Structure of the human systemic RNAi defective transmembrane protein 1 (hSIDT1) reveals the conformational flexibility of its lipid binding domain

Vikas Navratna[1, 2*], Arvind Kumar [3#], Jaimin K. Rana[1, 2#], and Shyamal Mosalaganti[1,2,4*]

[1] Life Sciences Institute, University of Michigan, Ann Arbor, Michigan, 48109, United States.

[2] Department of Cell and Developmental Biology, University of Michigan, Ann Arbor, Michigan, 48109, United States.

[3] Thermo Fisher Scientific, Waltham, Massachusetts, 02451, United States.

[4]Department of Biophysics, College of Literature, Science and the Arts, University of Michigan, Ann Arbor, Michigan, 48109, United States.

[#] Authors contributed equally

*Correspondence to: navratna@umich.edu or mosalaga@umich.edu

Abstract

In *C. elegans*, the inter-cellular transport of the small non-coding RNA that causes systemic RNA interference (RNAi) is mediated by the transmembrane protein SID1, encoded by the *sid1* gene in the systemic RNA interference-defective (*sid*) loci. SID1 shares remarkable structural and sequence similarity with cholesterol uptake protein 1 (CHUP1), and they are classified together as members of the cholesterol uptake family (ChUP). Although systemic RNAi is not an evolutionarily conserved process, the *sid* gene products are found across the animal kingdom, suggesting a potential for the existence of other novel gene regulatory mechanisms mediated by small non-coding RNAs. Human homologs of *sid* gene products – hSIDT1 and hSIDT2 – are believed to mediate contact-dependent lipophilic small non-coding dsRNA transport. Here, we report the structure of recombinant full-length human SIDT1. We find that the extra-cytosolic domain (ECD) of hSIDT1 adopts a double jelly roll fold, and the transmembrane domain (TMD) exists as two modules – a flexible lipid binding domain (LBD) and a rigid TMD core. Our structural analyses provide insights into the potential role of the lipid binding domain in the function of hSIDT1 and other cholesterol uptake (ChUP) family members.

1. Introduction

RNA interference (RNAi) is post-transcriptional regulation of protein expression, wherein small non-coding double-stranded (ds) RNA such as exogenous small interfering RNA (siRNA) and endogenous microRNA (miRNA) target and induce degradation of homologous intracellular mRNA [1-3]. Although this RNA homology-mediated gene silencing phenomenon is being explored to experimentally alter gene expression with academic and therapeutic applications, the mechanism of dsRNA transport has been poorly understood [4-7]. Most small non-coding dsRNA transport occurs in bulk by plasmodesmata in plants or clathrin-mediated endocytosis in other eukaryotes. Alternatively, specific small non-coding RNA transport processes mediated by high-density lipoprotein binding scavenger receptors or lipophilic transporters have been reported [8-10]. In *C. elegans*, systemic RNAi is mediated by the products of the genes in systemic RNA interference-defective (*sid*) loci. SID1 (systemic RNA interference-defective protein 1) is a transmembrane protein involved in the inter-cellular transport of small non-coding RNA that cause systemic RNAi [11, 12]. In humans, there are two known systemic RNAi defective transmembrane proteins (SIDTs) - hSIDT1 and hSIDT2 – that are implicated in hepatocellular lipid homeostasis, glucose tolerance, insulin secretion, tumor development, cancer chemoresistance, RNA translocation and degradation during autophagy, and activation of innate immunity [13-19].

hSIDT1 localizes on plasma membrane, endolysosomes, and endoplasmic reticulum [13, 17, 20-22]. *In vitro* overexpression of hSIDT1 results in increased siRNA transport, and subsequently siRNA mediated gene silencing [23]. Oncogenic microRNA, miR-21, induces resistance in tumors to gemcitabine, a nucleoside analog and inhibitor of DNA synthesis that is

popularly used cancer therapeutic [20]. hSIDT1 is believed to transport miR-21 into the tumor cell, thereby promoting chemoresistance to gemcitabine. hSIDT1 lining the gastric pit cells in the stomach is indicated to uptake exogenous dietary miRNA in a pH-dependent manner, thereby enabling the internalized miRNA to regulate host gene expression [13]. The homolog of hSIDT1, hSIDT2, transports siRNA into late endosomes and lysosomes. SIDT2 overexpression enhances siRNA transport, while knockdown reduces RNA degradation, and a knockout disturbs the glucose homeostasis [15, 24]. Like SIDT1, SIDT2 is also implicated in crucial biological processes such as lysosomal membrane permeability, hepatic lipid homeostasis, apoptosis, and tumor proliferation [14, 20, 25, 26]. Despite SIDTs being a potential target for therapeutics against tumor progression, liver diseases, and Type II diabetes, their exploration as potential drug targets has been impeded by the lack of *in vitro* transport kinetic studies and high-resolution 3D structures of SIDTs.

The transporter classification database categorizes SIDTs as α-type channels (TCDB: 1.A.79), belonging to the cholesterol uptake (ChUP) or dsRNA uptake family of transmembrane proteins whose function and stability is regulated by cholesterol [27]. In fact, ChUP family members possess conserved transmembrane cholesterol recognition/interaction amino acid consensus motif (CRAC), suggesting a role in cholesterol binding and transport, and mutations in the CRAC motifs affect the localization of SIDTs [28-30]. SIDTs are believed to operate as oligomers – while the full-length SIDTs have been shown to form dimers both *in vitro* and *in situ,* the N-terminal extra-cytosolic domain (ECD) of SIDTs have been shown to assemble as tetramers [31-36]. It is believed that ECD domains of the SIDTs are crucial for substrate recognition and nucleic acid binding [31, 32, 37]. Recently, Qian et al., reported the structure of human SIDT2. They show that SIDT2 exists as a dimer and has a β-strand-rich ECD and a transmembrane domain

(TMD) comprised of 11 transmembrane helices (TMs). The dimer interface is predominantly formed by ECDs and partially by three TMs. They also suggest that the TM region of SIDT2 is capable of zinc-dependent lipase activity [33]. Here, we use single particle cryo-electron microscopy (cryo-EM) to determine the structure of heterologously expressed full-length hSIDT1 at ~3.4 Å. We observe that hSIDT1 exists as a dimer, where the dimer interface is carved by both ECD and TMD. The β-strand-rich ECD adopts a double jelly roll conformation, and two such double jelly rolls, one from each protomer, are juxtaposed in a C2 symmetry to form a dimer. The TMD can be divided into a dynamic lipid binding domain and a TMD core that is stabilized by four tiers of interactions. We also identify a phenylalanine highway within the TMD core, that is often seen in ATP-binding cassette sterol transporters. Finally, based on our structure, we highlight the intra-chain and inter-chain dynamics within the ChUP family of membrane proteins.

2. Results

2.1. Expression, purification of dimeric hSIDT1, and structure determination by cryo-EM

Purification and structural characterization of full-length human SIDTs have remained unexplored for a long time. We expressed full-length hSIDT1 fused to GFP on the C-terminus in HEK293S GnTI- cells by baculoviral transduction, along with a C-terminal Strep-II tag for purification (Table S1). To identify the detergent that is most suitable for purification of hSIDT1-GFP, we solubilized transiently transfected mammalian cells expressing hSIDT1-GFP in 1% detergent and analyzed the solubilized lysates by fluorescence-detection size-exclusion chromatography (FSEC; Fig. S1A-E). We noticed that hSIDT1-GFP solubilized in digitonin and DMNG resulted in homogenous samples. For cryo-EM sample preparation, we used baculoviral transduction to express large quantities of full-length hSIDT1-GFP. As ChUP family members contain two CRAC

domains, and are believed to be a cholesterol transporter, we excluded cholesterol and its analogues from our purification buffer to obtain the structure of an apo-hSIDT1. We chose digitonin to solubilize the large-scale mammalian cell membranes expressing hSIDT1-GFP and performed all the subsequent purification steps in digitonin to produce homogenous dimeric hSIDT1-GFP fusion (Fig. S1F-H). Henceforth, hSIDT1-GFP discussed in this report will be referred to as hSIDT1. We used cryo-EM to determine the structure of hSIDT1 to a global resolution of ~3.4Å, with the N-terminal extra-cytosolic domain (ECD) at a better local resolution than the C-terminal transmembrane domain (TMD) (Table 1, Fig. S2, and Fig. S3A-C). We were able to model the entire protein except for the first 40 amino acids in the disordered N-terminus, the cytosolic loops CL1 and CL4, a portion of β7'-β8' hairpin (Q278-N286), and the GFP on the C-terminus (Fig. S3D). Amongst the regions that were modeled, the resolution for TMs6-9 was poor, so we used the unsharpened map to place the Cα traces of these helices into the density and truncated the side chains of amino acids in these regions (Fig. S3D inset).

2.2. Architecture of hSIDT1

hSIDT1 is a dimer where the protomers are related to each other by a 2-fold (C2) rotational symmetry, with the axis of rotation being perpendicular to the plane of the lipid bilayer (Fig. 1A-C). Each protomer has an N-terminal β-sheet rich ECD (black box) and a C-terminal all α-helical TMD, which can be divided into two parts – the TMD core (blue box) and the lipid binding domain (LBD; gray box) (Fig. 1E-F). All three regions participate in the assembly of the dimer (Fig. 2).

2.2.1. Extra-cytosolic domain (ECD)

hSIDT1 ECD contains a 260 amino acid domain (A42-I301) comprised of two tandem β-sandwiches arranged like a double jelly roll (Fig. 1E, Fig. 1G, and Fig. 2A). A classic jelly roll fold is an 8-stranded antiparallel β-sheet sandwich with the strands β1, β3, β6, and β8 forming the top sheet, such that strand order of top sheet is β1-β8-β3-β6 and of the bottom sheet is β2-β7-β4-β5 [38]. Within ECD of each SIDT1 protomer we see two such jelly rolls, JR1 (G45 to K167) and JR2 (R172 to I301), connected in tandem by a flexible linker (H168 to L171) such that the top sheet of JR1 and bottom sheet of JR2 are in the same plane (Fig. 2A). JR1 and JR2 arrangement allows the formation of an extensive network of aromatic ring stacking interactions between β6, β5-β6 hairpin, and β6-β7 hairpin of JR1 and β2', β7', β7'-β8' hairpin, and β4'-β5' hairpin of JR2. In addition, JR1 and JR2 are held together, towards the dimer interface, by a disulfide bond (C130-C222) between cysteines in β6-β7 hairpin and β4'-β5' hairpin (Fig. S4A and Fig. 2A). The ECD is also stabilized by three pairs of salt bridges (R98-D148, H178-E294, and R289-D268). The bottom strands of the double jelly roll line the SIDT1 dimer interface of the protein. ECD has five predicted N-glycosylation sites – N57, N67, N83, N136, and N282. We used HEK293SGnTI- cells that lack N-acetylglucosaminyltransferase I activity for recombinant hSIDT1 production, thus yielding a protein that lacks complex *N*-glycosylation. Although we observe density for the *N*-acetylglucosamine (NAG) at the Asn residues that are predicted glycosylation sites, our resolution is relatively poor to model NAG residues in the density unambiguously.

