol.

(E) Mechanistic basis for mutational impact on OCT1. Mutations are plotted based on their MPP$^+$ uptake scores (x axis). Dotted line indicates cutoffs from screen on these mutants. Mechanistic classifications inferred from their free energy landscapes (STAR Methods) grouped in colored circles.

which mutations alter the conformational ensembles in a manner that can lead to loss of function: for example, D149R and D149N at the TM2-TM11 interfaces destabilize the substrate-dependent exchange between outward and inward states by a "high kinetic barrier" (Figures 6C and S6J). In addition, D149R is no longer stabilized in the inward-facing state in the presence of bound MPP+, and we conclude that loss of function "destabilizes the inward-facing state." Alternatively, gain-of-function mutations can be examined, too. D303G changes the ensemble such that the energetic barriers between outward and inward states are dramatically reduced, which could increase the exchange between outward- and inward-facing states as in sugar transporters.[53] Combining mutational scanning and computational ensemble predictions allows us to learn the biophysical mechanisms behind how human polymorphisms alter substrate uptake.

An atlas of OCT1 variant effects within human populations

To understand how OCT1 variation changes across people and populations, we compared our mechanistic mutational scanning data with a large-scale public database that provides a comprehensive and diverse collection of human genetic variation, gnomAD. gnomAD v2.1.1 contains 160 synonymous, 392 missense, 3 in-frame deletions, and 2 in-frame insertion variants for OCT1 (Table S1).[22] Prior to this work, only 24 of these variants had been characterized, whereas now we can use our DMS experiment to functionally annotate nearly all observed variants (498 out of 515 unique protein-level variants with both expression and survivability scores, 7 with one or the other, and 11 with none) (Figures 7A and 7B). For the 23 out of 24 that had been previously studied and are present in our DMS, we find a strong concordance between published MPP$^+$ uptake experiments and our survivability ($p < 0.0001$, Pearson r = −0.9) and expression ($p < 0.0001$, Pearson R = 0.91) screens (Table S1). After evaluating 375 missense variants and 2 in-frame deletion human variants, we found that 91 out of 368 (25%) of them had significant loss-of-function phenotypes based on cytotoxicity, while 125 out of 377 (34%) had significant loss of protein expression (Figures 7B and 7C). Within the 86 human loss-of-function variants, expression is the primary driver with only 10 out of 86 variants having disrupted substrate uptake.

Because only 24 out of 392 known human variants had been experimentally characterized, it was previously difficult to determine which populations have greater or lesser total allele frequency of loss-of-function variants in OCT1. Now with our systematic dataset, we explored whether reduced function variants were evenly distributed across all populations. Surprisingly, when comparing the six loss-of-function variants with mean allele frequency (MAF) ≥1%, we found that they were exceedingly rare (MAF <0.0005%) in East Asian populations, which also had the lowest total allele frequency of common poor-functioning variants at 0.1% compared with African (7.4%), South Asian (14.8%), European (20%), non-Finnish European

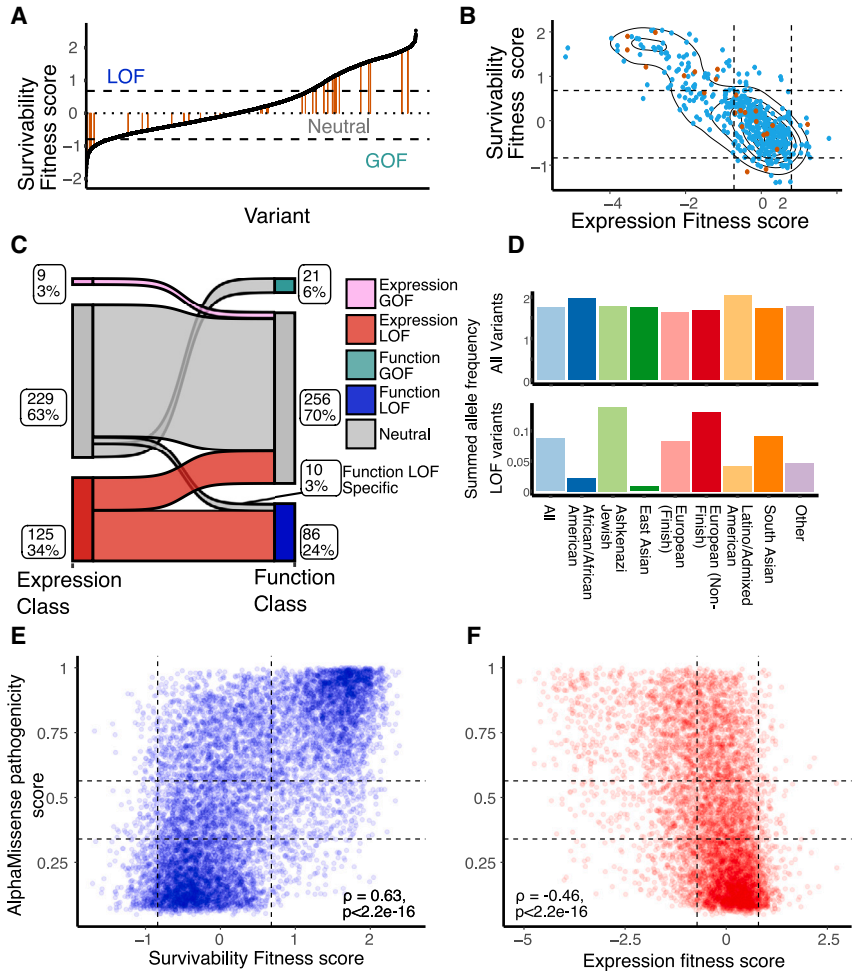


Figure 7. The impact of OCT1 polymorphisms in human populations

(A) Rank-ordered set of all 515 human polymorphisms from gnomAD (370 non-synonymous and 135 synonymous), with previously characterized variants indicated (39, orange). Cutoffs based on synonymous mutation distributions as dotted lines.

(B) Expression and survivability scores plotted for gnomAD variants, showing previously studied (orange) vs. unstudied (blue) variants (orange).

(C) Classification of non-synonymous human variants by predicted impact. Diagram shows how observed functional variation (right) is conditioned on expression impacts (left). Numbers and percentages of variants are next to classes.

(D) Allele frequencies of non-synonymous variants summed across populations by gnomAD ancestry. y axis: sum of minor allele frequencies for all (top) vs. loss-of-function (bottom) variants.

(E and F) Comparison of survivability (E) or VAMP-seq (F) scores (x axis) with AlphaMissense pathogenicity scores (y axis), with classification boundaries indicated with dotted lines. AlphaMissense classifies variants into benign, pathogenic, or ambiguous, corresponding to horizontal lines; vertical lines as in 1F. Spearman correlation coefficients (rho) and significance are plotted.

(27.4%), Ashkenazi (26.1%), and Latino (26.2%) populations, all of which had much higher total allele frequencies. Furthermore, when considering a total of 141 common and rare poor function mutations across all populations, East Asians had the lowest total allele frequency at 1.1%, followed by African (7.8%), South Asian (16.1%), European (20.1%), non-Finnish European (27.9%), Ashkenazi (26.5%), and Latino (26.5%) populations, all of which had much higher total allele frequencies (Figure 7D). A previous study observed the strongest divergence in the frequency of losing OCT1 activity between populations in Europe and East Asia but was limited in statistical power.[54]

Poor-functioning OCT1 variants broadly result in increased cholesterol levels in humans

We next aimed to understand the contribution of OCT1 variation to the diversity of human physiology. Poor-functioning OCT1 polymorphisms are associated with elevated low-density lipoprotein (LDL) cholesterol, total cholesterol, and triglycerides.[18] Based on this, we hypothesized that poor-functioning variants identified in our screens should also show an increase in these markers. To validate this, we used the UK Biobank (UKB), which contains exome sequencing and associated patient health records. Within OCT1, the UKB database contains 268 missense variants and 1 deletion variant, of which 264 are in our screens (Table S1). Of the 12 variants with a total MAF >0.1% in the UKB population, six have poor OCT1 function either validated in our experiments or from previous literature (R61C, C88R, S189L, G401S, M420del, and G465R). For all examined mutations, there is a strong association with elevated LDL, cholesterol, and triglycerides (Table S1). We used our DMS data to assign loss-of-function parameters to all UKB missense variants and in-frame deletion. In 180,000 individuals of European ancestry, variants that confer poor function were significantly associated with higher LDL ($p = 2.8 \times 10^{-10}$), total cholesterol ($p = 1.1 \times 10^{-9}$), and triglyceride ($p = 5.8 \times 10^{-11}$) levels (Table S1). In contrast, the 142 SNPs, which are similar to WT, exhibited weaker associations with LDL ($p = 0.04$), total cholesterol ($p = 0.088$), and triglyceride ($p=0.02$) levels. We also confirmed that these variants were not significantly associated with other non-OCT1 substrates, such as uric acid (Table S1).

DISCUSSION

By measuring how all possible missense mutations alter OCT1 expression and substrate uptake, we discover that loss-of-function mutations are caused by disruptions in expression. Using the expression screen, SLC22 subfamily conservation, and OCT1 alphafold2 structure, we discover the stability helix and ECD domain play a key role in biogenesis across SLC22 transporters. This implies that the results of expression-based screens would likely be similar across the SLC22 superfamily,

including other drug transporters (OCT2-3 and OAT1-3) and rare metabolic disease-associated transporters such as the carnitine transporter OCTN2. Additionally, we found that the N-terminal 6 TMs of OCT1 primarily contribute to expression, with the C-terminal 6 TMs contributing to transport. This agrees with prior *in vitro* biophysical studies of membrane proteins showing they co-translationally fold from N to C termini.[55,56] Mutations, therefore, could disrupt the co-translational folding of TMs 1–6, which may be the general rate-limiting step in MFS folding. By applying these approaches in other MFS transporters, we could determine whether this is a general phenomenon. If so, this could explain the evolutionary benefit of the fusion of two 6-TM domains, which has enabled MFS to play key roles in cell physiology across all kingdoms of life. This organization suggests a "division of labor" with the N-terminal TMs determining biogenesis, and C-terminal TMs modulating substrate uptake. Understanding MFS transporter biogenesis and function would have immense utility, as there are over 100 human MFS transporters, with many associated with disease or drug uptake, and in pathogenic bacteria MFS transporters are involved in antibiotic resistance.[57,58]

Besides a handful of intensely studied model membrane proteins, such as rhodopsin and CFTR,[39,59] we lack mechanistic models for how mutations disrupt membrane protein biogenesis to cause disease or alter drug efficacy.[60] Here, we show that mechanistic genetic screening provides a tractable, broadly applicable method to systematically identify which mutations affect biogenesis. Further, we find biogenesis underlies most functional defects, in line with previous results in a potassium channel, Kir2.1 and a more limited mutational study of human variants in OCTN2.[33,38,61] This approach could confer insight into membrane protein biogenesis and inform treatments by suggesting whether specific variants might be rescued by chaperones, for example.

Mechanistic variant effect predictive models are in their infancy. AlphaMissense, the current leading model, predicts a single variant pathogenicity score across the human proteome. In contrast, our screens reveal the multi-phenotypic effects of each mutation. We compared our abundance and cytotoxicity scores with AlphaMissense and found that they broadly agree, with a better fit for cytotoxicity than abundance (0.63 vs. −0.46 Spearman correlations, respectively, Figures 7E and 7F). AlphaMissense scores have low positional variance compared with our scores, implying lower sensitivity to mutational physiochemistry (Figure S7A). The lack of mechanistic foundation of these models limits their utility for understanding protein function and limits clinical genetics. For example, when interpreting the effects of mutations, we must acknowledge that most carriers of human variants are heterozygotes, as diploid organisms. Zygosity has been identified as a variant effect modulator for OCT1, which would be challenging to predict with existing data.[13,62,63] Also, dominant-negative mechanisms that alter ion channel function would be impossible to predict with existing models.[64] Without a mechanistic basis, we struggle to rationally develop targeted therapeutics or guide treatment.

Advances in structural biology allowed the membrane protein field to better understand how transporters work.[65,66] A powerful resource to assist understanding is the development of approaches that can predict entire conformational landscapes.[50,67,68] When we initiated this manuscript, there were no experimental structures within OCT1's SLC22 superfamily. Over the course of this study and manuscript preparation, 12 experimental models of OCTs were deposited, providing us the opportunity to compare our independently predicted models with experimental structure determination. One of the models (PDB: 8ET9) was even solved with the same substrate, MPP^+, we used for modeling, and we find remarkable similarities for the backbone, ligand, and side-chain orientations. This implies that, at least within the MFS family, existing computational modeling approaches are as accurate as experimental structure determination while providing a more extensive view of the conformational cycle. By complementing structural biology with mechanistic mutational scanning, we can address the often challenging "function" aspect in developing structure-function models.

Pharmacogenomic research on OCT1 unveiled population-specific variants, yet without systematic studies of variant effects it is not feasible to study the distribution of functional variants among populations.[54,69] Though OCT1 genetic polymorphisms are not yet routinely monitored in clinical pharmacogenetic analyses, there is a wealth of data suggesting that poor function variants collectively associate with differences in drug levels and response for drugs such as metformin,[13] fenoterol,[16] tramadol,[17] morphine,[62] and sumatriptan.[15] In this study, we demonstrate the feasibility of combining data from UKBiobank and DMS to investigate the effects of OCT1 variants on LDL, cholesterol, and triglycerides.

By integrating our DMS results with gnomAD to infer variant effects in populations and individuals, we find that loss of function in human variants most commonly is due to biogenesis defects. The distinct functionality of OCT1 across different populations can have clinical implications. For instance, morphine has been observed to exhibit stronger efficacy in East Asian and African populations, requiring significantly lower doses compared with Europeans.[70,71] An unbiased understanding for how variants impact drug transporters across populations could enable better clinical decision-making, customized dosage, and improved treatment.

Broadly applied approaches such as those in this study would yield a biophysics-based understanding of evolution, a deep understanding of protein biology, human genetic variation, and the diversity of life. The integration of mechanistic mutational scanning, biophysical models of folding and structure, and large-scale patient databases would allow a mechanistic prediction for the impact of variation. This would transform our basic biology for how proteins underlie our physiology, guide the diagnosis of disease, interindividual drug response, and development of therapeutics to better treat disease.

Limitations of the study

We acknowledge the limitations of our approach. For DMS experiments, we generated stable HEK293 cell lines with a variant downstream of a doxycycline-inducible promoter. This overexpression approach does not perfectly match within expression levels. Our VAMP-seq abundance readout indirectly measures this through a fluorescent protein proxy, in contrast to the

validation western blots. Finally, although we test variant effects of a subset of mutations on several medications, we do not test how all variants transport these medications, limiting our ability to determine the effects of all human variants in patients. Studies across multiple substrates will be required to determine how polymorphisms within OCT1 cause substrate-specific effects.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell cultures
- METHOD DETAILS
  - Library generation
  - Cell line generation and cell culture
  - Sequencing library preparation and genomic DNA extraction and data analysis
  - Transient transfection of the complementary remainder of the green fluorescent protein (mNG2-1-10)
  - Fluorescence-activated cell sorting for expression screen
  - Site-directed mutagenesis to create OCT1 variants
  - Transient transfection of plasmid containing OCT1 variants
  - Fluorescence microscopy
  - Western blotting
  - Radio-ligand uptake assay
  - SM73 Survivability Screen
  - Synthesis of SM73, a platinum anticancer agent
  - OCT1 AlphaFold2 structure mapping
  - Bioinformatics and sequence analysis
  - Molecular Dynamics simulations
  - Enhanced Sampling MD Simulations
  - Coevolution Driven Conformational Exploration
  - Mutational study in simulations
  - Protein network and computational importance analysis
  - Importance score profile peak assignment
  - Computational analysis of the biophysical basis of the effect mutations
  - GnomAD Database
  - UKBiobank total cholesterol, triglyceride and LDL levels
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2024.04.008.

ACKNOWLEDGMENTS

We are grateful to you for taking the time to read our manuscript. We are also grateful to Matthew Howard, James Fraser, Aashish Manglik, Mary Hennessy, Kliment Verba, Eric Greene, Margaux Pinney, Vijay Ramani, Robert Stroud, and Nicholas Reyes for helpful feedback and discussion as we developed and conducted this project, as well as extensive guidance on the manuscript. We are thankful to the reviewers for taking the time to give helpful comments and to Nadav Ahituv, Robert Edwards, Seth Shipman, and Anjali Jacob for giving us in-depth feedback on our revised draft. We also would like to thank the hard work of those in the UCSF Flow cytometry Core and Center for Advanced Technology, without whom none of the FACS or sequencing could have been possible. This work was supported by a Howard Hughes Medical Institute Hanna Gray Fellowship, UCSF QBI Fellowship, and Laboratory for Genomics Research pilot grant to W.C.-M.; the Knut and Alice Wallenberg Foundation, the Science for Life Laboratory, the Göran Gustafsson Foundation, the Swedish eScience Research Center, and the Swedish Research Council (VR 2019 - 02433) to L.D.; NIH 1F31AI157438 to D.D.T., NIH R01GM117163 to S.W.Y., NIH R01GM139875 to K.M.G., and NIH 1F32GM152977 to C.B.M. This work was also supported by the Noyce Initiative UC Partnerships in Computational Transformation Program to S.W.Y. Sequencing was performed at the UCSF CAT, supported by UCSF PBBR, RRP IMIA, and NIH 1S10OD028511-01 grants. The MD simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) on Beskow at the PDC Center for High-Performance Computing (PDC – HPC).

AUTHOR CONTRIBUTIONS

This study was designed by S.W.Y., C.B.M., D.M., M.L.K., L.D., K.M.G., and W.C.-M. S.W.Y., M.L.K., and W.C.-M. did the preliminary experiments for deciding on conditions for the screen. S.W.Y. made the libraries and conducted the mutational scans. X.Z. performed the FACS sorting. C.B.M. and W.C.-M. processed the data from the mutational scan and mapped these onto the structures with support from D.D.T. D.B.S. and S.W.Y., with support from P.R.G., made stable cell lines of mutants and conducted fluorescent microscopy. J.Y. and S.W.Y. did the radioligand uptake experiments. S.S.M. synthesized SM73. S.W.Y., C.B.M., L.K., and J.S.W. did the population genetics analysis. D.M. and L.D. conducted modeling based on the AlphaFold2 states and conducted MD simulations. All authors contributed to writing the manuscript and making figures.

DECLARATION OF INTERESTS

The cytotoxic ligand used in this study was patented in K.M.G. and S.M. (2015) "Platinum anticancer agents," US Patent US10392412B2. This compound, however, is not in commercial use.

Received: July 11, 2023
Revised: January 4, 2024
Accepted: April 15, 2024
Published: May 3, 2024

REFERENCES

1. Kell, D.B. (2021). The Transporter-Mediated Cellular Uptake and Efflux of Pharmaceutical Drugs and Biotechnology Products: How and Why Phospholipid Bilayer Transport Is Negligible in Real Biomembranes. Molecules *26*, 5629. https://doi.org/10.3390/molecules26185629.
2. Dobson, P.D., and Kell, D.B. (2008). Carrier-mediated cellular uptake of pharmaceutical drugs: an exception or the rule? Nat. Rev. Drug Discov. *7*, 205–220. https://doi.org/10.1038/nrd2438.
3. Kell, D.B., Dobson, P.D., and Oliver, S.G. (2011). Pharmaceutical drug transport: the issues and the implications that it is essentially carrier-mediated only. Drug Discov. Today *16*, 704–714. https://doi.org/10.1016/j.drudis.2011.05.010.
4. Ferrada, E., and Superti-Furga, G. (2022). A structure and evolutionary-based classification of solute carriers. iScience *25*, 105096. https://doi.org/10.1016/j.isci.2022.105096.
5. Lin, L., Yee, S.W., Kim, R.B., and Giacomini, K.M. (2015). SLC transporters as therapeutic targets: emerging opportunities. Nat. Rev. Drug Discov. *14*, 543–560. https://doi.org/10.1038/nrd4626.
6. International; Transporter Consortium, Giacomini, K.M., Huang, S.-M., Tweedie, D.J., Benet, L.Z., Brouwer, K.L.R., Chu, X., Dahlin, A., Evers, R., Fischer, V., et al. (2010). Membrane transporters in drug development. Nat. Rev. Drug Discov. *9*, 215–236. https://doi.org/10.1038/nrd3028.
7. Zamek-Gliszczynski, M.J., Sangha, V., Shen, H., Feng, B., Wittwer, M.B., Varma, M.V.S., Liang, X., Sugiyama, Y., Zhang, L., and Bendayan, R.; International Transporter Consortium (2022). Transporters in Drug Development: International Transporter Consortium Update on

Emerging Transporters of Clinical Importance. Clin. Pharmacol. Ther. *112*, 485–500. https://doi.org/10.1002/cpt.2644.