2.2.2. Transmembrane domain (TMD)

TMD of the SIDT1 protomer is about 500 amino acids long and comprises 11 TM helices. TMs 1-4, along with TMs 10 & 11, form the TMD core of the protein, and TMs 5-9 form the lipid binding domain (LBD) (Fig 1E-F). TM1 and TM2 are connected by a disordered cytoplasmic loop

(CL1) that is over 100 amino acids and is believed to form a cytosolic domain (CD) with an RNA binding ability. Owing to its flexibility, we do not see density for CL1 in our structure. ECD sits atop TMD core via JR2-TMD core interactions and does not interact with LBD.

TMD core

The TMD core is stabilized by four tiers of evolutionarily conserved interactions – the TMD-ECD interface, the TMD disulfides, a metal ion binding site, and a phenylalanine highway (Fig. 2B-C and Fig. 4E-F). Extensive network of electrostatic interactions and salt-bridges (K242-E777 and D477-R776) at the TMD-ECD interface between EL1, EL5, β3'-β4' hairpin, β5', β6', and β7'-β8' hairpin position the ECD over the TMD core such that the access to the metal binding site from the extra-cytosolic space is blocked by ECD (Fig. 1E, Fig. 2B, Fig. 2D, and Fig. 4B). Towards the TMD-ECD interface, the TMD core is held together by two conserved disulfides (C479-C565 and C485-C782) formed between EL1, EL2, and EL5, which stitch together the TMs2-4, TM10, and TM11 on the extra-cytosolic side (Fig. 1E, Fig. 2B, and Fig. 4E-F). The TMs3-5 and TM11 carve a conserved metal ion binding site lined by amino acids N501, S559, H563, D574, H791, and H795 (Fig. 2C). This site has been suggested to be a zinc site and is believed to be essential for the activity of SIDT like proteins [33-36]. While we observe density coordinated by H563, D574, H791, and H795 in our structure, due to a lack of convincing biochemical evidence, we have not modeled a metal ion at this site in the structure (Fig. 2C). ATP-binding cassette (ABC) sterol transporters such as ABCG family have a conserved structural motif called phenylalanine highway, where a series of phenylalanines line the dimer interface and enable binding of cholesterol [39]. Like ABCG, these phenylalanine highways have been suggested to exist in other cholesterol binding proteins such as NPC1, PTCH1, and ChUP family of proteins, albeit not always in the dimer interface [39]. In SIDT1, we notice a similar phenylalanine highway - TM11

(D786-L810) and TM1 (K302-R334) are studded with 6 and 5 phenylalanine residues, respectively. TM11 and TM1, along with TM3 and TM10, form a pouch facing the lipid bilayer lined with at least 10 phenylalanine residues (Fig. 4E). The aromatic ring stacking interactions in this pouch help stabilize the TM core along with the disulfides, metal ion binding site, and the TMD-ECD interface interactions. In addition, we notice a series of ordered lipids and detergent bound to this region in our cryo-EM density map (Fig. 4D). However, because of the relatively poor local resolution in this area, we have not modeled any lipid or detergent in this density (Fig. S3C).

Lipid binding domain (LBD)

TMs 5-9 form a conformationally flexible domain which is poorly resolved in our cryo-EM map (Fig. 1, Fig. S3, and Fig 4). Henceforth, we will refer to this region of the protein as the lipid binding domain (LBD). The local resolution for the LBD in our structure is poorest of all regions, especially for TMs 6-8 (Fig. 1D, Fig. S3C, and S3D). As a result, we have modeled only Cα traces for most part of the helices in this region. TM5 and TM6 of the LBD line the dimer interface along with TM2 of the TMD core. TMs 7-8 lie on the periphery of the protein and protrude into the lipid bilayer away from the TMD core. TM5 and TM9 align with TMD core, and at the interface of this LBD and TMD core we also notice unexplained ordered density which could either be lipid or detergent (Fig. 1D, Fig. 1G, and Fig. 4D).

2.2.3. hSIDT1 dimer interface

SIDT1 dimer interface spans ~2800 Å^2, and it can be divided into three sub-interfaces (SI) – 1, 2, and 3 (Fig. 1 and Fig. 2D). We name these sub-interfaces SI-1, SI-2, & SI-3. These are formed by

JR1 domains (JR1-A and JR1-B), JR2 domains (JR2-A and JR2-B), and the TMDs (TMD-A and TMD-B) of each protomers (A & B), respectively. Because of the protomers being related by C2 rotational symmetry, the secondary structure elements in chain A at the dimer interface are arranged as a mirror image of the same elements from chain B (Fig. 2D). SI-1 and SI-2 contribute to most of the interactions that stabilize the SIDT1 dimer and constitute 65% of the dimer interface. SI-1 is formed by the strands β2, β4, β5, and β7 from JR1-A and JR1-B, which are juxtaposed at the dimer interface and are stabilized by a series of electrostatic and hydrophobic interactions contributed primarily by amino acid stretches N90-S105 and F146-M152 (Fig. 2D and Fig. S4C). A set of three charged residues, E61, R98, and D148, from both protomers form a network of salt bridges that stabilize SI-1. SI-2 is formed by a β4'-β5' hairpin, that also stabilizes JR-1 and JR-2 interaction (Fig. 2A and 2D). β4'-β5' hairpin from JR2-A and JR2-B are held together by aromatic ring stacking interactions between F233 of JR2-A and H229 of JR2-B. N230 of both protomers also contribute to the SI-2 formation (Fig. 2D and Fig. S4C). We noticed that only a few amino acids - (F454, Y455, Q461, Y466, and S569) from TM 2, TM 6, and EL2 – within each protomer contribute to form a small hydrophobic SI-3 that is buried within the transmembrane region (Fig. 2D and Fig. S4D). The TM2s from both protomers are curved inwards at the center like the surfaces of a double concave lens, and the cytosolic halves of TM2s move away from each other and the dimer interface (Fig. 1G and Fig. 2D). However, the local resolution of SI-3, primarily, cytosolic half of TM2 and the TM6 are poor. As a result, we could not model all the side chains in this region (Fig. 1D, Fig. S3C and S3D).

2.3. Comparison of hSIDT1 structure with its homologs

While we were compiling this communication, five research groups either published structures of SIDT1 and its homologs or deposited the unpublished structures in the PDB (RCSB PDB: 8JUL, 8JUN, 8HIP, 8HKE, 7Y63, 7Y68, and 7Y69) [33-36]. Here, we compare the available PDBs of ChUP family structures and highlight the flexibility within these proteins, especially in the LBD that is believed to bind to a variety of lipids and possess lipid hydrolysis activity (Fig. 3, Fig. 5, and Table S2). Because the TMD core is stabilized by four tiers of interactions, including a metal ion binding site, we used TMD core of these proteins to superpose chain A of each structure onto our hSIDT1 map (RCSB PDB: 8V38), and analyzed the Cα RMSDs of the superposition to study intra-chain flexibility (Fig. 3 and Table S2). Our structure superposition analysis suggests that LBD is the most dynamic region of the protein, as we see varying degrees of motion in the LBD in all the structures compared to ours. The LBD position in our structure is nearly identical to the E555Q-hSIDT1, a mutant that displays abrogated lipase activity (Fig. 3B- D). Overall, the LBD flexibility appears insensitive to pH changes (Fig. 3E-G) or RNA binding (Fig. 3C and 3D). Although the SIDT2 structure obtained at pH 5.0 in the presence of RNA did not yield a map where RNA could be modeled unambiguously, Zheng et al., demonstrate binding biochemically [36]. Even in SIDT2, the LBD dynamics remain unaffected by RNA binding (Fig. 3F and 3G). Based on the *C. elegans* SID1-RNA complex structure (8HKE), the interface of JR1 and JR2, mainly β5-β6 hairpin and β7'-β8' hairpin, seems to be the primary site of RNA interaction. We notice non-trivial movement in this region in our structure, compared to *C. elegans* SID1-RNA complex structure (Fig. 3D and Table S2). Interestingly, JR1 seems to be the second most dynamic region after LBD across different structures, with significant movement in the *C. elegans* SID1 structure (Fig. 3D). Between *C. elegans* SID1 and human SIDT1 structures, of all the helices, TM2, the primary helix at the dimer interface, is most dynamic (Fig. 3C and 3D).

Upon mapping the center of masses (COMs) of chain A and chain B of hSIDT1, we noticed that the whole chain COMs are ~35Å from each other, and the ECD COMs are ~28Å apart (Fig. 5A-C). To study the inter-chain flexibility amongst the structures, we superposed chain A of all the available structures and measured the distance of COMs of chain B from the superposed chain A center. The distance of whole chain COMs from each other remains similar, with only up to a ~3Å movement. However, the vertical movement of chain B in reference to chain A alters in a non-trivial fashion (Fig. 5C). For example, this difference is ~9.7Å between the hSIDT1 structures and the *C. elegans* SID1 structures (Fig. 5C). We notice a similar difference in the position of chain B ECD in reference to chain A ECD for the *C. elegans* structures. In conjunction, we also notice that the difference in the distance between COMs of chain A and chain B ECDs in the *C. elegans* structures compared to hSIDT1 structures is almost twice (~7Å) that of the difference in distance between full-length chain A and chain B COMs (~3Å). These observations suggest a difference in the directionality of motion between the TMDs and ECDs in one protomer in reference to the other protomer in a dimer. A noteworthy observation is the position of COMs of the ECD and full chains of the E555Q hSIDT1 mutant (8JUN) in comparison with wild-type hSIDT1 (Fig. 5B and 6C). The glutamic acid (E555) residue in TM4 lies at the interface of TMD core and LBD, facing the TM5 and the metal ion binding site. A mutation of E555Q resulted in a decreased lipase activity, suggesting that the LBD-TMD core interface plays a crucial role in either binding or hydrolysis of the lipid [34]. While the ECD distances remain unchanged, and the intrachain motion is limited to just TM5, chain B moves down in reference to chain A, suggesting a motion in the TMD-B region (Fig. 3A, 3B, and 5C). Perhaps synergistic, inter-chain TMD and intra-chain LBD dynamics are essential for dsRNA transport.