8. Zhang, L., Dresser, M.J., Gray, A.T., Yost, S.C., Terashita, S., and Giacomini, K.M. (1997). Cloning and functional expression of a human liver organic cation transporter. Mol. Pharmacol. *51*, 913–921. https://doi.org/10.1124/mol.51.6.913.
9. Gorboulev, V., Ulzheimer, J.C., Akhoundova, A., Ulzheimer-Teuber, I., Karbach, U., Quester, S., Baumann, C., Lang, F., Busch, A.E., and Koepsell, H. (1997). Cloning and characterization of two human polyspecific organic cation transporters. DNA Cell Biol. *16*, 871–881. https://doi.org/10.1089/dna.1997.16.871.
10. Yee, S.W., and Giacomini, K.M. (2021). Emerging Roles of the Human Solute Carrier 22 Family. Drug Metab. Dispos. *50*, 1193–1210. https://doi.org/10.1124/dmd.121.000702.
11. Chen, L., Shu, Y., Liang, X., Chen, E.C., Yee, S.W., Zur, A.A., Li, S., Xu, L., Keshari, K.R., Lin, M.J., et al. (2014). OCT1 is a high-capacity thiamine transporter that regulates hepatic steatosis and is a target of metformin. Proc. Natl. Acad. Sci. USA *111*, 9983–9988. https://doi.org/10.1073/pnas.1314939111.
12. Haberkorn, B., Fromm, M.F., and König, J. (2021). Transport of Drugs and Endogenous Compounds Mediated by Human OCT1: Studies in Single- and Double-Transfected Cell Models. Front. Pharmacol. *12*, 662535. https://doi.org/10.3389/fphar.2021.662535.
13. Shu, Y., Sheardown, S.A., Brown, C., Owen, R.P., Zhang, S., Castro, R.A., Ianculescu, A.G., Yue, L., Lo, J.C., Burchard, E.G., et al. (2007). Effect of genetic variation in the organic cation transporter 1 (OCT1) on metformin action. J. Clin. Invest. *117*, 1422–1431. https://doi.org/10.1172/JCI30558.
14. Meyer, M.J., Neumann, V.E., Friesacher, H.R., Zdrazil, B., Brockmöller, J., and Tzvetkov, M.V. (2019). Opioids as Substrates and Inhibitors of the Genetically Highly Variable Organic Cation Transporter OCT1. J. Med. Chem. *62*, 9890–9905. https://doi.org/10.1021/acs.jmedchem.9b01301.
15. Matthaei, J., Kuron, D., Faltraco, F., Knoch, T., Dos Santos Pereira, J.N., Abu Abed, M., Prukop, T., Brockmöller, J., and Tzvetkov, M.V. (2016). OCT1 mediates hepatic uptake of sumatriptan and loss-of-function OCT1 polymorphisms affect sumatriptan pharmacokinetics. Clin. Pharmacol. Ther. *99*, 633–641. https://doi.org/10.1002/cpt.317.
16. Tzvetkov, M.V., Matthaei, J., Pojar, S., Faltraco, F., Vogler, S., Prukop, T., Seitz, T., and Brockmöller, J. (2018). Increased Systemic Exposure and Stronger Cardiovascular and Metabolic Adverse Reactions to Fenoterol in Individuals with Heritable OCT1 Deficiency. Clin. Pharmacol. Ther. *103*, 868–878. https://doi.org/10.1002/cpt.812.
17. Stamer, U.M., Musshoff, F., Stüber, F., Brockmöller, J., Steffens, M., and Tzvetkov, M.V. (2016). Loss-of-function polymorphisms in the organic cation transporter OCT1 are associated with reduced postoperative tramadol consumption. Pain *157*, 2467–2475. https://doi.org/10.1097/j.pain.0000000000000662.
18. Liang, X., Yee, S.W., Chien, H.-C., Chen, E.C., Luo, Q., Zou, L., Piao, M., Mifune, A., Chen, L., Calvert, M.E., et al. (2018). Organic cation transporter 1 (OCT1) modulates multiple cardiometabolic traits through effects on hepatic thiamine content. PLoS Biol. *16*, e2002907. https://doi.org/10.1371/journal.pbio.2002907.
19. Long, T., Hicks, M., Yu, H.-C., Biggs, W.H., Kirkness, E.F., Menni, C., Zierer, J., Small, K.S., Mangino, M., Messier, H., et al. (2017). Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. Nat. Genet. *49*, 568–578. https://doi.org/10.1038/ng.3809.
20. Yousri, N.A., Fakhro, K.A., Robay, A., Rodriguez-Flores, J.L., Mohney, R.P., Zeriri, H., Odeh, T., Kader, S.A., Aldous, E.K., Thareja, G., et al. (2018). Whole-exome sequencing identifies common and rare variant metabolic QTLs in a Middle Eastern population. Nat. Commun. *9*, 333. https://doi.org/10.1038/s41467-017-01972-9.
21. Hoffmann, T.J., Theusch, E., Haldar, T., Ranatunga, D.K., Jorgenson, E., Medina, M.W., Kvale, M.N., Kwok, P.-Y., Schaefer, C., Krauss, R.M., et al. (2018). A large electronic-health-record-based genome-wide study of serum lipids. Nat. Genet. *50*, 401–413. https://doi.org/10.1038/s41588-018-0064-5.
22. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.
23. Fowler, D.M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. Nat. Methods *11*, 801–807. https://doi.org/10.1038/nmeth.3027.
24. Frazer, J., Notin, P., Dias, M., Gomez, A., Min, J.K., Brock, K., Gal, Y., and Marks, D.S. (2021). Disease variant prediction with deep generative models of evolutionary data. Nature *599*, 91–95. https://doi.org/10.1038/s41586-021-04043-8.
25. Brandes, N., Goldman, G., Wang, C.H., Ye, C.J., and Ntranos, V. (2023). Genome-wide prediction of disease variant effects with a deep protein language model. Nat. Genet. *55*, 1512–1522. https://doi.org/10.1038/s41588-023-01465-0.
26. Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L.H., Zielinski, M., Sargeant, T., et al. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science *381*, eadg7492. https://doi.org/10.1126/science.adg7492.
27. Sidransky, E. (2006). Heterozygosity for a Mendelian disorder as a risk factor for complex disease. Clin. Genet. *70*, 275–282. https://doi.org/10.1111/j.1399-0004.2006.00688.x.
28. Gerasimavicius, L., Livesey, B.J., and Marsh, J.A. (2022). Loss-of-function, gain-of-function and dominant-negative mutations have profoundly different effects on protein structure. Nat. Commun. *13*, 3895. https://doi.org/10.1038/s41467-022-31686-6.
29. Sivakumaran, S., Agakov, F., Theodoratou, E., Prendergast, J.G., Zgaga, L., Manolio, T., Rudan, I., McKeigue, P., Wilson, J.F., and Campbell, H. (2011). Abundant pleiotropy in human complex diseases and traits. Am. J. Hum. Genet. *89*, 607–618. https://doi.org/10.1016/j.ajhg.2011.10.004.
30. Faure, A.J., Domingo, J., Schmiedel, J.M., Hidalgo-Carcedo, C., Diss, G., and Lehner, B. (2022). Mapping the energetic and allosteric landscapes of protein binding domains. Nature *604*, 175–183. https://doi.org/10.1038/s41586-022-04586-4.
31. Amorosi, C.J., Chiasson, M.A., McDonald, M.G., Wong, L.H., Sitko, K.A., Boyle, G., Kowalski, J.P., Rettie, A.E., Fowler, D.M., and Dunham, M.J. (2021). Massively parallel characterization of CYP2C9 variant enzyme activity and abundance. Am. J. Hum. Genet. *108*, 1735–1751. https://doi.org/10.1016/j.ajhg.2021.07.001.
32. Markin, C.J., Mokhtari, D.A., Sunden, F., Appel, M.J., Akiva, E., Longwell, S.A., Sabatti, C., Herschlag, D., and Fordyce, P.M. (2021). Revealing enzyme functional architecture via high-throughput microfluidic enzyme kinetics. Science *373*, eabf8761. https://doi.org/10.1126/science.abf8761.
33. Coyote-Maestas, W., Nedrud, D., He, Y., and Schmidt, D. (2022). Determinants of trafficking, conduction, and disease within a K+ channel revealed through multiparametric deep mutational scanning. eLife *11*, e76903. https://doi.org/10.7554/eLife.76903.
34. Matreyek, K.A., Starita, L.M., Stephany, J.J., Martin, B., Chiasson, M.A., Gray, V.E., Kircher, M., Khechaduri, A., Dines, J.N., Hause, R.J., et al. (2018). Multiplex assessment of protein variant abundance by massively parallel sequencing. Nat. Genet. *50*, 874–882. https://doi.org/10.1038/s41588-018-0122-z.
35. Feng, S., Sekine, S., Pessino, V., Li, H., Leonetti, M.D., and Huang, B. (2017). Improved split fluorescent proteins for endogenous protein labeling. Nat. Commun. *8*, 370. https://doi.org/10.1038/s41467-017-00494-8.
36. Macdonald, C.B., Nedrud, D., Grimes, P.R., Trinidad, D., Fraser, J.S., and Coyote-Maestas, W. (2023). DIMPLE: deep insertion, deletion, and missense mutation libraries for exploring protein variation in evolution, disease, and biology. Genome Biol. *24*, 36. https://doi.org/10.1186/s13059-023-02880-6.

37. Matreyek, K.A., Stephany, J.J., Chiasson, M.A., Hasle, N., and Fowler, D.M. (2020). An improved platform for functional assessment of large protein libraries in mammalian cells. Nucleic Acids Res. *48*, e1. https://doi.org/10.1093/nar/gkz910.

38. Koleske, M.L., McInnes, G., Brown, J.E.H., Thomas, N., Hutchinson, K., Chin, M.Y., Koehl, A., Arkin, M.R., Schlessinger, A., Gallagher, R.C., et al. (2022). Functional genomics of OCTN2 variants informs protein-specific variant effect predictor for Carnitine Transporter Deficiency. Proc. Natl. Acad. Sci. USA *119*, e2210247119. https://doi.org/10.1073/pnas.2210247119.

39. McKee, A.G., McDonald, E.F., Penn, W.D., Kuntz, C.P., Noguera, K., Chamness, L.M., Roushar, F.J., Meiler, J., Oliver, K.E., Plate, L., and Schlebach, J.P. (2022). Elucidation of Global Trends in the Effects of VX-661 and VX-445 on the Expression of Clinical CFTR Variants. Preprint at bioRxiv. https://doi.org/10.1101/2022.10.14.512300.

40. Penn, W.D., McKee, A.G., Kuntz, C.P., Woods, H., Nash, V., Gruenhagen, T.C., Roushar, F.J., Chandak, M., Hemmerich, C., Rusch, D.B., et al. (2020). Probing biophysical sequence constraints within the transmembrane domains of rhodopsin by deep mutational scanning. Sci. Adv. *6*, eaay7505. https://doi.org/10.1126/sciadv.aay7505.

41. Young, H.J., Chan, M., Selvam, B., Szymanski, S.K., Shukla, D., and Procko, E. (2021). Deep Mutagenesis of a Transporter for Uptake of a Non-Native Substrate Identifies Conformationally Dynamic Regions. Preprint at bioRxiv. https://doi.org/10.1101/2021.04.19.440442.

42. Muhammad, A., Calandranis, M.E., Li, B., Yang, T., Blackwell, D.J., Harvey, M.L., Smith, J.E., Chew, A.E., Capra, J.A., Matreyek, K.A., et al. (2023). High-throughput functional mapping of variants in an arrhythmia gene, *KCNE1*, reveals novel biology. Preprint at bioRxiv. https://doi.org/10.1101/2023.04.28.538612.

43. Yariv, B., Yariv, E., Kessel, A., Masrati, G., Chorin, A.B., Martz, E., Mayrose, I., Pupko, T., and Ben-Tal, N. (2023). Using evolutionary data to make sense of macromolecules with a "face-lifted" ConSurf. Protein Sci. *32*, e4582. https://doi.org/10.1002/pro.4582.

44. Eddy, S.R. (2004). Where did the BLOSUM62 alignment score matrix come from? Nat. Biotechnol. *22*, 1035–1036. https://doi.org/10.1038/nbt0804-1035.

45. Burckhardt, G., and Wolff, N.A. (2000). Structure of renal organic anion and cation transporters. Am. J. Physiol. Ren. Physiol. *278*, F853–F866. https://doi.org/10.1152/ajprenal.2000.278.6.F853.

46. Zeng, Y.C., Sobti, M., Quinn, A., Smith, N.J., Brown, S.H.J., Vandenberg, J.I., Ryan, R.M., O'Mara, M.L., and Stewart, A.G. (2023). Structural basis of promiscuous substrate transport by Organic Cation Transporter 1. Nat. Commun. *14*, 6374. https://doi.org/10.1038/s41467-023-42086-9.

47. Suo, Y., Wright, N.J., Guterres, H., Fedor, J.G., Butay, K.J., Borgnia, M.J., Im, W., and Lee, S.-Y. (2023). Molecular basis of polyspecific drug and xenobiotic recognition by OCT1 and OCT2. Nat. Struct. Mol. Biol. *30*, 1001–1011. https://doi.org/10.1038/s41594-023-01017-4.

48. Khanppnavar, B., Maier, J., Herborg, F., Gradisch, R., Lazzarin, E., Luethi, D., Yang, J.-W., Qi, C., Holy, M., Jäntsch, K., et al. (2022). Structural basis of organic cation transporter-3 inhibition. Nat. Commun. *13*, 6714. https://doi.org/10.1038/s41467-022-34284-8.