Discussion

siRNA-mediated RNA interference begins with the poorly understood mechanism of transport of small noncoding dsRNA into a cell. Once inside the cell, the dsRNA is processed in a manner that is conserved across different kingdoms of life. Briefly, the incoming dsRNA is trimmed to 21-23 bp short interfering RNA by the RNase III family of enzymes. The processed dsRNA is unwound by an RNA-induced silencing complex (RISC) that mediates the hybridization of the single-stranded antisense RNA with target mRNA and eventual degradation of the homologous mRNA [40, 41]. A large portion of the existing knowledge on the complicated specific cellular transport of small non-coding RNAs has been obtained by studying RNAi in *C. elegans* [2, 11, 12]. Human homologs of the ChUP family of proteins – hSIDT1 and hSIDT2 - are believed to be lipophilic dsRNA transporters [17, 18, 29]. However, the exact physiological role of SIDTs has been a longstanding debate, with various activities such as the transport of cations, cholesterol, DNA, ATP-dependent protein and peptides, and most recently, lipid hydrolysis being attributed to them in the literature [18, 24, 25, 29, 42-44]. The emerging consensus on the function is that SIDTs are cholesterol-dependent dsRNA transporters with a divalent metal ion dependent lipase activity within the transmembrane region [29, 33-36, 45].

While preparing this manuscript, four other research groups reported the structural and functional characterization of recombinant human SIDTs [33-36]. All four reports have non-trivial differences compared to the sample preparation protocol employed here (Table S1). Qian et al., and Sun et al., studies used large-scale transient transfections in HEK293F cells to produce full-length hSIDT2 and hSIDT1 [33, 34]. Both studies used a combination of maltose-neopentyl glycol (MNG) and cholesteryl hemisuccinate (CHS) to solubilize, purify the protein in glyco-diosgenin (GDN), and concentrate it to > 12 mg/ml for sample vitrification. Zheng et al., used baculoviral

transduction of insect cells to express human SIDTs with a portion of 40-50 amino acid long cytosolic loop (CL1) between TM1 and TM2 deleted [36]. They, too, used CHS in their purification buffers. Hirano et al., used baculoviral transduction of mammalian cells to express full-length hSIDT1 in the absence of CHS [35]. Interestingly, all the reports incorporated a FLAG tag for purification. All three groups maintained Vitrobot chamber at a temperature of 6-8°C for sample blotting, and we used a sample blotting temperature of 18°C. These differences in the purification and grid preparation conditions might contribute to the differences we see in our structure compared to the ChUP family member structures reported by other groups (Table S1).

ECD and the disordered cytosolic domain (CD) connecting TM1 and TM2 have been shown to bind to RNA in a pH-dependent manner, with more efficient binding at low pH [31-37, 46]. In hSIDT2, even though the cryo-EM sample at low pH contained RNA, it could not be observed in the cryo-EM map. However, in the presence of RNA, an unexplained density appeared towards the cytosolic side - perhaps a better resolved CD [33]. In hSIDT1, deletion of CD resulted in an expression construct with diminished RNA binding ability [35]. Mendez-Acevedo et al., demonstrated that SIDTs primarily transport cholesterol, and they transport RNA only upon being conjugated to cholesterol [29]. SIDTs are believed to have three CRAC motifs (L/V-$X_{(1-5)}$-Y/F/W-$X_{(1-5)}$-R/K) located between L155-K165 (β8 of JR1), L584-R593 (TM5), and L643-R652 (TM7). In the unpublished RNA-bound *C. elegans* SID1 structure, RNA has been modeled at the JR1-JR2 interface, closer to the first CRAC motif in JR1 (RCSB PDB: 8HKE). It is possible that cholesterol facilitates the binding of RNA to the ECD. In all the structures of ChUP family, lipid interactions have been observed at the LBD or LBD-TMD core interface, and the LBD is believed to possess lipase activity [33-36]. However, Hirano et al., suggest that hSIDT1 lipase activity has been observed in the presence of methyl β-cyclodextrin, emphasizing that removing cholesterol

enhances the lipid hydrolysis [35]. In our experimentally determined hSIDT1 structure, we observe that LBD is poorly resolved as we did not use cholesterol or its analogs in our sample preparation. Although we see density for ordered lipids/detergents in our structure, they are predominantly located at the extra-cytosolic side of the phenylalanine highway (Fig. 4D). In the previously reported hSIDT1 structures, LBD is better resolved either due to the presence of cholesterol or due to the E555Q mutant that reduces the lipase activity [33-36]. These observations, combined with our structure comparison analysis (Fig. 3), suggest that LBD dynamics facilitates the lipase activity and that cholesterol acts as an allosteric regulator of the activity by stabilizing these LBD dynamics.

Grishin and colleagues, using sensitive sequence similarity searches, classified a group of putative transmembrane receptors and transporters into a superfamily called CREST (alkaline **c**eramidase, PAQ **r**eceptor, P**e**r1, **S**ID-1, and **T**MEM8) [47]. CREST superfamily of membrane proteins was suggested to be comprised of metal-dependent lipid-modifying hydrolases that possess a conserved S-H-D-H-H metal ion coordination site formed by the juxtaposition of three TMs. These residues are conserved in the TMD core of ChUP family of proteins. As a result, all the previous ChUP family structure reports probed and demonstrated zinc-dependent lipid hydrolysis activity in SIDT-like proteins [33-36]. In our hSIDT1 structure, we find unexplained density in this site carved by S559, H563, D574, H791, and H795 (Fig. 2C). Due to a nominal resolution of ~3.4Å, and due to lack of evidence for zinc in our experiments, we left this site unmodeled. H795F mutation, like E555Q, in hSIDT1, reduces the lipid hydrolysis activity [34]. Also, the analogous mutations S559I and H791Y in *C. elegans* attenuate systemic RNAi [46]. These observations show that a metal ion is not only essential to maintain the TMD core integrity but is also crucial for lipase hydrolysis and RNA transport.

Interestingly, we also notice that the amino acids which upon being mutated in *C. elegans* abrogate the systemic RNAi, and are conserved in the ChUP family of proteins, all predominantly lie in the vicinity of the path traced by the cavity in our MOLE analysis (Fig. 4B and Fig. 5D) [35, 46, 48]. We analyzed these mutations in the context of hSIDT1 structure using FoldX to model the mutations. P109S mutation in the β5-β6 hairpin, at the JR1-JR2 interface crucial for ECD architecture, lies near the RNA binding site as per the RNA-bound *C. elegans* SID1 structure (RCSB PDB: 8HKE) (Fig. 5D). C479 at the TMD-ECD interface is a part of the C479-C565 disulfide essential for stabilizing the TMD core by stitching EL1 and EL2 together (Fig. 4F). C479Y mutation breaks this disulfide and, as a result, could destabilize the TMD-ECD interface (Fig. 2B). R593C mutation occurs in the hinge region at the LBD-TMD core interface (Fig. 5D). Arg to Cys substitution leads to a drastic charge change, potentially destabilizing the hinge region primarily stabilized by a series of salt bridges and electrostatic interactions. Four other conserved residues that abrogate systemic RNAi are G503, G508, S559, and H791 (Fig. 5D). All these residues are situated in the TMD core with H791 being a part of metal ion coordination. H791Y will destabilize the metal ion binding site that is suggested to be essential for lipase activity and protein stability. Although it does not coordinate the metal ion directly, S559 is a part of the binding site. An Ile at this position will affect the TM4-EL2 connection that hosts H563, another residue involved in metal ion coordination (Fig. 2C). Finally, G503R and G508R are drastic mutations that change both the charge and size of the side chain, affecting the stability of the core towards the cytosolic side (Fig. 5D).

The exposed face of ECD is covered with amino acids containing basic side chains, especially at the interface of JR1 and JR2 (Fig. 4A). In the RNA-bound *C. elegans* SID1 structure (RCSB PDB: 8HKE), the RNA has been modeled at the JR1-JR2 interface. MOLE analysis of the

hSIDT1 structure revealed the presence of two cavities in tandem, each of about 110-120 Å in length, which extend from the JR1-JR2 interface to the cytosolic opening of the protomer passing through the LBD-TMD core interface (Fig. 4B). It has been suggested, based on the recent lipase activity observations of SIDT1-like proteins, that RNA transport is dependent on lipase activity. Finally, our structure and its comparative analysis with other reported structures reveal that the LBD domain is the most flexible region of the protein, followed by JR1 and TM2 (Fig. 3 and Table S2). Based on these observations, we hypothesize that the RNA binding at the JR1-JR2 interface induces a rearrangement of JR1 and JR2 which allows for the formation of an opening that can feed RNA to the TMD. Then, the metal ion dependent lipase activity in the TMD region disturbs the lipid bilayer around the protein and allows for the LBD and TM2 to rearrange, enabling the formation of a cavity large enough to allow the passage of RNA to cytosol.

Materials and Methods

Cell culturing and protein expression

The codon optimized gene encoding full length human systemic RNAi-defective transmembrane family member 1 (hSIDT1) was synthesized by GenScript and was then cloned into the pEG BacMam expression vector to be expressed *via* baculoviral transduction in HEK293S GnTI$^-$ cells as a fusion protein containing a *C*-terminal GFP-Strep-tag-II for large scale protein expression [49]. For small scale detergent screening, adherent HEK293S GnTI$^-$ cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) with 10% fetal bovine serum (FBS) at 37°C. Before transfection, the cells were washed with 1X PBS and were supplied with fresh pre-warmed DMEM containing 10% FBS. Cells were transfected at 80% confluency. Approximately 1x10^6 cells were

transfected with 1μg DNA using TurboFect (ThermoFisher Scientific), in serum-free DMEM, as per the manufacturer's protocol. The cells expressing hSIDT1 were grown at 37°C and 5% CO_2 for 8-10 hr, following which the culture media was replaced with fresh pre-warmed DMEM containing 10% FBS and 10 mM of sodium butyrate. Subsequently, the cells were grown at 32°C and 5% CO_2 for an additional 24-36 hr, before being harvested.

Baculovirus preparation was done in Sf9 cells grown in Sf-900 III serum-free medium. Briefly, DH10Bac cells (ThermoFisher Scientific) were transformed with hSIDT1 expression vector, and $lacZ^-$ colonies were selected for bacmid DNA isolation. $1x10^6$ Sf9 cells were transfected with 1 ug of DNA using Cellfectin II reagent as per manufacturer's protocol. Transfected cells were grown at 27°C for 96 hr. The supernatant from the cell culture was harvested, filtered through 0.2 μm filter, and used as P1 virus. 100 μL of P1 virus was added to 1L of Sf9 cells at a cell density of $1x10^6$/mL in serum-free Sf-900 III media. The transduced cells were grown at 96 hr at 27°C while shaking at 120 rpm, spun down at 4000 xg for 20 min, and the supernatant media was filtered through 0.2 μm filter and used as P2 baculovirus.

For large-scale expression of hSIDT1, baculoviral transduction of HEK293S $GnTI^-$ cells was performed in FreeStyle 293 expression media containing 2% FBS. Mammalian cells at a density of $2.5x10^6$ cells/mL were transduced using P2 virus at a multiplicity of infection of 1.5-2, and was incubated on an orbital shaker at 37°C and 5% CO_2 for 8-10 hr. Following this, the cells were supplemented with 10 mM sodium butyrate and were incubated in a shaker at 32°C and 5% CO_2 for additional 38-40 hr. The harvested cell pellet was stored at -80°C until further use.