49. McComas, S.E., Reichenbach, T., Mitrovic, D., Alleva, C., Bonaccorsi, M., Delemotte, L., and Drew, D. (2023). Determinants of sugar-induced influx in the mammalian fructose transporter GLUT5. eLife *12*, e84808. https://doi.org/10.7554/eLife.84808.

50. Mitrovic, D., McComas, S.E., Alleva, C., Bonaccorsi, M., Drew, D., and Delemotte, L. (2023). Reconstructing the transport cycle in the sugar porter superfamily using coevolution-powered machine learning. eLife *12*, e84805. https://doi.org/10.7554/eLife.84805.

51. Jardetzky, O. (1966). Simple allosteric model for membrane pumps. Nature *211*, 969–970. https://doi.org/10.1038/211969a0.

52. Meyer, M.J., Schreier, P.C.F., Basaran, M., Vlasova, S., Seitz, T., Brockmöller, J., Zdrazil, B., and Tzvetkov, M.V. (2022). Amino acids in transmembrane helix 1 confer major functional differences between human and mouse orthologs of the polyspecific membrane transporter OCT1. J Biol Chem. *298*, 101974. https://doi.org/10.1016/j.jbc.2022.101974.

53. Qureshi, A.A., Suades, A., Matsuoka, R., Brock, J., McComas, S.E., Nji, E., Orellana, L., Claesson, M., Delemotte, L., and Drew, D. (2020). The molecular basis for sugar import in malaria parasites. Nature *578*, 321–325. https://doi.org/10.1038/s41586-020-1963-z.

54. Seitz, T., Stalmann, R., Dalila, N., Chen, J., Pojar, S., Dos Santos Pereira, J.N., Krätzner, R., Brockmöller, J., and Tzvetkov, M.V. (2015). Global genetic analyses reveal strong inter-ethnic variability in the loss of activity of the organic cation transporter OCT1. Genome Med. *7*, 56. https://doi.org/10.1186/s13073-015-0172-0.

55. Curnow, P., and Booth, P.J. (2009). The transition state for integral membrane protein folding. Proc. Natl. Acad. Sci. USA *106*, 773–778. https://doi.org/10.1073/pnas.0806953106.

56. Choi, H.-K., Min, D., Kang, H., Shon, M.J., Rah, S.-H., Kim, H.C., Jeong, H., Choi, H.-J., Bowie, J.U., and Yoon, T.-Y. (2019). Watching helical membrane proteins fold reveals a common N-to-C-terminal folding pathway. Science *366*, 1150–1156. https://doi.org/10.1126/science.aaw8208.

57. Kim, J., Cater, R.J., Choy, B.C., and Mancia, F. (2021). Structural Insights into Transporter-Mediated Drug Resistance in Infectious Diseases. J. Mol. Biol. *433*, 167005. https://doi.org/10.1016/j.jmb.2021.167005.

58. Fluman, N., and Bibi, E. (2009). Bacterial multidrug transport through the lens of the major facilitator superfamily. Biochim. Biophys. Acta *1794*, 738–747. https://doi.org/10.1016/j.bbapap.2008.11.020.

59. Fiedorczuk, K., and Chen, J. (2022). Mechanism of CFTR correction by type I folding correctors. Cell *185*, 158–168.e11. https://doi.org/10.1016/j.cell.2021.12.009.

60. Marinko, J.T., Huang, H., Penn, W.D., Capra, J.A., Schlebach, J.P., and Sanders, C.R. (2019). Folding and misfolding of human membrane proteins in health and disease: From single molecules to cellular proteostasis. Chem. Rev. *119*, 5537–5606. https://doi.org/10.1021/acs.chemrev.8b00532.

61. Zhang, L., Sarangi, V., Ho, M.-F., Moon, I., Kalari, K.R., Wang, L., and Weinshilboum, R.M. (2021). SLCO1B1: Application and Limitations of Deep Mutational Scanning for Genomic Missense Variant Function. Drug Metab. Dispos. *49*, 395–404. https://doi.org/10.1124/dmd.120.000264.

62. Tzvetkov, M.V., dos Santos Pereira, J.N., Meineke, I., Saadatmand, A.R., Stingl, J.C., and Brockmöller, J. (2013). Morphine is a substrate of the organic cation transporter OCT1 and polymorphisms in OCT1 gene affect morphine pharmacokinetics after codeine administration. Biochem. Pharmacol. *86*, 666–678. https://doi.org/10.1016/j.bcp.2013.06.019.

63. Tzvetkov, M.V., Saadatmand, A.R., Lötsch, J., Tegeder, I., Stingl, J.C., and Brockmöller, J. (2011). Genetically polymorphic OCT1: another piece in the puzzle of the variable pharmacokinetics and pharmacodynamics of the opioidergic drug tramadol. Clin. Pharmacol. Ther. *90*, 143–150. https://doi.org/10.1038/clpt.2011.56.

64. Gargus, J.J. (2003). Unraveling monogenic channelopathies and their implications for complex polygenic disease. Am. J. Hum. Genet. *72*, 785–803. https://doi.org/10.1086/374317.

65. Cheng, Y. (2018). Membrane protein structural biology in the era of single particle cryo-EM. Curr. Opin. Struct. Biol. *52*, 58–63. https://doi.org/10.1016/j.sbi.2018.08.008.

66. Del Alamo, D., Sala, D., Mchaourab, H.S., and Meiler, J. (2022). Sampling alternative conformational states of transporters and receptors with AlphaFold2. eLife *11*, e75751. https://doi.org/10.7554/eLife.75751.

67. Fleetwood, O., Matricon, P., Carlsson, J., and Delemotte, L. (2020). Energy Landscapes Reveal Agonist Control of G Protein-Coupled Receptor Activation via Microswitches. Biochemistry *59*, 880–891. https://doi.org/10.1021/acs.biochem.9b00842.

68. Fleetwood, O., Carlsson, J., and Delemotte, L. (2021). Identification of ligand-specific G protein-coupled receptor states and prediction of

downstream efficacy via data-driven modeling. eLife *10*, e60715. https://doi.org/10.7554/eLife.60715.