Purification of hSIDT1

The sonicated lysate of HEK293S $GnTI^-$ cells expressing *C*-terminal GFP fusion of hSIDT1-FL was subjected to ultra-centrifugation at 185,000 xg for 1 hr at 4°C to harvest membranes.

Membranes were resuspended using a Dounce homogenizer in 25 mM Tris-HCl, pH 7.5, 200 mM NaCl, 1 mM PMSF, 0.8 μM aprotinin, 2 μg/mL leupeptin, and 2 μM pepstatin A. To this membrane suspension, equal volume of 2% digitonin solution prepared in membrane suspension buffer was added. The membrane suspension was solubilized at 4ºC for 90 min. The solubilized membrane suspension was centrifuged at 185,000 xg for 1 hr at 4ºC. 2 ml of Strep-Tactin affinity resin was packed into a column, per every liter of cell culture, and the solubilized supernatant was passed through it at a flow rate of ~0.3-0.5 ml/min. Affinity resin saturated with hSIDT1-FL was washed with 5-6 column volumes of 0.5% digitonin, 25 mM Tris-HCl, pH 7.5, and 200 mM NaCl. The bound protein was eluted using 5 mM D-desthiobiotin prepared in the wash buffer. The homogeneity of purified hSIDT1 was confirmed by performing fluorescence-detection size-exclusion chromatography (FSEC) [50]. We confirmed the purity of hSIDT1 by SDS-PAGE and peptide mass fingerprinting. The elution fractions containing homogeneous and pure protein were pooled and concentrated to ~2 mg/mL. hSIDT1 was further purified by size-exclusion chromatography (SEC) using Superdex S200 pre-equilibrated with 0.5% digitonin, 25 mM Tris-HCl, pH 7.5, and 200 mM NaCl. Fractions corresponding to the dimer hSIDT1 peak were analyzed for their purity and homogeneity on SDS-PAGE and FSEC, and were concentrated to 1 mg/ml. The concentrated protein was incubated with 100 uM fluorinated octyl-maltoside (FOM) for 20-30 min on ice, centrifuged at 18000 xg for 15 minutes, and the supernatant was used for preparing cryo-EM grids.

Detergent screening by fluorescence-detection size-exclusion chromatography

Briefly, 100,000 transiently transfected cells expressing hSIDT1-GFP were solubilized in 100 μl of 1% of detergent prepared in 25 mM Tris-HCl, pH 7.5, 200 mM NaCl, 1 mM PMSF, 0.8 μM

aprotinin, 2 μg/mL leupeptin, and 2 μM pepstatin A. The solubilized lysate was centrifuged at 185,000 xg for 1 hr at 4ºC and the supernatant was analyzed by FSEC on Superose 6 Increase 10/300 GL column pre-equilibrated with 0.15 mM LMNG, 25 mM Tris-HCl, pH 7.5, and 200 mM NaCl. To evaluate the quality of the purified protein during the protein purification process, and to estimate the molecular weight of the dimer, ~10 μL of elution fractions were analyzed in a similar fashion (Fig. S1).

Cryo-EM sample preparation and data collection

UltrAuFoil holey-gold 300 mesh 1.2/1.3 μm size/hole space grids (UltrAuFoil, Quantifoil) were glow discharged, using a PELCO glow discharger, for 1 min at 15 mA current and 0.26 mBar air pressure before sample vitrification. 2.5 μL of hSIDT1 at 1 mg/ml was applied to the glow-discharged grids, and subsequently blotted for 2.5 s at 18°C and 100% humidity using a Vitrobot (mark IV, ThermoFisher Scientific). The grids were plunge-frozen in liquid ethane, without any wait time. Movies were recorded on a Gatan K3 direct electron detector in super-resolution counting mode with a binned pixel size of 0.85 Å per pixel using Serial EM on a FEI Titan Krios G4i transmission electron microscope operating at an electron beam energy of 300 KeV with a Gatan Image Filter slit width set to 20 eV. Exposures of ~2s were dose fractionated into 50 frames, attaining a total dose of ~50 e^{-} Å^{-2}, and the defocus values were varied from −1.2 to −2.5 μm.

Cryo-EM data processing

Image processing was done entirely using CryoSPARC (v 4.2.1) [51]. Briefly, motion correction was carried out on the raw movies, and subsequently, the contrast transfer function (CTF) was estimated using patch-CTF correction on dose-weighted motion-corrected averages within

CryoSPARC (Fig. S2). A total of 20,764 micrographs were collected, of which 15,391 were selected for further processing. Elliptical gaussian blob-picker was used to pick 3,204,173 particles. Particles were extracted initially using a box size of 336 pixels. A subset of the particles picked by the blob picker was used to generate a low-resolution *ab initio* reconstruction, which was later used as a reference for heterogenous refinement. The particles were subjected to multiple rounds of reference-free 2D classification and heterogenous refinement to get a cleaned subset of particles which resulted in classes with recognizable features. The cleaned particles were re-extracted with a box size of 360 pixels, resulting in the particle stack of 302,224 particles that were further processed stringently by heterogeneous refinement and 2D classification. Finally, only the particles belonging to micrographs with CTF fit <4 Å were selected for subsequent refinements. A final subset of 122,683 particles was used for NU-refinement, yielding a 3.5 Å map (Fourier shell coefficient (FSC) = 0.143 criterion) [52]. Local refinement with a focused mask on the entire protein was performed to improve local density, resulting a final map of 3.4 Å. Both sharpened and unsharpened maps were used to finalize the fit of the side chains during model building and refinement (Fig S3D). The final 3D model was validated using the sharpened map for deposition, and local resolutions were estimated on the sharpened map using the RESMAP software [53].

Model building, refinement, and structure analysis

ModelAngelo was used to build hSIDT1 ECD into the final map [54]. AlphaFold hSIDT1 model (UniProt ID: Q9NXL6) was used to build the TMD core [55, 56]. LBD domain was built manually, and the entire structure was analyzed for the Ramachandran and geometry outliers in Coot [57]. The final model refinement was performed in Phenix [58]. ChimeraX was used for making the figures [59]. RNA transport channel prediction on hSIDT1 was performed using MOLEonline

server [60]. To analyze the RNAi attenuating *C. elegans* mutations in hSIDT1 context, we used FoldX to generate mutant models of hSIDT1 [61].

Acknowledgements

We thank the University of Michigan (UM) cryo-EM facility staff for assistance with cryo-EM data collection, and UM BSI and UM LSI for generously supporting the UM cryo-EM facility. All members of the Mosalaganti lab and the Baldridge lab, and Dr. Jonathan Coleman, Dr. Farzad Jalali-Yazdi, and Dr. Prashant Rao are acknowledged for their helpful comments on our work. The research presented here was funded by the Klatskin-Sutker Discovery Fund award to VN and the start-up funding provided to SM by the University of Michigan. SM is also a recipient of the NIH Director's New Innovator Award (DP2) (1DP2GM150019-01).

Author contributions

VN designed the project. VN and JKR performed the biochemistry. VN collected the cryo-EM data, analyzed the structure, and wrote the complete first draft of the manuscript. VN and AK built the model and deposited the map and coordinates. All the authors processed the cryo-EM data, prepared the figures, and edited the final draft of the manuscript. AK and JKR contributed equally to this work and have been listed alphabetically.

Conflict of interest

The authors declare no conflict of interest.

Data availability

The correspondence for material and data presented in this manuscript should be addressed to navratna@umich.edu or mosalaga@umich.edu. The cryo-EM maps and 3D coordinates of human SIDT1, along with the half-maps and the masks used for refinement, have been deposited under the accession codes EMD-42943 and PDB-8V38 in the Electron Microscopy Data Bank (EMDB) and Protein Data Bank (PDB).

References:

1. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature. 1998;391(6669):806-11. doi: 10.1038/35888. PubMed PMID: 9486653.

2. Tabara H, Grishok A, Mello CC. RNAi in C. elegans: soaking in the genome sequence. Science. 1998;282(5388):430-1. doi: 10.1126/science.282.5388.430. PubMed PMID: 9841401.

3. Timmons L, Fire A. Specific interference by ingested dsRNA. Nature. 1998;395(6705):854. doi: 10.1038/27579. PubMed PMID: 9804418.

4. Lucas T, Bonauer A, Dimmeler S. RNA Therapeutics in Cardiovascular Disease. Circ Res. 2018;123(2):205-20. doi: 10.1161/CIRCRESAHA.117.311311. PubMed PMID: 29976688.

5. Chernikov IV, Vlassov VV, Chernolovskaya EL. Current Development of siRNA Bioconjugates: From Research to the Clinic. Front Pharmacol. 2019;10:444. Epub 20190426. doi: 10.3389/fphar.2019.00444. PubMed PMID: 31105570; PubMed Central PMCID: PMCPMC6498891.

6. Pottash AE, Kuffner C, Noonan-Shueh M, Jay SM. Protein-based vehicles for biomimetic RNAi delivery. J Biol Eng. 2019;13:19. Epub 20190226. doi: 10.1186/s13036-018-0130-7. PubMed PMID: 30891095; PubMed Central PMCID: PMCPMC6390323.

7. Rai MF, Pan H, Yan H, Sandell LJ, Pham CTN, Wickline SA. Applications of RNA interference in the treatment of arthritis. Transl Res. 2019;214:1-16. Epub 20190710. doi: 10.1016/j.trsl.2019.07.002. PubMed PMID: 31351032; PubMed Central PMCID: PMCPMC6848781.

8. Chen X, Rechavi O. Plant and animal small RNA communications between cells and organisms. Nat Rev Mol Cell Biol. 2022;23(3):185-203. Epub 20211027. doi: 10.1038/s41580-021-00425-y. PubMed PMID: 34707241; PubMed Central PMCID: PMCPMC9208737.

9. Cappelle K, de Oliveira CF, Van Eynde B, Christiaens O, Smagghe G. The involvement of clathrin-mediated endocytosis and two Sid-1-like transmembrane proteins in double-stranded

RNA uptake in the Colorado potato beetle midgut. Insect Mol Biol. 2016;25(3):315-23. Epub 20160309. doi: 10.1111/imb.12222. PubMed PMID: 26959524.

10. Jose AM, Hunter CP. Transport of sequence-specific RNA interference information between cells. Annu Rev Genet. 2007;41:305-30. doi: 10.1146/annurev.genet.41.110306.130216. PubMed PMID: 17645412; PubMed Central PMCID: PMCPMC7377521.

11. Winston WM, Molodowitch C, Hunter CP. Systemic RNAi in C. elegans requires the putative transmembrane protein SID-1. Science. 2002;295(5564):2456-9. Epub 20020207. doi: 10.1126/science.1068836. PubMed PMID: 11834782.