69. Shu, Y., Leabman, M.K., Feng, B., Mangravite, L.M., Huang, C.C., Stryke, D., Kawamoto, M., Johns, S.J., DeYoung, J., Carlson, E., et al. (2003). Evolutionary conservation predicts function of variants of the human organic cation transporter, OCT1. Proc. Natl. Acad. Sci. USA *100*, 5902–5907. https://doi.org/10.1073/pnas.0730858100.
70. Letchuman, V., He, L., Mummaneni, P.V., Agarwal, N., Campbell, L.J., Shabani, S., Chan, A.K., Abrecht, C.R., Miller, C., Sankaran, S., et al. (2023). Racial Differences in Postoperative Opioid Prescribing Practices in Spine Surgical Patients. Neurosurgery *92*, 490–496. https://doi.org/10.1227/neu.0000000000002227.
71. Konstantatos, A.H., Imberger, G., Angliss, M., Cheng, C.H.K., Meng, A.Z.Y., and Chan, M.T.V. (2012). A prospective cohort study comparing early opioid requirement between Chinese from Hong Kong and Caucasian Australians after major abdominal surgery. Br. J. Anaesth. *109*, 797–803. https://doi.org/10.1093/bja/aes261.
72. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209. https://doi.org/10.1038/s41586-018-0579-z.
73. Bushnell, B. (2014). BBTools software package. https://jgi.doe.gov/data-and-tools/software-tools/bbtools/.
74. Van der Auwera, G.A., and O'Connor, B.D. (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (O'Reilly Media, Inc.).
75. Rubin, A.F., Gelman, H., Lucas, N., Bajjalieh, S.M., Papenfuss, A.T., Speed, T.P., and Fowler, D.M. (2017). A statistical framework for analyzing deep mutational scanning data. Genome Biol. *18*, 150. https://doi.org/10.1186/s13059-017-1272-5.
76. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.
77. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview, Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191. https://doi.org/10.1093/bioinformatics/btp033.
78. Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLoS Comput. Biol. *7*, e1002195. https://doi.org/10.1371/journal.pcbi.1002195.
79. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and 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.
80. Wu, E.L., Cheng, X., Jo, S., Rui, H., Song, K.C., Dávila-Contreras, E.M., Qi, Y., Lee, J., Monje-Galvan, V., Venable, R.M., et al. (2014). CHARMM-GUIMembrane Buildertoward realistic biological membrane simulations. J. Comput. Chem. *35*, 1997–2004. https://doi.org/10.1002/jcc.23702.
81. Yu, W., He, X., Vanommeslaeghe, K., and MacKerell, A.D., Jr. (2012). Extension of the CHARMM General Force Field to sulfonyl-containing compounds and its utility in biomolecular simulations. J. Comput. Chem. *33*, 2451–2468. https://doi.org/10.1002/jcc.23067.
82. Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., and Olson, A.J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J. Comput. Chem. *30*, 2785–2791. https://doi.org/10.1002/jcc.21256.
83. Hagberg, A., Swart, P., and Daniel, S.C. (2008). Exploring network structure, dynamics, and function using NetworkX. In Proceedings of the 7th Python in Science Conference.
84. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.
85. Grant, B.J., Skjaerven, L., and Yao, X.-Q. (2021). The Bio3D packages for structural bioinformatics. Protein Sci. *30*, 20–30. https://doi.org/10.1002/pro.3923.
86. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. Am. J. Hum. Genet. *89*, 82–93. https://doi.org/10.1016/j.ajhg.2011.05.029.
87. Peterson, R. (2021). Finding optimal normalizing transformations via bestNormalize. R J. *13*, 310. https://doi.org/10.32614/RJ-2021-041.
88. Nedrud, D., Coyote-Maestas, W., and Schmidt, D. (2021). A large-scale survey of pairwise epistasis reveals a mechanism for evolutionary expansion and specialization of PDZ domains. Proteins *89*, 899–914. https://doi.org/10.1002/prot.26067.
89. Yee, S.W., Vora, B., Oskotsky, T., Zou, L., Jakobsen, S., Enogieru, O.J., Koleske, M.L., Kosti, I., Rödin, M., Sirota, M., and Giacomini, K.M. (2021). Drugs in COVID-19 Clinical Trials: Predicting Transporter-Mediated Drug-Drug Interactions Using In Vitro Assays and Real-World Data. Clin. Pharmacol. Ther. *110*, 108–122. https://doi.org/10.1002/cpt.2236.
90. Zhang, S., Lovejoy, K.S., Shima, J.E., Lagpacan, L.L., Shu, Y., Lapuk, A., Chen, Y., Komori, T., Gray, J.W., Chen, X., et al. (2006). Organic cation transporters are determinants of oxaliplatin cytotoxicity. Cancer Res. *66*, 8847–8857. https://doi.org/10.1158/0008-5472.CAN-06-0769.
91. Giandomenico, C.M., Abrams, M.J., Murrer, B.A., Vollano, J.F., Rheinheimer, M.I., Wyer, S.B., Bossard, G.E., and Higgins, J.D. (1995). Carboxylation of Kinetically Inert Platinum(IV) Hydroxy Complexes. An Entr.acte.ee into Orally Active Platinum(IV) Antitumor Agents. Entr.acte.ee. Inorg. Chem. *34*, 1015–1021. https://doi.org/10.1021/ic00109a004.
92. Giacomini, K.M. (2019). Platinum anticancer agents. filed November 2, 2016, and published August 27, 2019. US Patent 10392412B2.
93. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.
94. UniProt Consortium (2023). UniProt: the Universal Protein knowledgebase in 2023 (2023). Nucleic Acids Res. *51*, D523–D531. https://doi.org/10.1093/nar/gkac1052.
95. Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi, M.L., Bork, P., Bridge, A., Colwell, L., et al. (2023). InterPro in 2022. Nucleic Acids Res. *51*, D418–D427. https://doi.org/10.1093/nar/gkac993.
96. Wheeler, T.J., Clements, J., and Finn, R.D. (2014). Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. BMC Bioinformatics *15*, 7. https://doi.org/10.1186/1471-2105-15-7.
97. NCBI (2004). BLOSUM90 Matrix. NCBI blast ftp site. https://ftp.ncbi.nih.gov/blast/matrices/BLOSUM90.
98. Szustakowski, J.D., Balasubramanian, S., Kvikstad, E., Khalid, S., Bronson, P.G., Sasson, A., Wong, E., Liu, D., Wade Davis, J., Haefliger, C., et al. (2021). Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. Nat. Genet. *53*, 942–948. https://doi.org/10.1038/s41588-021-00885-0.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-SLC22A1 antibody (2C5)	Novus	Cat#NBP1-51684; RRID:AB_11037784
Mouse monoclonal anti-GAPDH antibody	Millipore Sigma	Cat#MAB374; RRID:AB_2107445
Peroxidase-conjugated AffiniPure Goat Anti-Mouse IgG	Jackson ImmunoResearch	Cat#115-035-003; RRID:AB_10015289
Bacterial and Virus Strains		
MegaX DH10B $T1^R$ Electrocompetent Cells	ThermoFisher	Cat#C640003
Chemicals, Peptides, and Recombinant Proteins		
SM73	This paper	N/A
[^{3}H]-MPP^+	American Radiolabeled Chemicals	Cat#ART 0150
[^{14}C]-TEA	Perkin Elmer	Cat#NEC298050UC
[^{14}C]-metformin	Moravek	Cat#MC 2043)
[^{3}H]-sumatriptan	American Radiolabeled Chemicals	Cat#ART 1619
MicroScint-20	Perkin Elmer	Cat#6013621
Wheat Germ Agglutinin (WGA) Alexa Fluor 647 conjugate	Invitrogen	Cat#W32466
CellTiter-Glo	Promega	Cat#G7570
PrimeStar GXL DNA polymerase	Takara	Cat#R050A
NEBridge Golden Gate Assembly Kit (BsaI-HF v2)	NEB	Cat#E1601S
AP1903	MedChemExpress	Cat#HY-16046
SPRIselect beads	Beckman Coulter	Cat#B23318
Hoechst 33342 Solution	ThermoFisher	cat#62249
Puromycin	Life Technologies Corporation	Cat#NC2380542
Lipofectamine LTX	Thermo Fisher	Cat#15338500
SlowFade Gold Antifade mountant	Thermo Fisher	Cat#S36936
Protease inhibitor cocktail	Thermo Scientific	Cat#PI87786
Pierce ECL Western Blotting Substrate	Thermo Scientific	Cat#PI32209
Critical Commercial Assays		
Quick-DNA Microprep Plus Kit	Zymo Research	Cat#D4074
Quick-DNA Miniprep Plus Kit	Zymo Research	Cat#D4068
Nextera XT DNA Library Preparation Kit	Illumina	Cat#FC-131-1096
IDT for Illumina Nextera DNA Unique Dual Indexes Set D	Illumina	Cat#20042667
Q5 Site-Directed Mutagenesis Kit	NEB	Cat#E0554S
Pierce BCA Protein Assay Kits	Thermo Scientific	Cat# 23225
Deposited Data		
Raw sequencing reads	This paper	NCBI BioProject PRJNA980726
gnomAD v2.1.1	Karczewski et al.[22]	gnomAD v2.1.1 SLC22A1
SLC22a1 mutant imaging files	This paper	Mendeley Data: http://www.doi.org/10.17632/2pktbnf7kn.1
UK Biobank	Bycroft et al.[72]	https://www.ukbiobank.ac.uk/
Experimental Models: Cell Lines		
Human: HEK-293T	This paper	CRL-3216
Human: HEK-293T LLP-iCasp9-Blast Clone 4	Matreyek et al.[37]	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
Primers for OCT1 library generation, see Table S2, tab 1	This paper	N/A
Oligos for OCT1 library generation, see Table S2, tab 2	This paper	N/A
Landing_pad_for: GGAGCAATTCCACAACACTTTTGTC	This paper	N/A
P2A_cell_line_rev: CGGCGAAAGCAGCTGCC	This paper	N/A
Recombinant DNA		
Plasmid: pCAG-NLS-BxB1	Matreyek et al.[37]	Addgene # 51271
Plasmid: landing pad vector	Macdonald et al.[36]	N/A
Plasmid: pcDNA3.1(+)-GFP	Genscript	N/A
Software and Algorithms		
BBTools	Bushnell[73]	N/A
GATK	Van der Auwera and O'Connor[74]	N/A
Enrich2	Rubin et al.[75]	N/A
ConSurf	Yariv et al.[43]	N/A
MAFFT	Katoh and Standley.[76]	N/A
Jalview	Waterhouse et al.[77]	N/A
HMMER3	Eddy[78]	N/A
GROMACS2022	Abraham et al.[79]	N/A
CHARMM-GUI	Wu et al.[80]	N/A
cGENFF	Yu et al.[81]	N/A
AutoDock	Morris et al.[82]	N/A
NetworkX	Hagberg et al.[83]	N/A
Fiji	Schindelin et al.[84]	N/A
Bio3D R package	Grant et al.[85]	N/A
SKAT R package	Wu et al.[86]	N/A
bestNormalizeR package	Peterson[87]	N/A
Other		
Code for DMS pipeline, data analysis and figure generation, raw data and models	This paper	https://github.com/odcambc/OCT1_DMS, [https://doi.org/10.5281/zenodo.6962677]
FACSAria II cell sorter	BD	N/A
TapeStation 4200	Agilent	Cat#G2991BA
NovaSeq 6000	Illumina	N/A
Nikon Ti Microscope (Inverted) with Yokogawa CSU-22 Spinning Disk Confocal	Nikon	N/A

RESOURCE AVAILABILITY

Lead contact

Lead Contact: Willow Coyote-Maestas (willow.coyote-maestas@ucsf.edu)

Materials availability

Reagents generated within this study including variant libraries, stable cell lines, and mutant constructs will be shared upon request to the lead contact.

Data and code availability

- Sequencing data has been deposited at the NCBI Sequence Read Archive as BioProject PRJNA980726 and are publicly available as of publication. Final scores and processed data are included in the supplemental tables.
- All original code and all models have been deposited at Zenodo and is publicly available as of publication. Full imaging files have been deposited at Mendeley and are publicly available as of publication. DOIs listed in key resources table.

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell cultures

The cell lines generated for this study were previously described.[33,37] The HEK293T-based landing pad strains were a gift from Dr. Kenneth Matreyek. To make the cell lines, 1500 ng of library landing pad constructs (described below) were co-transfected with 1500ng of a BxB1 expression construct (pCAG-NLS-BxB1) using 10.5μL of lipofectamine LTX in 10 wells of a 6 well plate. All cells were cultured in 1X DMEM, 10% FBS, 1% sodium pyruvate, and 1% penicillin/streptomycin (D10). All cells were cultured at 37 °C at 5% CO2. The HEK293T-based cell line has a tetracycline induction cassette upstream of a BxB1 recombination site and split rapamycin analog inducible dimerizable Casp-9. Two days following transfection, expression of integrated genes or iCasp-9 selection system is induced by the addition of doxycycline (2 μg/μL, Sigma-Aldrich) to D10 media. Two days after induction with doxycycline, AP1903 is added (10nM, MedChemExpress) to cause dimerization of Casp9. Successful recombination shifts iCasp-9 out of frame, so only non-recombined cells will die from iCasp-9 induced apoptosis following the addition of AP1903. After two days of AP1903-Casp9 selection the media is changed back to D10 with doxycycline and cells are allowed to recover for two days. After allowing cells to recover for two days, media was changed to D10 with doxycycline and puromycin (2 μg/ml, Life Technologies Corporation), as an additional selection step to remove non-recombined cells. Cells remained in D10 plus doxycycline and puromycin for at least two days until cells stopped dying. Following puromycin treatment cells are detached, mixed, and seeded on two T75 flasks. Cells were then allowed to grow until they reached near confluence, then frozen in aliquots in a cryoprotectant media (2X HyClone, HyCryo, Cryopreservation Reagent). Landing pad cell lines were checked for mycoplasma using a detection kit.

METHOD DETAILS

Library generation

The OCT1 library generation was performed using the previously described method, DIMPLE (Deep Insertion, Deletion, and Missense Mutation Libraries for Exploring Protein Variation in Evolution, Disease, and Biology).[36] This method was adapted from SPINE and was used to design oligo to create mutations, insertions, and deletions.[88] The code was used to allow one to mutate each amino acid position to the other 19 amino acids, mutate to a synonymous codon, and a deletion (see code deposited at https://github.com/coywil26/DIMPLE). Primers and oligos that are designed and used for generating OCT1 libraries are listed in Table S2. Oligos for SLC22A1 was synthesized by Agilent (SurePrint Oligonucleotide, Agilent Technologies) to give 10 pmol of lyophilized DNA, which was then resuspended in 1 X TE buffer. Vector containing SLC22A1 was synthesized by Twist Bioscience in the High Copy Number Kanamycin backbone, the lyophilized plasmid DNA was resuspended to 10 ng/μL in 1 x TE buffer.

Sublibraries of different regions of SLC22A1 were PCR amplified using primer-specific and polymerase (PrimeStar GXL DNA polymerase) (Takara Bio). A total of 11 regions were PCR amplified, each in 50 μL reactions using 1 μL of the total OLS library as template and 30 cycles of PCR. 5 μL of the reactions was assessed by running on an agarose gel. The reactions were cleaned up using Zymo Clean and Concentrate kits (Zymo Research) and eluted in 10 μL of TE buffer. For each 11 regions, the plasmid was amplified to add on Golden Gate compatible type IIS restriction sites complementary to those encoded within the sublibrary oligos using Primestar GXL polymerase according to the manufacturer's instructions in 50 μL reactions using 1 μL of the template vector and 30 cycles of PCR. The entire PCR reaction was run on a 1% agarose gel and gel purified using a Zymoclean Gel DNA Recovery Kit.

Target gene backbone PCR product and the corresponding oligo sublibrary were assembled using BsaI-mediated Golden Gate cloning. Each 40 μL reaction was composed of 300 ng of backbone DNA, 50 ng of oligo sublibrary DNA, 2 μL BsaI-HF v2 Golden Gate enzyme mixture (New England Biolabs), 4 μL 10x T4 Ligase buffer, and brought up to a total volume of 40 μL with nuclease free water. These reactions were placed in a thermocycler with the following program: (i) 5 min at 37°C, (ii) 5min at 16°C, (iii) repeat (i) and (ii) 29 times, (iv) 5 minutes at 60°C, (v) hold at 10°C. Reactions were cleaned using Zymo Clean and Concentrate kits, eluted into 10 μL nucleus free water, and transformed into MegaX DH10B electrocompetent cells (Thermo Fisher) according to manufacturer's instructions.

MegaX DH10B cells were recovered for one hour at 37°C. A small subset of the transformed cells were plated at varying cell density to assess transformation efficiency. All transformations had at least 100x the number of transformed colonies compared to the library size. The remaining cell outgrowth was added to 30 mL LB with 50 μg/mL kanamycin and grown at 37°C with shaking until the OD reached 0.6 - 0.7. Library DNA was isolated by miniprep (Zymo Research). Sublibrary concentration was assessed using Qubit. Each sub-library of a given gene was pooled together at an equimolar ratio. These mixed libraries were assembled with a landing pad cell line compatible backbone containing a Carbenicillin resistance cassette and GSGSGS-mNeonGreen Fragment P2A-Puromycin cassette for positive selection.

Libraries were cloned into a landing pad vector containing a BxB1-compatible *attB* recombination site using BsmBI mediated golden gate cloning. We kept track of transformation efficiency to maintain library diversity that was at least 100x the size of a given

library. We designed the landing pad vector which we recombined the library into to contain BsmBI cutsites with compatible overhangs for the library to have an N terminal Kozak sequence and in-frame with a C-terminal GSGSGS-mNeonGreen Fragment P2A-Puromycin cassette for positive selection. The golden gate protocol we used was 42°C for 5 minutes then 16°C for 5 minutes repeated for 29 cycles followed by 60°C for 5 minutes before being stored at 4°C. This landing pad backbone was generated using Q5 site-directed mutagenesis, according to the manufacturer's suggestions.[36] Reactions were cleaned using Zymo Clean and Concentrate kits, eluted into 10 μL nucleus free water, and transformed into MegaX DH10B electrocompetent cells (Thermo Fisher) according to manufacturer's instructions. MegaX DH10B cells were recovered for one hour at 37°C. A small subset of the transformed cells was plated at varying cell densities to assess transformation efficiency. All transformations had at least 100x the number of transformed colonies compared to the library size. The remaining of the cells were plated into two large square plates (245mm x 245mm) containing 200 mL LB Amp.

Cell line generation and cell culture

The cell lines generated for this study were previously described.[33,37] To make the cell lines, 1500 ng of library landing pad constructs (described in previous section) were co-transfected with 1500ng of a BxB1 expression construct (pCAG-NLS-BxB1) using 10.5μL of lipofectamine LTX in 10 wells of a 6 well plate. All cells were cultured in 1X DMEM, 10% FBS, 1% sodium pyruvate, and 1% penicillin/streptomycin (D10). The HEK293T based cell line has a tetracycline induction cassette upstream of a BxB1 recombination site and split rapamycin analog inducible dimerizable Casp-9. Two days following transfection, expression of integrated genes or iCasp-9 selection system is induced by the addition of doxycycline (2 μg/μL, Sigma-Aldrich) to D10 media. Two days after induction with doxycycline, AP1903 is added (10nM, MedChemExpress) to cause dimerization of Casp9. Successful recombination shifts iCasp-9 out of frame, so only non-recombined cells will die from iCasp-9 induced apoptosis following the addition of AP1903. After two days of AP1903-Casp9 selection the media is changed back to D10 with doxycycline and cells are allowed to recover for two days. After allowing cells to recover for two days, media was changed to D10 with doxycycline and puromycin (2 μg/ml, Life Technologies Corporation), as an additional selection step to remove non-recombined cells. Cells remained in D10 plus doxycycline and puromycin for at least two days until cells stopped dying. Following puromycin treatment cells are detached, mixed, and seeded on two T75 flasks. Cells were then allowed to grow until they reached near confluence, then frozen in aliquots in a cryoprotectant media (2X HyClone, HyCryo, Cryopreservation Reagent).

Sequencing library preparation and genomic DNA extraction and data analysis

Genomic DNA was extracted using a Quick-DNA™ Microprep Plus Kit (Zymo Research) from cells sorted into four different GFP intensities. Genomic DNA was extracted using a Quick-DTNA™ Miniprep Plus Kit (Zymo Research) from cells treated with or without SM73 (1 μM) at different timepoints. The extracted genomic DNA from the miniprepped or micro kit prepped plasmid library were amplified using Landing_pad_backbone_for and P2A_cell_line_rev primers (Table S2).

Amplicons were prepared for sequencing using the Nextera XT DNA Library kit from Illumina with 1 ng of DNA input. Samples were indexed using the IDT for Illumina Nextera DNA Unique Dual Indexes Set D (96 Indexes) and SPRISelect beads (Beckman Coulter) at a 0.9x ratio were used for cleanup and final size selection. Each indexed tagmented library was quantified with Qubit HS as well as Agilent 4200 TapeStation. Samples were then pooled and sequenced on a NovaSeq 6000 SP300 flowcell in paired-end mode, generating fastq files for each sample after demultiplexing. Each fastq was then processed in parallel using the following workflow: adapter sequences and contaminants were removed using BBDuk, then paired reads were error corrected with BBMerge and then mapped to the reference sequence using BBMap with 15-mers (all from BBTools[73]). Variants in the mapped SAM file were called using the AnalyzeSaturationMutagenesis tool in GATK v4.[74] The output of this tool is a CSV containing the genotype of each distinct variant as well as the total number of reads. This was then further processed using a python script, which filtered out sequences that were not part of the designed variants, then formatted input files for Enrich2.[75] Enrichment scores were calculated from the collected processed files using weighted least squares and normalized using wild-type sequences. The final scores were then processed and plotted using R. Read counts are reported within Table S1. Enrich2 scores for all replicates and overall are reported in Table S1. Our libraries and cell lines show good coverage and low bias, as desired, with >97% of variants identified in our pre-screen sample and even representation across positions (Figure S1).

Transient transfection of the complementary remainder of the green fluorescent protein (mNG2-1-10)

HEK293T cells with the OCT1 library of variants are transfected with a construct containing green fluorescent protein (mNG2-1-10). This construct contains a CMV-mNG2-1-10-P2A-mCherry allowed selecting cells for sorting that have high expression of this transiently transfected gene. 600,000 - 700,000 cells/well (6-well poly-d-lysine coated) were seeded two days before the transfection. A total of 12 wells were plated including a well containing HEK293 cells without the landing pad integration site. The mNG2-1-10 constructs (3000 ng/well) were mixed with OptiMEM media (0.5 mL/well) (Life Technologies) and Lipofectamine LTX (10.5 μL/well). Samples were mixed by pipetting up and down several times, allowed to stand at room temperature for 15-20 min. The media with the cells seeded the day before were removed and fresh media (2 mL, DMEM-H21, 10% FBS) were added to each well. After 15-20 min incubation time of the mixture above, then add 0.5 mL of the mixture to each well of the 6-well plate. After 24 hours, cell media were removed, and fresh media were added to each well. (DMEM-H21, 10% FBS). After additional 24 h, cells are ready for subsequent experiments with fluorescent-activated cell sorting (FACS) to sort the cells into different GFP abundance.

Fluorescence-activated cell sorting for expression screen

In VAMP-seq, we add a small fragment of a split-fluorescent protein, to the cytoplasmic C-terminus of OCT1 (OCT1-mNG2-11) then cytoplasmically co-express the complementary remainder of the fluorescent protein (mNG2-1-10). Folded OCT1-mNG2-11 efficiently assembles with mNG2-1-10 to generate fluorescence, allowing fluorescence to serve as a FACS-seq-compatible proxy for OCT1 abundance. After conducting a 24-hour transient transfection of CMV-mNG2-1-10-P2A-mCherry (as described in the previous section), all cells were sorted using a BD FACSAria II cell sorter. Our gating strategy involved selecting single cells, then only sorting cells with high mCherry fluorescence thereby ensuring that the complementary of mNeonGreen fragment was not limiting (Figure S2). mNeongreen fluorescence was excited with a 488 nm laser and recorded with a 530/30 nm band pass filter, and mCherry fluorescence with a 561 nm laser and 585/42 nm band pass filter. The remaining pool was sorted into four subpopulations based on their mNeonGreen fluorescence levels. As the fluorescence distribution showed a skewed pattern, we designed gates for even bins by cell count (Figure S2). Subsequently, each cell population underwent the extraction of genomic DNA and the preparation of a library for sequencing, as described in the previous method section on sequencing DNA.

Site-directed mutagenesis to create OCT1 variants

Twenty-six missense variants were selected to validate the OCT1 function and abundance. Six of the variants (R61C, C88R, P117L, G401S, M420Del, G465) in the landing pad vector backbone generated using site-directed mutagenesis. To create each mutation, NEBaseChanger tool was used to design primers and Q5 Site-Directed Mutagenesis Kit was used to perform mutagenesis. For the other twenty of the OCT1 variants, we used the site-directed mutagenesis service offered by GenScript (New Jersey, USA). These twenty OCT1 variants are in a pcDNA3.1(+)-GFP (C-terminal) vector. For all the constructs, Sanger sequencing was used to confirm sequence.

Transient transfection of plasmid containing OCT1 variants

Twenty-six missense variants were selected to validate the OCT1 function and abundance. Six of the variants (R61C, C88R, P117L, G401S, M420Del, G465) and reference OCT1 were transfected into HEK293T cells containing the landing pad and were used to create stable cell lines. The six OCT1 variants are in the landing pad vector backbone. Twenty of the variants and reference OCT1 were transfected into HEK293 cells (UCSF Cell Culture Facility) to create transient expressing cell lines. The twenty OCT1 variants are in the pcDNA3.1(+)-GFP (C-terminal). Transient transfection of constructs encoding OCT1 reference and OCT1 variants was achieved by reverse transfection using Lipofectamine LTX transfection reagent (Thermo Fisher Scientific). 50,000 cells/well (96-well) were used in the reverse transfection. Constructs (100 ng/well) were mixed with OptiMEM media (20 μL/well) (Life Technologies) and Lipofectamine LTX (0.2 μL/well). Samples were mixed by pipetting up and down several times, allowed to stand at room temperature for 15-20 min, and then added to each well of the 96-well plate (20 μL) (poly-d-lysine coated). HEK293 cells were counted and seeded into wells at a density of 45,000 cells/well (100 μL/well) (96-well). After 24 hours, cell media were removed, and fresh media were added to each well. (DMEM-H21, 10% FBS.) After additional 24 h, cells are ready for subsequent experiments (uptake assays). To investigate the effects of missense variants on OCT1 expression, stable HEK293 cells expressing the variants or reference OCT1 were created. Specifically, eight missense variants (W64R, V135I, D149N, D149R, R175H, E284A, E386K, R486W) and a reference OCT1 sequence in the pcDNA3.1(+)-GFP (C-terminal) expression vector were transfected into HEK293 cells to generate stable cell lines. We subsequently analyzed OCT1 expression in these cells using confocal imaging.