12. Hunter CP, Winston WM, Molodowitch C, Feinberg EH, Shih J, Sutherlin M, et al. Systemic RNAi in Caenorhabditis elegans. Cold Spring Harb Symp Quant Biol. 2006;71:95-100. doi: 10.1101/sqb.2006.71.060. PubMed PMID: 17381285.

13. Chen Q, Zhang F, Dong L, Wu H, Xu J, Li H, et al. SIDT1-dependent absorption in the stomach mediates host uptake of dietary and orally administered microRNAs. Cell Res. 2021;31(3):247-58. Epub 20200817. doi: 10.1038/s41422-020-0389-3. PubMed PMID: 32801357; PubMed Central PMCID: PMCPMC8026584.

14. Chen X, Gu X, Zhang H. Sidt2 regulates hepatocellular lipid metabolism through autophagy. J Lipid Res. 2018;59(3):404-15. Epub 20180123. doi: 10.1194/jlr.M073817. PubMed PMID: 29363559; PubMed Central PMCID: PMCPMC5832936.

15. Gao J, Gu X, Mahuran DJ, Wang Z, Zhang H. Impaired glucose tolerance in a mouse model of sidt2 deficiency. PLoS One. 2013;8(6):e66139. Epub 20130611. doi: 10.1371/journal.pone.0066139. PubMed PMID: 23776622; PubMed Central PMCID: PMCPMC3679015.

16. Gao J, Yu C, Xiong Q, Zhang Y, Wang L. Lysosomal integral membrane protein Sidt2 plays a vital role in insulin secretion. Int J Clin Exp Pathol. 2015;8(12):15622-31. Epub 20151201. PubMed PMID: 26884831; PubMed Central PMCID: PMCPMC4730044.

17. Nguyen TA, Smith BRC, Elgass KD, Creed SJ, Cheung S, Tate MD, et al. SIDT1 Localizes to Endolysosomes and Mediates Double-Stranded RNA Transport into the Cytoplasm. J Immunol. 2019;202(12):3483-92. Epub 20190506. doi: 10.4049/jimmunol.1801369. PubMed PMID: 31061008.

18. Nguyen TA, Smith BRC, Tate MD, Belz GT, Barrios MH, Elgass KD, et al. SIDT2 Transports Extracellular dsRNA into the Cytoplasm for Innate Immune Recognition. Immunity. 2017;47(3):498-509 e6. Epub 20170912. doi: 10.1016/j.immuni.2017.08.007. PubMed PMID: 28916264; PubMed Central PMCID: PMCPMC5679266.

19. Xiong QY, Xiong CQ, Wang LZ, Gao JL. Effect of sidt2 Gene on Cell Insulin Resistance and Its Molecular Mechanism. J Diabetes Res. 2020;2020:4217607. Epub 20200911. doi:

10.1155/2020/4217607. PubMed PMID: 32964053; PubMed Central PMCID: PMCPMC7502120.

20. Elhassan MO, Christie J, Duxbury MS. Homo sapiens systemic RNA interference-defective-1 transmembrane family member 1 (SIDT1) protein mediates contact-dependent small RNA transfer and microRNA-21-driven chemoresistance. J Biol Chem. 2012;287(8):5267-77. Epub 20111215. doi: 10.1074/jbc.M111.318865. PubMed PMID: 22174421; PubMed Central PMCID: PMCPMC3285307.

21. Takahashi M, Seki M, Nashimoto M, Kabuta T. Perturbing the Normal Level of SIDT1 Suppresses the Naked ASO Effect. J Nucleic Acids. 2021;2021:2458470. Epub 20211116. doi: 10.1155/2021/2458470. PubMed PMID: 34824869; PubMed Central PMCID: PMCPMC8610720.

22. Morell M, Varela N, Castillejo-Lopez C, Coppard C, Luque MJ, Wu YY, et al. SIDT1 plays a key role in type I IFN responses to nucleic acids in plasmacytoid dendritic cells and mediates the pathogenesis of an imiquimod-induced psoriasis model. EBioMedicine. 2022;76:103808. Epub 20220119. doi: 10.1016/j.ebiom.2021.103808. PubMed PMID: 35065421; PubMed Central PMCID: PMCPMC8784643.

23. Duxbury MS, Ashley SW, Whang EE. RNA interference: a mammalian SID-1 homologue enhances siRNA uptake and gene silencing efficacy in human cells. Biochem Biophys Res Commun. 2005;331(2):459-63. doi: 10.1016/j.bbrc.2005.03.199. PubMed PMID: 15850781.

24. Aizawa S, Fujiwara Y, Contu VR, Hase K, Takahashi M, Kikuchi H, et al. Lysosomal putative RNA transporter SIDT2 mediates direct uptake of RNA by lysosomes. Autophagy. 2016;12(3):565-78. doi: 10.1080/15548627.2016.1145325. PubMed PMID: 27046251; PubMed Central PMCID: PMCPMC4836006.

25. Beck A, Fecher-Trost C, Wolske K, Philipp SE, Flockerzi V, Wissenbach U. Identification of Sidt2 as a lysosomal cation-conducting protein. FEBS Lett. 2017;591(1):76-87. Epub 20170102. doi: 10.1002/1873-3468.12528. PubMed PMID: 27987306.

26. Nguyen TA, Bieging-Rolett KT, Putoczki TL, Wicks IP, Attardi LD, Pang KC. SIDT2 RNA Transporter Promotes Lung and Gastrointestinal Tumor Development. iScience. 2019;20:14-24. Epub 20190910. doi: 10.1016/j.isci.2019.09.009. PubMed PMID: 31546103; PubMed Central PMCID: PMCPMC6817685.

27. Saier MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, Iddamsetty V, et al. The Transporter Classification Database (TCDB): 2021 update. Nucleic Acids Res. 2021;49(D1):D461-D7. doi: 10.1093/nar/gkaa1004. PubMed PMID: 33170213; PubMed Central PMCID: PMCPMC7778945.

28. Fantini J, Barrantes FJ. How cholesterol interacts with membrane proteins: an exploration of cholesterol-binding sites including CRAC, CARC, and tilted domains. Front Physiol.

2013;4:31. Epub 20130228. doi: 10.3389/fphys.2013.00031. PubMed PMID: 23450735; PubMed Central PMCID: PMCPMC3584320.

29. Mendez-Acevedo KM, Valdes VJ, Asanov A, Vaca L. A novel family of mammalian transmembrane proteins involved in cholesterol transport. Sci Rep. 2017;7(1):7450. Epub 20170807. doi: 10.1038/s41598-017-07077-z. PubMed PMID: 28785058; PubMed Central PMCID: PMCPMC5547113.

30. Sampieri A, Asanov A, Mendez-Acevedo KM, Vaca L. SIDT2 Associates with Apolipoprotein A1 (ApoA1) and Facilitates ApoA1 Secretion in Hepatocytes. Cells. 2023;12(19). Epub 20230926. doi: 10.3390/cells12192353. PubMed PMID: 37830567; PubMed Central PMCID: PMCPMC10571540.

31. Pratt AJ, Rambo RP, Lau PW, MacRae IJ. Preparation and characterization of the extracellular domain of human Sid-1. PLoS One. 2012;7(4):e33607. Epub 20120411. doi: 10.1371/journal.pone.0033607. PubMed PMID: 22509261; PubMed Central PMCID: PMCPMC3324469.

32. Li W, Koutmou KS, Leahy DJ, Li M. Systemic RNA Interference Deficiency-1 (SID-1) Extracellular Domain Selectively Binds Long Double-stranded RNA and Is Required for RNA Transport by SID-1. J Biol Chem. 2015;290(31):18904-13. Epub 20150611. doi: 10.1074/jbc.M115.658864. PubMed PMID: 26067272; PubMed Central PMCID: PMCPMC4521010.

33. Qian D, Cong Y, Wang R, Chen Q, Yan C, Gong D. Structural insight into the human SID1 transmembrane family member 2 reveals its lipid hydrolytic activity. Nat Commun. 2023;14(1):3568. Epub 20230615. doi: 10.1038/s41467-023-39335-2. PubMed PMID: 37322007; PubMed Central PMCID: PMCPMC10272179.

34. Sun CR, Xu D, Yang F, Hou Z, Luo Y, Zhang CY, et al. Human SIDT1 mediates dsRNA uptake via its phospholipase activity. Cell Res. 2023. Epub 20231106. doi: 10.1038/s41422-023-00889-x. PubMed PMID: 37932444.

35. Yoshinori Hirano UO, Ryota Sato, Kensuke Miyake, Toshiyuki Shimizu, . Cryo-EM analysis reveals human SID1 transmembrane family member 1 dynamics underlying lipid hydrolytic activity. Research Square https://doi.org/10.21203/rs.3.rs-3267442/v12023.

36. Zheng L, Yang T, Guo H, Qi C, Lu Y, Xiao H, et al. Cryo-EM structures of human SID-1 transmembrane family proteins and implications for their low-pH-dependent RNA transport activity. Cell Res. 2023. Epub 20231106. doi: 10.1038/s41422-023-00893-1. PubMed PMID: 37932445.

37. Hase K, Contu VR, Kabuta C, Sakai R, Takahashi M, Kataoka N, et al. Cytosolic domain of SIDT2 carries an arginine-rich motif that binds to RNA/DNA and is important for the direct transport of nucleic acids into lysosomes. Autophagy. 2020;16(11):1974-88. Epub 20200116. doi:

10.1080/15548627.2020.1712109. PubMed PMID: 31944164; PubMed Central PMCID: PMCPMC7595612.

38. Krupovic M, Makarova KS, Koonin EV. Cellular homologs of the double jelly-roll major capsid proteins clarify the origins of an ancient virus kingdom. Proc Natl Acad Sci U S A. 2022;119(5). doi: 10.1073/pnas.2120620119. PubMed PMID: 35078938; PubMed Central PMCID: PMCPMC8812541.

39. Farhat D, Rezaei F, Ristovski M, Yang Y, Stancescu A, Dzimkova L, et al. Structural Analysis of Cholesterol Binding and Sterol Selectivity by ABCG5/G8. J Mol Biol. 2022;434(20):167795. Epub 20220818. doi: 10.1016/j.jmb.2022.167795. PubMed PMID: 35988751.

40. Hannon GJ. RNA interference. Nature. 2002;418(6894):244-51. doi: 10.1038/418244a. PubMed PMID: 12110901.

41. Mello CC, Conte D, Jr. Revealing the world of RNA interference. Nature. 2004;431(7006):338-42. doi: 10.1038/nature02872. PubMed PMID: 15372040.

42. Aizawa S, Contu VR, Fujiwara Y, Hase K, Kikuchi H, Kabuta C, et al. Lysosomal membrane protein SIDT2 mediates the direct uptake of DNA by lysosomes. Autophagy. 2017;13(1):218-22. Epub 20161115. doi: 10.1080/15548627.2016.1248019. PubMed PMID: 27846365; PubMed Central PMCID: PMCPMC5245770.

43. Fujiwara Y, Kabuta C, Sano T, Murayama S, Saito Y, Kabuta T. Pathology-associated change in levels and localization of SIDT2 in postmortem brains of Parkinson's disease and dementia with Lewy bodies patients. Neurochem Int. 2022;152:105243. Epub 20211118. doi: 10.1016/j.neuint.2021.105243. PubMed PMID: 34800582.

44. Chang G, Yang R, Cao Y, Nie A, Gu X, Zhang H. SIDT2 is involved in the NAADP-mediated release of calcium from insulin secretory granules. J Mol Endocrinol. 2016;56(3):249-59. Epub 20160107. doi: 10.1530/JME-15-0227. PubMed PMID: 26744456.

45. Feinberg EH, Hunter CP. Transport of dsRNA into cells by the transmembrane protein SID-1. Science. 2003;301(5639):1545-7. doi: 10.1126/science.1087117. PubMed PMID: 12970568.

46. Whangbo JS, Weisman AS, Chae J, Hunter CP. SID-1 Domains Important for dsRNA Import in Caenorhabditis elegans. G3 (Bethesda). 2017;7(12):3887-99. Epub 20171204. doi: 10.1534/g3.117.300308. PubMed PMID: 29025917; PubMed Central PMCID: PMCPMC5714486.

47. Pei J, Millay DP, Olson EN, Grishin NV. CREST--a large and diverse superfamily of putative transmembrane hydrolases. Biol Direct. 2011;6:37. Epub 20110706. doi: 10.1186/1745-6150-6-37. PubMed PMID: 21733186; PubMed Central PMCID: PMCPMC3146951.

48. Wolfrum C, Shi S, Jayaprakash KN, Jayaraman M, Wang G, Pandey RK, et al. Mechanisms and optimization of in vivo delivery of lipophilic siRNAs. Nat Biotechnol. 2007;25(10):1149-57. Epub 20070916. doi: 10.1038/nbt1339. PubMed PMID: 17873866.

49. Goehring A, Lee CH, Wang KH, Michel JC, Claxton DP, Baconguis I, et al. Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. Nat Protoc. 2014;9(11):2574-85. Epub 2014/10/10. doi: 10.1038/nprot.2014.173. PubMed PMID: 25299155; PubMed Central PMCID: PMCPMC4291175.

50. Kawate T, Gouaux E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. Structure. 2006;14(4):673-81. Epub 2006/04/18. doi: 10.1016/j.str.2006.01.013. PubMed PMID: 16615909.

51. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods. 2017;14(3):290-6. Epub 20170206. doi: 10.1038/nmeth.4169. PubMed PMID: 28165473.

52. Punjani A, Zhang H, Fleet DJ. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. Nat Methods. 2020;17(12):1214-21. Epub 20201130. doi: 10.1038/s41592-020-00990-8. PubMed PMID: 33257830.

53. Kucukelbir A, Sigworth FJ, Tagare HD. Quantifying the local resolution of cryo-EM density maps. Nat Methods. 2014;11(1):63-5. Epub 20131110. doi: 10.1038/nmeth.2727. PubMed PMID: 24213166; PubMed Central PMCID: PMCPMC3903095.

54. Jamali K, Kall L, Zhang R, Brown A, Kimanius D, Scheres SHW. Automated model building and protein identification in cryo-EM maps. bioRxiv. 2023. Epub 20230516. doi: 10.1101/2023.05.16.541002. PubMed PMID: 37292681; PubMed Central PMCID: PMCPMC10245678.

55. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9. Epub 20210715. doi: 10.1038/s41586-021-03819-2. PubMed PMID: 34265844; PubMed Central PMCID: PMCPMC8371605.

56. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022;50(D1):D439-D44. doi: 10.1093/nar/gkab1061. PubMed PMID: 34791371; PubMed Central PMCID: PMCPMC8728224.

57. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 4):486-501. Epub 20100324. doi: 10.1107/S0907444910007493. PubMed PMID: 20383002; PubMed Central PMCID: PMCPMC2852313.

58. Liebschner D, Afonine PV, Baker ML, Bunkoczi G, Chen VB, Croll TI, et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Struct Biol. 2019;75(Pt 10):861-77. Epub 20191002. doi: 10.1107/S2059798319011471. PubMed PMID: 31588918; PubMed Central PMCID: PMCPMC6778852.

59. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, et al. UCSF ChimeraX: Tools for Structure Building and Analysis. Protein Sci. 2023:e4792. Epub 20230929. doi: 10.1002/pro.4792. PubMed PMID: 37774136.

60. Pravda L, Sehnal D, Tousek D, Navratilova V, Bazgier V, Berka K, et al. MOLEonline: a web-based tool for analyzing channels, tunnels and pores (2018 update). Nucleic Acids Res. 2018;46(W1):W368-W73. doi: 10.1093/nar/gky309. PubMed PMID: 29718451; PubMed Central PMCID: PMCPMC6030847.

61. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. Nucleic Acids Res. 2005;33(Web Server issue):W382-8. doi: 10.1093/nar/gki387. PubMed PMID: 15980494; PubMed Central PMCID: PMCPMC1160148.

62. Kremer JR, Mastronarde DN, McIntosh JR. Computer visualization of three-dimensional image data using IMOD. J Struct Biol. 1996;116(1):71-6. doi: 10.1006/jsbi.1996.0013. PubMed PMID: 8742726.

Figures

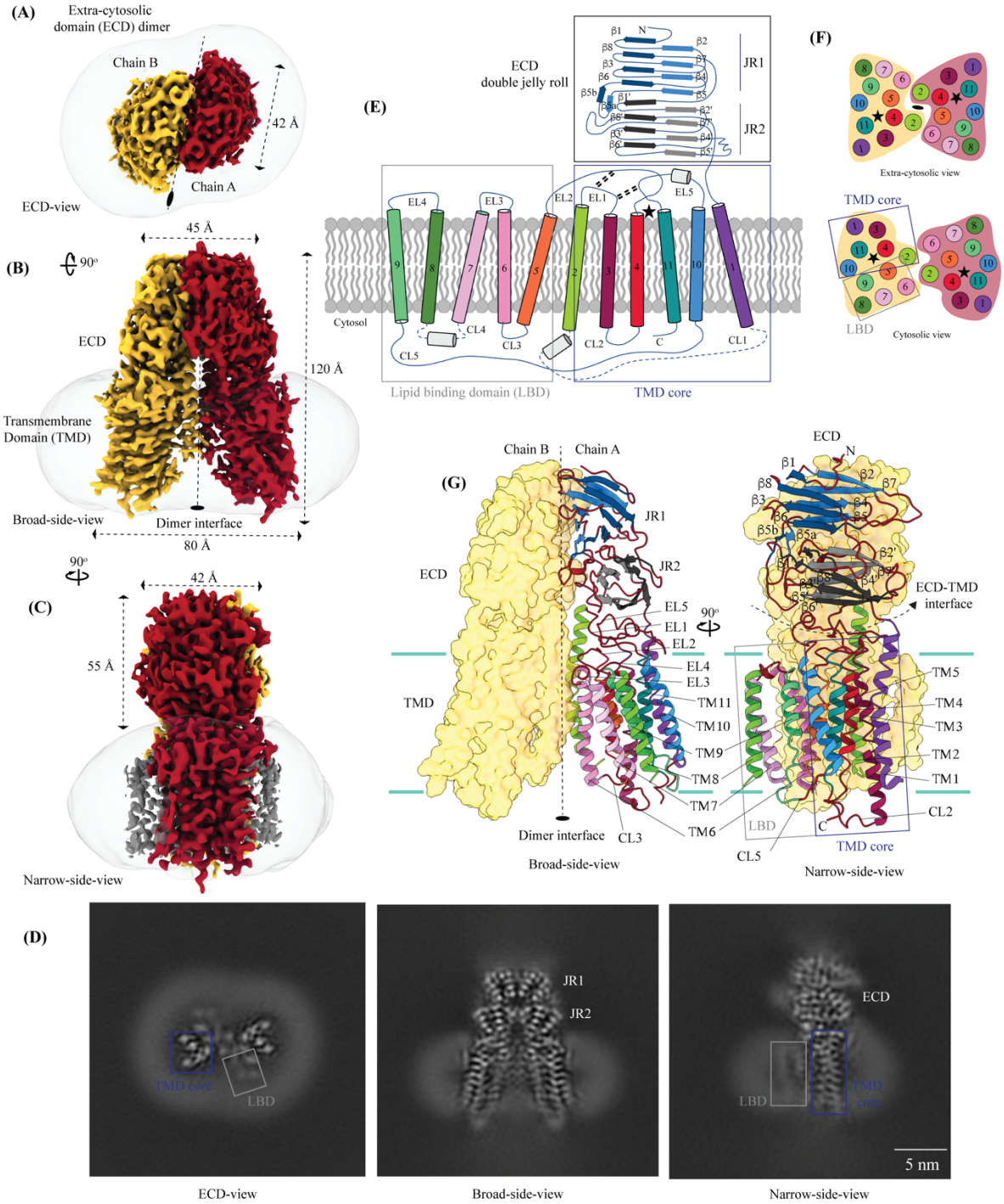

Figure 1: Architecture of hSIDT1

Panels **(A)**, **(B),** and **(C)** show the extra-cytosolic domain (ECD) view, broad-side-view, and narrow-side-view respectively, of the ~3.4Å cryo-EM map of hSIDT1 dimer. Chain A and chain

B are displayed in red and yellow, respectively, at level 0.14 in ChimeraX [59]. The LBD density is highlighted as a gray mesh in panel (c). Micelle is shown in transparent gray. **(D)** Cross-sections of the final 3D map from CryoSPARC, in three views, displayed using IMOD, indicating the poor resolution of the lipid binding domain (LBD, gray box) compared to the transmembrane domain core (TMD core, blue box) or ECD [51, 62]. **(E)** 2D topology of the ChUP family of proteins. Transmembrane helices (TMs) that form the LBD are highlighted by a gray box, the TMs that form TMD core by a blue box, and the ECD by a black box. The ECD forms a double jelly roll where each jelly roll (JR1 and JR2) is a β-sandwich made of two 4-stranded anti-parallel β-sheets. The cytosolic loops (CLs) and the extra-cytosolic loops (ELs) are numbered based on the primary structure. **(F)** The TMD dimer is depicted as cartoon, showing the TM arrangement as seen from extra-cytosolic space and the cytosol. The star indicates the metal ion binding site. **(G)** Two side-views of hSIDT1 highlight the dimer interface and ECD-TMD interface. Chain B is displayed as a surface, and chain A is shown as a cartoon with β-strands and TMs colored as they are in the topology diagram. The two-fold rotation axis is displayed as a dashed line with an ellipsoid.