Fluorescence microscopy

For the immunostaining experiments, HEK293 stable cell lines were plated onto poly-D-lysine treated 12-well plates with sterile coverslips at a density of 200,000 cells per well. Two days post-seeding, when cells reached 90-100% confluency, they were stained. On the day of staining, the cell media was removed and cells were washed with cold Hank's Balanced Salt Solution (HBSS, Thermo Fisher Scientific Inc.). The plasma membrane was stained first with Wheat Germ Agglutinin (WGA) Alexa Fluor 647 conjugate (Invitrogen Life Sciences Corporation), diluted in HBSS at 1:500, for 15 minutes at room temperature (RT). After staining, the solution was removed and cells were washed three times with HBSS. Cells were fixed with 3.7% formaldehyde in HBSS for 20 minutes, and after aspiration, cells were washed again three times with HBSS. The nucleus was then stained with Hoechst solution (Thermo Fisher Scientific Inc.), diluted at 1:2000 in HBSS, for 20 minutes at RT in darkness. After staining, the solution was aspirated and cells were washed twice with HBSS. Coverslips were carefully mounted on Superfrost Plus Microscope Slides (Thermo Fisher Scientific Inc.) with a drop of SlowFade Gold Antifade mountant (Thermo Fisher Scientific Inc.). Slides were left to dry overnight in darkness, and then imaged on an inverted Nikon Ti microscope equipped with a CSU-22 spinning disk confocal. All images were captured with the following channel settings; DAPI at 300ms exposure time and 50% laser power, FITC at 300ms exposure time and 25% laser power, and CY5 at 100ms exposure time and 5% laser power. The images were overlapped using Fiji software.[84]

Western blotting

HEK293 stable cells containing individual OCT1 variants were rinsed with PBS and lysed on ice with RIPA buffer (Thermo Scientific, Cat#89900) supplemented with a protease inhibitor cocktail (Thermo Scientific, Cat#PI87786). The lysed samples were centrifuged at 12,000 x g for 10 minutes at 4°C, and the supernatant was collected. The protein concentration of each sample in the supernatant

was determined with the Pierce BCA Protein Assay Kits (Thermo Scientific, Cat# 23225). Proteins were separated by electrophoresis using Blot Bis-Tris Mini Protein Gels (Invitrogen, Cat#NW04125BOX) and then transferred onto a polyvinylidene difluoride (PVDF) membrane (Millipore, Cat#IPFL00010). The transferred membrane was blocked with 5% skim milk (Bio-Rad, Cat#1706404) in 1 x TBST buffer (Thermo Scientific, Cat#28360) for 1 hour at RT. The membrane was then incubated with SLC22A1 antibody (2C5) (Novus, Cat#NBP1-51684) at 1:1000 dilution, for overnight at 4°C, followed by incubation with the corresponding horseradish peroxidase-labeled secondary antibody for 1 hour at RT. After detection of OCT1, the membrane was stripped in stripping buffer (Thermo Scientific, Cat#46430) at RT for 10 minutes, blocked again, and reprobed with anti-GAPDH antibody at a 1:1000 dilution (Millipore, Cat#MAB374) as a loading control. The proteins were detected using the Pierce ECL Western Blotting Substrate (Thermo Scientific, Cat#PI32209), and Western blot images were captured by use of the Amersham ImageQuant™ 800 Western blot imaging systems.

Radio-ligand uptake assay

The *in vitro* uptake assays were performed using methods developed in our laboratory as previously described.[38,89] After 48 hours of transient transfection of OCT1 variants in 96 well plates (poly-d-lysine coated), culture medium was removed, and cells were washed once with 250 μL HBSS. Four radiolabeled substrates of OCT1, [^{3}H]-MPP$^+$ (American Radiolabeled Chemicals, #ART 0150), [^{14}C]-TEA (Perkin Elmer, #NEC298050UC), [^{14}C]-metformin (Moravek, #MC 2043) and [^{3}H]-sumatriptan (American Radiolabeled Chemicals, #ART 1619), were used in the assay. To each well, 80 μL of Hanks' Balanced Salt solution (HBSS) (Gibco, #14025092) containing trace amounts of radioligand was added to each well and then incubated at 37 °C for 15 min. After a 15 min incubation period, the reaction mix was aspirated and the cells were washed 2x with ice-cold HBSS (250 μL). Next, 100 μL of MicroScint-20 (Perkin Elmer, #6013621) were added to each well and the plate was shaken at room temperature for 60-90 min. Radioactivity in each well was measured on a MicroBeta2 Microplate Counter for Radiometric Detection (PerkinElmer). To determine the function of each variant, each variant was normalized to OCT1 reference and expressed as a percentage after background uptake measured in the GFP vector was subtracted from both, calculated as follows: (OCT1 variant minus GFP) *100/ (OCT1 reference minus GFP). Each variant was assayed in triplicate on a 96 well plate and measured in three or more biological replicates.

SM73 Survivability Screen

Cytotoxicity of SM73, a platinum analog, in OCT1 cells was adapted from our previous study.[90] To determine inhibition potencies of SM73 to kill 50% of the cells (IC_{50}), cells were seeded on a 96-well plate (poly-d-lysine coated) at density 4000 cells/well. After 16-24 hours, the cells were incubated with different concentrations of SM73, starting from 100 μM to 1.7 nM, for 72 hours. Doxycycline (2 μg/mL) was incubated in the media with SM73 to induce the expression levels of OCT1. After 72 hours, media was removed and 50 μL of media was added to each well. Then 50 μL of Promega CellTiter-Glo luminescent reagent was added to each well. After 10-15 min incubation, transfer 80 μL to a 96 well plate (white, opaque). The luminescent cell viability was read on the Promega plate reader. The assay is based on quantification of the ATP present in each well, which relates to the number of viable cells in each well. The luminescent signal from cells treated with DMSO alone was considered the maximal signal (i.e. 100% cell growth). The percent cell growth of each OCT1 reference and mutants and at each concentration of SM73 were calculated. IC_{50} were determined using GraphPad (Prism v9.0).

For the deep mutational scanning, the cells (cells transfected with OCT1 library) were seeded into two T75 flasks. One flask only had doxycycline (2 μg/mL) and the other flask contained SM73 (1 μM) and doxycycline (2 μg/mL). After 48 hours, cells in both T75 flasks were split and transferred into new T75 flasks and treated with doxycycline (2 μg/mL) with or without SM73 1 μM. Approximately 6 million cells were seeded in the T75 flask, and the T75 flask treated with SM73 1 μM, was seeded with approximately 10 million cells as many will be killed after 48 hours. After another 48 hours, the above process was repeated. Every 48 hours, when the cells were split, approximately 1.5-3.0 million cells were collected for genomic DNA extraction. After another 48 hours, the above process was repeated. The cells in T75 flask are exposed to doxycycline (2 μg/mL) and SM73 1 μM or doxycycline (2 μg/mL) only for a total of 144 hours.

Synthesis of SM73, a platinum anticancer agent

The methodology of Christen and Higgins was adopted for the synthesis of the 4-phenylpiperazinyl platinum analog SM73.[91] The key species $[PtIICl_3NH_3]^-$ was obtained as its potassium salt, K[Pt"Cl_3NH_3] (Figure S7B), complex **1**) from cisplatin. Reaction of complex **1** with NaI followed by 1-phenylpiperazine afforded the mixed-halo Pt(II) complex **2**, which in turn was converted to the dichloride by formation of the aqua species with silver nitrate and the mixture exposed to HCl resulting in complex **3** (SM73). The synthesis of SM73 was previously described in a patent by Giacomini and More.[92] (SP-4-3)-Amminedichloro(cyclohexanamine)platinum(II) (Scheme 1, complex **3**, SM73). Synthesis of complex **1** was performed per the procedure described previously.[91] To an aqueous solution of **1** (0.375 g, 1.05 mmol) was added 0.4 g (2.0 mmol) of NaI in 1 mL of H_2O and stirred for 30 min at ambient temperature. Then, 0.55 mL of 1-phenylpiperazine (0.1 mL, 0.58 mmol) was added to the reaction mixture and stirred for an additional 4 hours at room temperature. The precipitated yellow solid was washed with water and ethanol. Additional washing with acetone provided compound **2** as a yellow solid, which was filtered and dried in vacuo (0.81 g, 52% yield).To a stirred suspension of **2** (0.50 g, 0.93 mmol) in 30 mL H_2O was added 0.253 g (1.49 mmol, 1.6 equiv) of $AgNO_3$ in the dark. The reaction was allowed to stir for 4 hours before being decolorized by activated charcoal. The precipitated AgCl was filtered and about 10 mL of concentrated HCl was added to the filtrate.

The solution was left undisturbed at room temperature for 2 h and then at 4°C overnight. The product was collected by vacuum filtration, followed by repeated washings with water, ethanol and ether to obtain SM73 as a light-yellow solid (0.42 g, 54% yield). Anal. calcd for $C_{10}H_{17}N_3Cl_2Pt$: C, 26.98; H, 3.85; N, 9.44. Found: C, 26.82; H, 3.47; N, 9.41. ^{1}H NMR (DMF-d_7): 7.16 (2H, t), 6.90–6.73 (3H, m), 4.12 (3H, b), 3.58–3.31 (2H, m), 3.19–3.03 (4H, m), 2.84–2.62 (2H, m).

OCT1 AlphaFold2 structure mapping

The AlphaFold2 predicted OCT1 structure was used for initial investigation.[93] We created scores for specific functional and abundance importance of each residue based on our DMS results. The abundance importance score at a given position was defined as the sum of the absolute values of the GFP scores for variants at that position with GFP scores >= 0.68 (corresponding to loss of function) and is reported in Table S1. The functional importance score for a given position was defined as the sum of the absolute values of the cytotoxicity scores for variants at that position with GFP scores < 0.68 (corresponding to non-loss of function) and is reported in Table S1. These scores were mapped onto the B-factors of the AlphaFold2 model using the Bio3D R package for visualization.[85]

Bioinformatics and sequence analysis

Conservation of OCT1 positions was determined using ConSurf.[43] Human and other specific sequences were obtained from UniProt.[94] They were aligned with MAFFT using L-INS-i mode.[76] The alignment was visualized in Jalview.[77] For broader sequence analysis, first all sequences matching the SLC22 InterPro family (IPR004749) were downloaded.[95] Similar sequences in the unaligned set were removed by removing redundancy at 50% with Jalview, then the resultant set was aligned with MAFFT using L-INS-i mode. An HMM profile was generated from the alignment using HMMER3,[78] then visualized with Skylign.[96] The BLOSUM90 values were obtained from the NCBI blast ftp site.[97]

Molecular Dynamics simulations

The Molecular Dynamics (MD) simulations of this work were done using GROMACS2022.[79] The AlphaFold2 model of human OCT1 in an inward-facing open conformation was downloaded and prepared using the CHARMM-GUI membrane builder,[80] where the ligands were parametrized using PDB coordinates and cGENFF.[81] Partial charges were forced by QM cluster calculations in implicit solvent using the PCM model and were calculated using the 6-31G basis set with Hartree-Fock level of theory. Moreover, CHARMM-GUI was also used to prepare the mutated OCT1 simulations by automatically choosing reasonable rotamers, and for all OCT1 simulations to create disulfide bridges between three pairs of cysteine residues in the soluble extracellular domain. The systems consisted of OCT1 embedded in a POPC and solvated in a 10mM KCl solution with an ion imbalance such that the net charge was zero. The initial PBC box size was set to 90x90x100 Å^3, and the systems were equilibrated according to the standard CHARMM-GUI protocol in the NVT ensemble. The production simulations were carried out using a 2 femtosecond timestep and a Nose-Hoover thermostat with separately coupled protein with ligand, POPC and water. For the barostat, the C-rescale scheme was used using a compressibility of 4.5e-5 and a tau of 5.0. The linear constraint solver (LINCS) algorithm was used to constrain all hydrogen-involving bonds, and particle mesh Ewald (PME) was utilized to calculate the long-range electrostatics above 12Å. For the LJ potential we used a neighbor-list cutoff with a switching function at 10Å and r_{vdw} of 12Å in consistency with the PME calculations.

The substrate MPP^+ was docked using AutoDock with default parameters and a 10x10x10 Å^2 box centered around the centre of mass of the entire TM region.[82] The top 100 poses were identical in most part thanks to the strong electrostatic forces.

Enhanced Sampling MD Simulations

In this work, we used both non-equilibrium steered MD with umbrella potentials as an additional restraint in collective variable space, as well as the accelerated weight histogram method (AWH) to drive the conformational change. The expression for the CVs stemmed from the coevolutionary analysis (see below - coevolution driven conformational exploration) and was implemented in the GROMACS transformation pull coordinate module. For the calculation of free energy surfaces along two collective variables, we used the accelerated weight histogram (AWH) method as natively implemented in GROMACS. For all simulations, we used an adaptive target distribution with the value of 1 below the cut-off of 120 kJ/mol and 0 above the cut-off with a smooth switching function between 100-140 kJ/mol. Additionally, we used the convolved bias potential to bias the reference coordinate and a harmonic coupling restraint between the reference and physical coordinates with varying force constants between 5,000-40,000 kJ/mol, Å.^{2}

AWH consists of two stages – the initial stage where the bias update size is decreased with each covering with the growth factor (set by default to 3.0), and after a criterion measuring the samples in the weight histogram and the actual number of collected samples is met the Wang-Landau algorithm is applied to achieve convergence. We recognized that our starting states may not be representative of the deepest minima, and thus we altered the growth factor to 2.0, which prolonged the initial stage such that more leeway would be given to account for time-dependent degrees of freedom to equilibrate. We could thus cover the conformational landscape several more times (<20) until the simulations exited the initial phase. We have heuristically seen that these estimations are more long-term reliable and produce flatter distributions without walkers getting stuck.

Coevolution Driven Conformational Exploration

The core idea behind the computational methodology is to combine evolutionary information with physical information through machine learning. To achieve this, we further refined the procedure described in Mitrovic et al..[50] To infer coevolving contacts between residue pairs, we chose the pre-constructed multiple sequence alignment of the PFAM transporter family PF00083. We then trained a Potts model using the direct coupling analysis (DCA) approach of the pseudo-maximum likelihood GREMLIN method, as described in Mitrovic et al.[50]. From the trained Potts model, we calculated the L2-norms of the pairwise coupling parameter matrices. After average product correction (APC) and alignment to the target human OCT1 sequence, we could assign these coevolutionary coupling scores to all matched residue-residue pairs in the human OCT1 protein.

We then aligned the top 500 coupling pairs with the contact map extracted from the retrieved AlphaFold2 model of human OCT1, where we identified the positions presenting a high coevolution score but no inter-residue contact as so-called false positives (FPs). We then clustered the FPs according to coupling strength and created five separate coordinates for exploration with the accelerated weight histogram (AWH) enhanced sampling algorithm in GROMACS2022. After 5ns of sampling with 4 walkers sharing a bias, we trained a linear kernel SVM to distinguish the most extreme end states from the starting state. We could extract the top 50 distances and their coefficients and analogously generate a new CV set to be used in AWH again by retraining the SVM model with the new data. To avoid retaining previously identified and explored degrees of freedom we restrained the already sampled degrees of freedom through a 10-times higher regularization parameter. We repeated this process 5 times, each time increasing the sampling towards an outward-open state. In total, we used 25ns per walker. These AWH simulations were run with a constant target distribution, and unrestrictive bounds. Using the last collective variable set, we defined bounds such that both start and end states would be sampled and set a cut-off of 120 kJ/mol. We did not decrease the histogram size until the histogram equilibrated (meaning each point must have sampled within 80% of the target distribution at that point) and started two walkers from the start- and end states.

After achieving convergence of the free energy landscapes (Figure S8), defined as the free energy surface that changes less than 1 kJ/mol over 50 ns per point, we estimated the errors of the free energy surface by the procedure described in Mitrovic et al.[50]. In essence, we counted the transition imbalance between neighboring points, which under equilibrium conditions and a well-estimated PMF, should be 0. By measuring the deviation from 0, we could estimate the size of the pointwise error in the free energy estimation (Figure S8). Additionally, for a reliable free energy estimate, all walkers need to have sufficient overlap in their sampling regions. Thus, we measured the effective probability distributions after convergence and the overlap between walkers (Figure S8D)

Mutational study in simulations

While 200-250 nanoseconds of simulation per walker were necessary to equilibrate, explore and converge the free energy landscape of the full conformational transition of the WT human OCT1, we instead focused our investigation of the mutants on the rocker switch mechanism, without sampling the opening of extra- and intracellular gates. We did this to alleviate the computational cost of running 18 mutants under 2 conditions each for a total of 36 separate simulation conditions. Under these conditions, we could iterate only 3 sets of explorations as described above. To enable a comparison between simulations, we used the optimal collective variable as derived for the WT system for the final AWH round converging the final free energy surfaces.

For each mutational condition, we visually inspected both the 3D structures from each basin to verify the functional assignment of each free energy basin, as well as the 2D landscapes themselves and quantitatively labeled each mutant based on this analysis on each.

Protein network and computational importance analysis

While the collective variables inferred above may capture the essential degrees of freedom for moving between states, it is neither unique in its description, nor can it describe all of the 3N intricacies of the conformational change. Thus, after convergence, we prolonged the simulations (with a nearly static bias) until an additional 100 coverings of the conformational landscape were made. Based on this sampling, we obtained the frame-wise free energy estimate as a function of the collective variable:

$$w(t) = e^{\frac{-U(\xi(t))}{RT}} \sum_{i}^{N_{frames}} u(i,t,\xi)$$

U is the pointwise free energy estimate, RT the thermodynamic constant at 298 K, and u(i,t,ξ) the binning procedure. Moreover, ξ(t) represents the value of the reaction coordinate at time t. Given a new reaction coordinate x, we could calculate each frame's projection according to X2:

$$E(x_i) = -RTln\left(\frac{\sum_{t=0}^{N_{frames}} u(i,t,x)w(t)}{\sum_{t=0}^{N_{frames}} w(t)}\right)$$

With this simple reweighting scheme accounting for the thermodynamic weight of each frame, we could reweight the free energy landscape onto any other degree of freedom. In this case, we systematically projected the free energy surface onto every minimum residue distance pair throughout the simulation. From these estimates, we measured the coupling between residues as the height of the highest free energy barrier along that coordinate, seeing as tightly coupled interactions should have two or more well-defined states with high barriers in-between, and less coupled pairs should have more spread-out distributions.

Additionally, we utilized the same reweighting procedure to back-map the 2D free energy landscape onto the determined SVM hyperplane normal vector, which would optimally describe the conformational change in 1D. This was primarily conducted for visualization purposes.

Importance score profile peak assignment

With a robust conformation-wide measure of coupling, we constructed a protein network in which edges between residues are estimated as coupling strength mentioned above. We then analyzed the network and calculated betweenness centrality with the shortest-path finder algorithm as implemented in python's NetworkX package.[83] In simple terms, the quantity calculated is the number of shortest paths that pass through each node in the network. Given that the centrality values correlated well with the experimentally determined mutational sensitivity (as determined by a confusion matrix), we anticipated that we could possibly provide a molecular-level rationalization of the effect of mutations, and we set out to identify which processes each of the peaks in the centrality plot were involved in. We focused on six different observable quantities, in particular residues which were primarily involved in:

1. Substrate dependency, determined by analyzing the difference in centrality values between the apo and substrate bound simulations.
2. Extracellular communication, determined by measuring the percentage of shortest paths that connected the ECD and TM domains.
3. The Rocker Switch mechanism, determined by mapping where the involved barriers localized on the free energy landscape. If the highest residue-residue pair barrier fell in the barrier region, we associated this with the main conformational change.
4. Gating contacts, determined by mapping the contacts to either the extracellular or intracellular gate opening and occlusion processes.
5. Inward-specific contacts, as residues involved in contacts only in the inward-facing states.
6. Outward-specific contacts, as residues involved in contacts only in the outward-facing states.

Computational analysis of the biophysical basis of the effect mutations

Using visual analysis of the free energy surfaces themselves and 3D structures of each basin, we categorized each mutational condition into a functional phenotype that could be linked to the experimentally determined fitness (Figure 3). In detail, the categories we identified were:

1. High kinetic barrier, which was determined by measuring whether the kinetic barrier between inward- and outward facing states was more than 5 kJ/mol higher than in WT simulations.
2. Low kinetic barrier, which was determined by measuring whether the kinetic barrier between inward- and outward facing states was more than 5 kJ/mol lower than in WT simulations.
3. Destabilizes inward state, which was identified in cases where the inward facing state is rarely accessed compared to the outward facing state.
4. Disrupts substrate binding, which was identified in the case of E386K, where the substrate did not occupy the correct binding mode, rather it was the only condition in which the substrate would spontaneously unbind from the protein.
5. Stabilizes inward state, which was identified in cases where the inward facing state was more stabilized than the outward-facing state.
6. Substrate dependence loss, which was determined by examining whether new basins appeared, disappeared, or changed in a significant capacity between substrate bound and Apo simulations.

In some cases (I449T, S372G) no label could be assigned. The folding or trafficking-defect mutations were not probed computationally since a functional localization never could be obtained experimentally.

GnomAD Database

To determine the frequencies of coding variants in human populations, we utilized the comprehensive gnomAD browser v2.1.1. Specifically, we extracted the variants for the SLC22A1 gene (transcript: ENST00000366963.4, NM_003057.3). We considered a variant to be functionally impaired if it had a score beyond two standard deviations of the synonymous variant distribution (Figure 1F). This corresponds to a cytotoxicity score of >= 0.68 or a GFP score of <= -0.72.

UKBiobank total cholesterol, triglyceride and LDL levels

In this study, we utilize exome sequencing results available from ~200,000 UK Biobank (UKB) participants[72,98] to perform genetic association analysis of coding missense and in-frame deletion variants in SLC22A1 with total cholesterol and LDL levels of European ancestry and other ethnic populations. This research was conducted with approved access to UK Biobank data under application number 14105 (PI: J.S. Witte) and in accordance with the UK Biobank Ethics and Governance Framework. UK Biobank data are publicly available by request from https://www.ukbiobank.ac.uk. Individual exome sequencing data, LDL cholesterol and total cholesterol were extracted from the 199,933 UKB participants. The LDL cholesterol and total cholesterol levels were normalized

using the R package bestNormalize.[87] The association analysis and rare variant analysis were performed using RStudio (2022.07.02). R package SKAT (version 2.2.4) was used for rare variant analysis.[86] In SKAT analyses, variants that have a cytotoxicity score of 0.6798075 or higher, or an abundance score of -0.7207321 or lower, are considered to have poor function.

We performed several association analyses to determine the significant association of OCT1 missense variant with total cholesterol and LDL levels and adjusted with body mass index (BMI), age, gender and for population structure (principal component analysis). The association analyses include:

1. Linear regression model to determine significant association of each missense variant with LDL and total cholesterol levels in European, South Asian and African populations (Table S1).
2. Combined the known poor function OCT1 missense variants with LDL and total cholesterol levels in European, South Asian and African populations. There are 5 variants that have poor OCT1 function in our study or from previous literature: p.R61C, p.C88R, p.G401S, p.M420del, p.G465R (Table S1).
3. Combined the common and less common poor function OCT1 missense variants that poor function and previously characterized: S29L, R61C, C88R, S189L, R206C, G220V, G401S, M420del, G465R (Table S1).
4. Performed SKAT analyses, including Burden test, which examines functional variants in SLC22A1 as causal factors with effects in the same direction (Table S1).

QUANTIFICATION AND STATISTICAL ANALYSIS

Details of statistical analyses are in text and figure legends corresponding to specific analysis. For DMS experiments, three biological replicates were performed and sequenced in a single run. Standard error reported by Enrich2 is available in Figure S3 and Table S1.