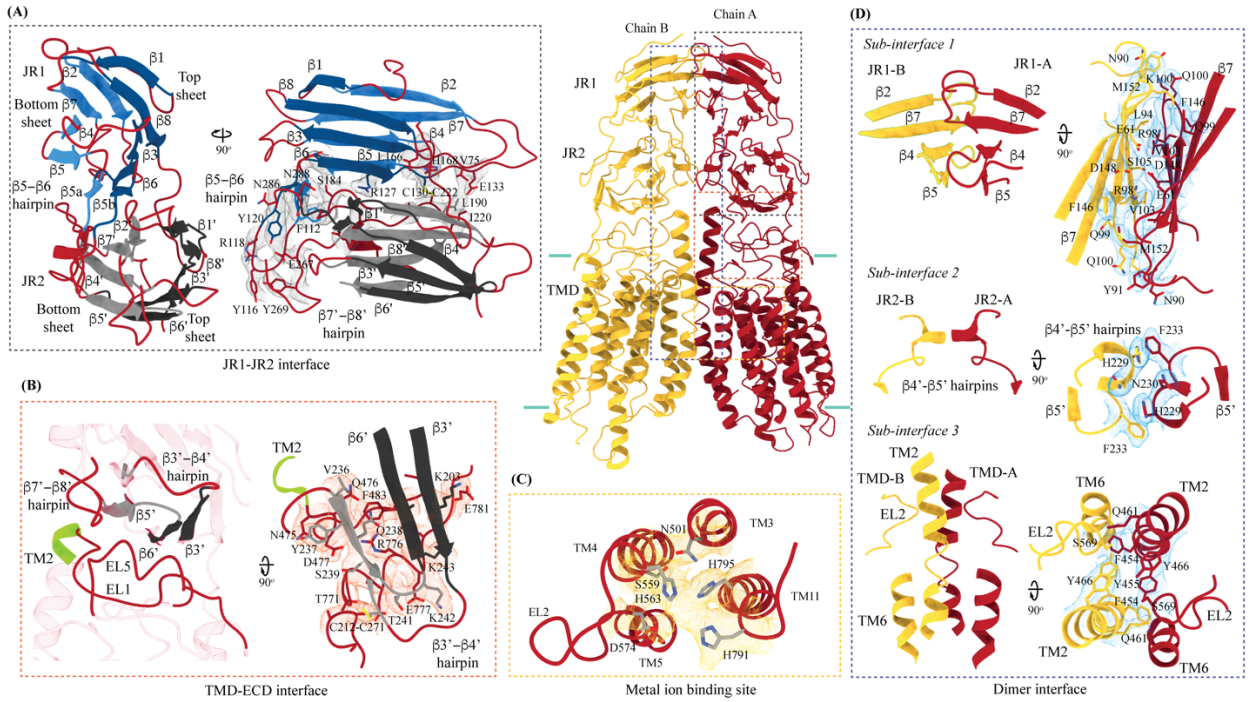

Figure 2: Critical inter-chain and intra-chain interactions of hSIDT1

Chain A and Chain B are displayed in red and yellow, respectively, at the center. The lipid bilayer is highlighted in all the figures as cyan bars. The regions highlighted in the panels A-D have been indicated on the overall structure with dashed boxes, gray for the JR1-JR2 interface, yellow for the metal ion binding site, orange for the ECD-TMD interface, and blue for the dimer interface. The densities for interactions have been displayed in colors that match their inset highlights (display level 0.25 for the ECD region and 0.1 for the TMD region in ChimeraX). **(A)** Double jelly roll arrangement of the ECD and the interactions that stabilize JR1 atop JR2. **(B)** Interaction of JR2 with the ECD face of the TMD core allows the ECD to seclude the TMD core from extra-cytosolic space. **(C)** Ensconced within the TMD core is the evolutionarily conserved metal ion binding site, which is highlighted in yellow. Side chains of amino acids that carve the metal coordination site have been highlighted. **(D)** Three sub-interfaces of the dimer interface, each made by JR1, JR2, and TMD, respectively, have been highlighted in the blue inset.

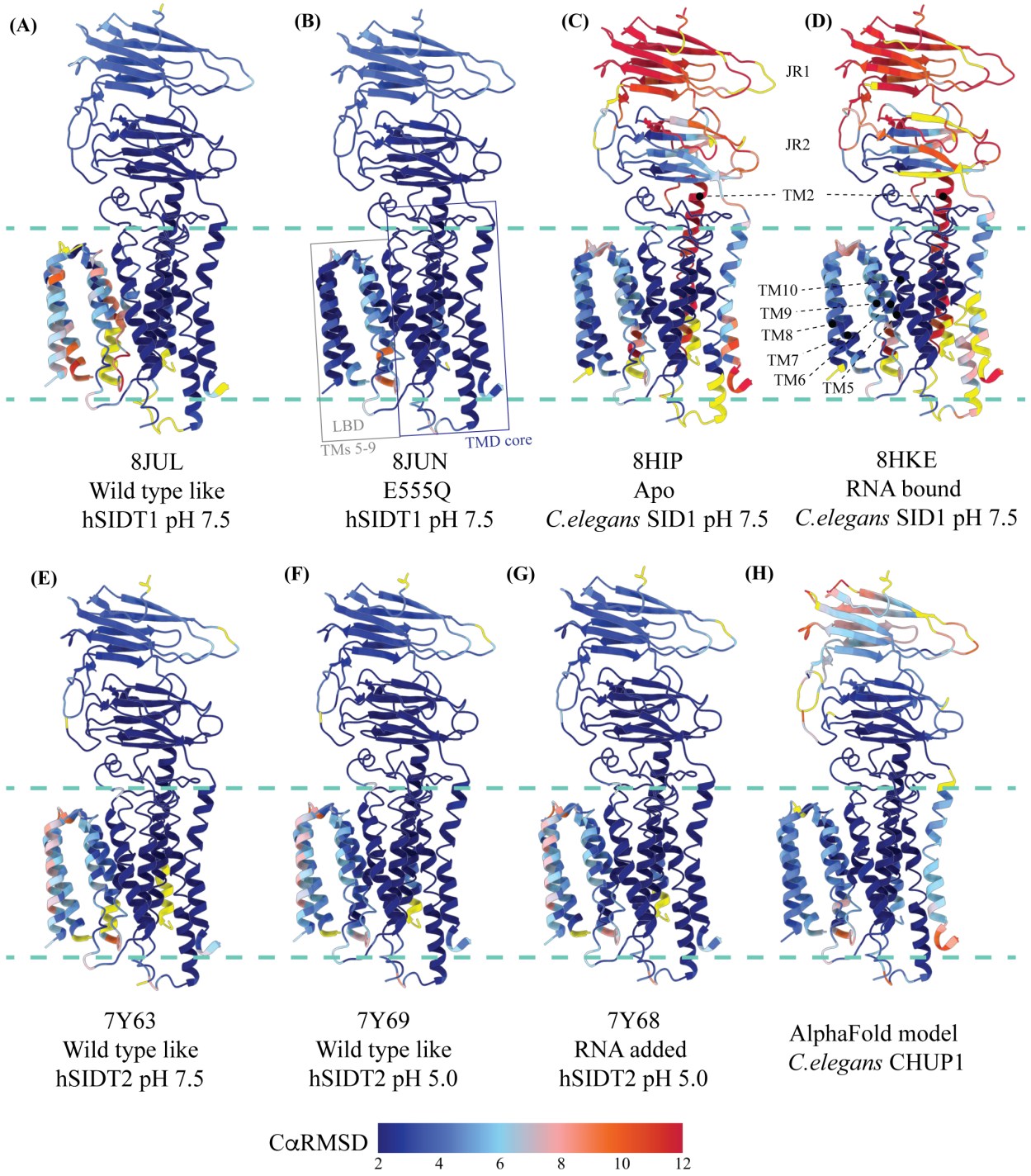

Figure 3: Comparison of hSIDT1 structure with other ChUP family homologs

We used structure based pairwise alignment in ChimeraX to color hSIDT1 structure in this study using the Cα RMSD obtained upon aligning it with various structures (Table S2). The TMD core

of chain A from all the structures was used for alignments. The color key denotes the range of Cα RMSD of the pairwise alignments, with blue for lowest RMSD, red highest, and yellow for regions missing from the alignment. The panels show hSIDT1 colored based on the RMSDs of alignments with **(A)** so-called wild type like (8JUL), **(B)** E555Q mutant (8JUN), **(C)** apo *C. elegans* SID1 (8HIP), **(D)** RNA bound *C. elegans* SID1 (8HKE), **(E)** wild-type like hSIDT2 at pH 7.5, **(F)** wild-type like hSIDT2 at pH 5.0, **(G)** RNA added hSIDT2 at pH 5.0, and **(H)** AlphaFold model of *C. elegans* cholesterol uptake protein CHUP1 (UniProt ID: Q9GYF0) [33, 34]. All the structures used in alignments A-D have been obtained at pH 7.5. Based on the alignments, hSIDT1 structure reported in this study appears to be similar to the E555Q hSIDT1, and the LBD flexibility does not appear to be pH related.

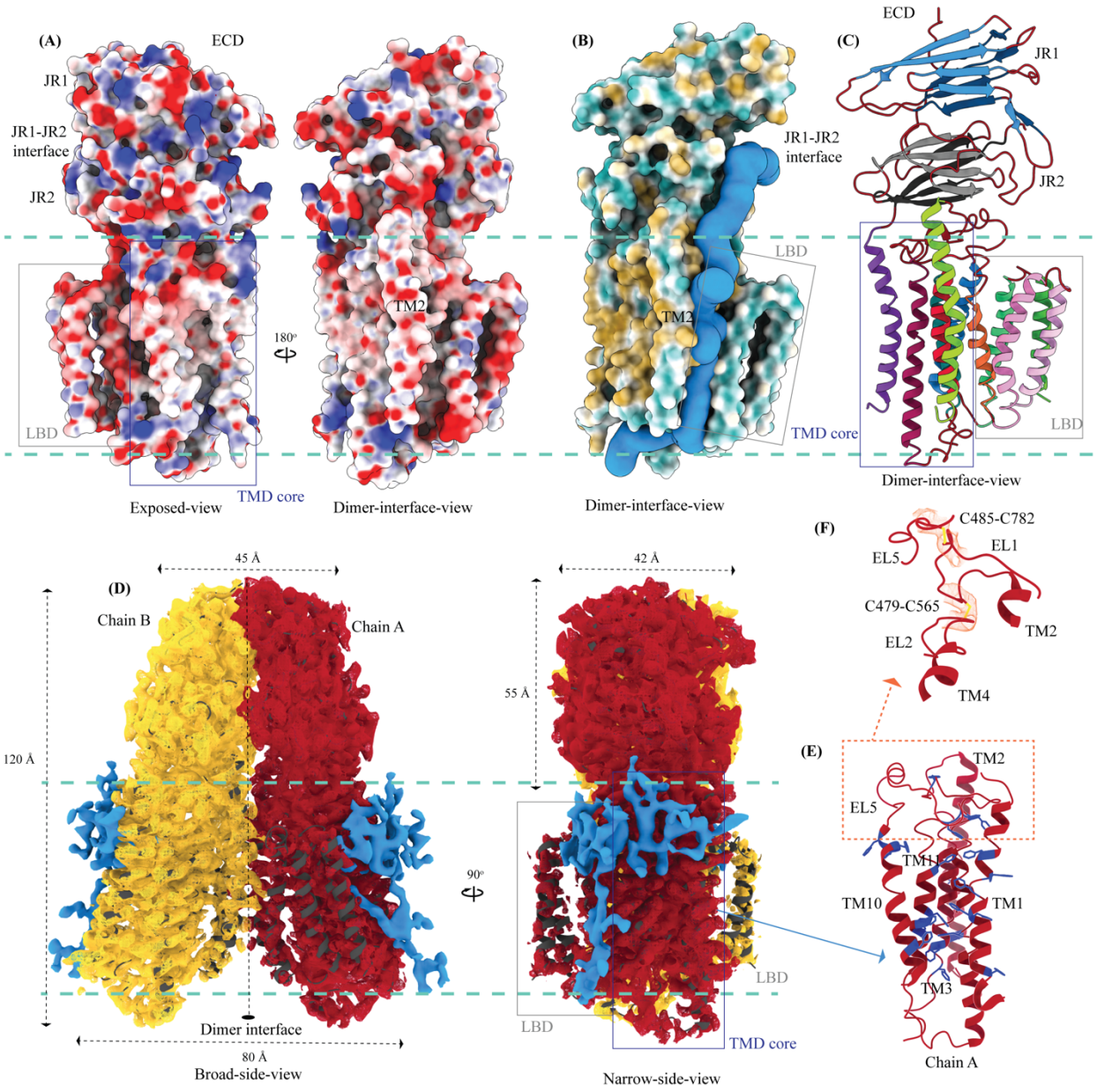

Figure 4: Potential RNA and lipid binding sites

(A) Electrostatic potential and surface charge representation of hSIDT1 protomer. The display color is contoured on the potential from -10 kT (red) to +10 kT (blue). Exposed surface of ECD, especially the JR1-JR2 interface shows a high density of basic residues conducive for RNA binding. The dimer interface of the TMD is lined with uncharged polar residues. **(B)** Surface representation of hSIDT1 protomer colored based on hydrophobic (orange) and hydrophilic (cyan)

amino acids. In blue is the channel that represents a potential path of RNA transport, which was predicted by MOLEonline server using a probe radius of 20Å [60]. A cartoon representation of hSIDT1 protomer in the same orientation is shown in the panel **(C)**. **(D)** Ordered lipid or detergent density observed in the final cryo-EM map is displayed in blue, and density for chain A and chain B are shown in red and yellow respectively (display level 0.12 of the C2 refine map as mesh in ChimeraX). We have not modeled either lipid or digitonin in the blue density because we observe relatively poor local resolution for LBD and the LBD-TMD core interface. **(E)** The TMD core that harbors the phenylalanine highway and metal ion binding site. Highlighted in blue are the phenylalanine residues, primarily from TM11 and TM1, that indicate the presence of phenylalanine highway motif in hSIDT1, as observed in other cholesterol transporters such as ABCG1, NPC1, and PTCH1. **(F)** The disulfides that stitch together EL1, EL2, and EL5 of the TMD core towards the ECD-TMD interface.

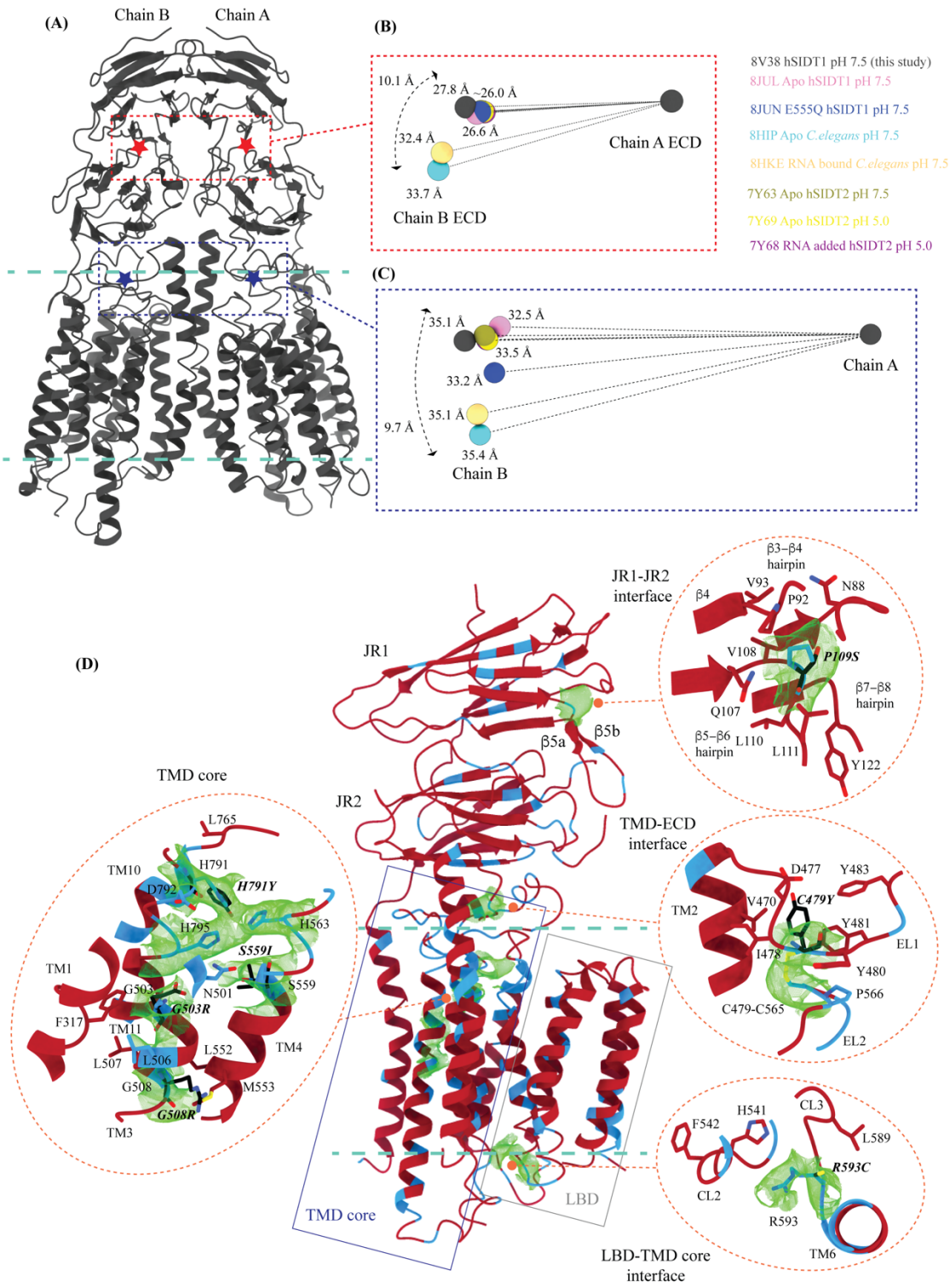

Figure 5: Inter-chain dynamics in hSIDT1 and mutations that attenuate RNAi in *C. elegans*

(A) hSIDT1 dimer (gray) with ECD (red star) center of masses (COMs) and TMD COMs (blue star) highlighted by red and blue dashed boxes. **(B)** The relative position of the COMs of chain B

ECDs of the homologs compared to hSIDT1 from this study (gray sphere). The relative positions of these COMs were calculated by superposing the chain A ECDs of all the homologs. **(C)** The relative position of the COMs of the entire chain B of the homologs compared to hSIDT1 from this study (gray sphere) was calculated by superposing the chain A of all the homologs. In panels (B) and (C) only hSIDT1 COM is displayed for clarity. The distances on the spheres denote the distance of chain B from chain A. The double-ended arrow denotes the maximum distance by which chain B COM has moved with respect to hSIDT1 chain B upon being superposed using chain A. **(D)** hSIDT1 protomer colored based on sequence alignment of ChUP family members. The amino acids that are absolutely conserved are marked in blue. Equivalent mutations that strongly attenuate systemic RNAi mediated by SID1 in *C. elegans* are highlighted in black. The density for those residues in hSIDT1 is displayed (display level between 0.12-0.25 in ChimeraX), and the residue environment is elaborated in the insets. These conserved residue mutations either lie in the channel predicted by our MOLE analysis or in the vicinity of the metal ion coordination site, suggesting the essential role of these sites in RNA transport function.

Tables

Table 1: hSIDT1 cryo-EM data collection, processing, and validation statistics

	EMDB-42943 and PDB-8V38
Data collection and processing	
Magnification	105,000x
Voltage (kV)	300
Data collection mode	Super-resolution
Electron exposure (e–/Å^2)	50
Defocus range (μm)	-1.2 to -2.5
Physical Pixel size (Å)	0.848
Symmetry imposed	C2
Initial particle images (no.)	3204173
Final particle images (no.)	122683
Map resolution (unmasked, Å) at $FSC_{0.143}$	3.98
Map resolution (masked, Å) at $FSC_{0.143}$	3.44
Map resolution range (Local resolution)	2.0-8.0
Refinement	
Map sharpening B factor (Å^2)	-30
Model composition	
Chains	2
Atoms	9373 (Hydrogens: 0)
Residues	Protein: 1270 Nucleotide: 0
Water	0
Ligands	0
Bonds (RMSD)	
Length (Å) (# > 4sigma)	0.006 (0)
Angles (°) (# > 4sigma)	1.081 (28)
MolProbity score	1.61
Clash score	2.28
Ramachandran plot (%)	
Outliers	0.64
Allowed	11.56
Favored	87.8
Rama-Z (Ramachandran plot Z-score, RMSD)	
whole (N = 1254)	-3.33 (0.21)
helix (N = 482)	-1.28 (0.21)
sheet (N = 254)	-0.30 (0.33)
loop (N = 518)	-4.24 (0.21)
Rotamer outliers (%)	0.00
Cβ outliers (%)	0.00
Peptide plane (%)	
Cis proline/general	4.3/0.0
Twisted proline/general	4.3/0.0

Cα BLAM outliers (%)	8.08
ADP (B-factors)	
Iso/Aniso (#)	9373/0
Protein (min/max/mean)	10.24/153.29/76.08
Model vs. Data	
CC (mask)	0.8
CC (box)	0.6
CC (peaks)	0.59
CC (volume)	0.77