Intramembranal disulfide cross-linking illucidates the super-quaternary structure of mammalian CatSperms

Christopher Bystroff
Department of Biological Sciences
Rensselaer Polytechnic Institute
Troy, NY 12180

ABSTRACT

CatSper is a voltage-gated calcium channel located in the plasma membrane of the sperm flagellum and is responsible for triggering hyperactive motility. A homology model for the transmembrane region was built in which the arrangement of the subunits around the pseudo-four-fold symmetry axis was deduced by the pairing of conserved transmembranal cysteines across mammals. Directly emergent of the predicted quaternary structure is an architecture in which tetramers polymerize through additional, highly conserved cysteines, creating one or more double-rows channels extending the length of the principal piece of the mammalian sperm tail. The few species that are missing these cysteines are eusocial or otherwise monogamous, suggesting that sperm competition is selective for a disulfide-crosslinked macromolecular architecture. The model suggests testable hypotheses for how CatSper channels open in response to capacitation and mechanical force. A source of the concomitant disulfide isomerase activity is found in CatSper-associated proteins β, δ and ε.

INTRODUCTION

Successful transmission of genetic material during sexual reproduction in mammals provides strong selective pressure for sperm motility, especially in species where sperm competition takes place. CatSper is a sperm-associated voltage-gated calcium channel that is required for the hyperactive motility that is required to drive the sperm through the outer layers of the oocyte (1). Knock-out of any of the four subunits of CatSper (CatSper1-4) in mouse produces infertile males but has no phenotypic effect on the female and no additional effect on the male (2). CatSper is a member of the ion transport protein family (pfam00520), having six transmembrane helices and an ion selective pore. Several atomic-resolution structures exist for the homologous voltage-gated cation channel sub-families NaV, Kv and CaV (3). Each consists of four transmembrane helical segments of the voltage-sensing (VS) domain orbiting like satellites around the central pore domain (P). The P-domain is made up of four copies of the two C-terminal transmembrane segments with an intervening, pore-defining “P-loop”. The P-loop of CatSper contains a conserved sequence motif, TxDxW, that is associated with a calcium ion selective channel (4).

CatSper is the only known natively heterotetrameric CaV. All others are homotetramers or a single chain fusion of four transmembrane domains (5), with the special exception of Xenopus epithelial Ca^{2+} channels TRPV5 and TRPV6 which, when co-expressed, assemble into functional but non-native heteromeric calcium channels with different ion selectivities (6). The requirement for four CatSper genes in mammals suggests that there is a structural selective pressure that goes beyond the form and function of the individual channel, since one gene would do if sperm required only a voltage-gated and ion selective channel. The arrangement of CatSper into four lines extending the length of the principal piece of the sperm flagellum (60), and the observed
association of defects in this architecture with reduced motility (7), suggests a connection between the CatSper macromolecular architecture and hyperactive motility.

CatSper is a target for immune-based contraception. A reliable model will help to identify antigenic sites. Also, a model for the architecture of CatSper may help us to understand the relationship between its macromolecular form and its function in capacitation-induced hyperactive motility. Subunit-specific sequence conservation in the context of structure provides hints as to the basis of the selective pressure. Here we discuss conserved cysteine residues in mammalian CatSpers. Their cross-linking ability may serve to dictate the relative positioning of the four subunits. Here we describe a quaternary structure and an emergent higher-order (super-quaternary) structure that carries implications with regard to the mechanisms of voltage-gated, pH-induced calcium channel opening in capacitated mammalian sperm.

MATERIALS AND METHODS

Bioinformatics

Homolog sequences for each CatSper subunit were identified by searching the RefSeq (11) database using BLASTp (43), restricting the search to mammals, and requiring the Entrez (44) search string "cation channel sperm-associated protein [n]" where n was 1,2,3,4 respectively in four BLASTp searches. Sequences with the annotation "LOW QUALITY SEQUENCE" or which lacked any of the six transmembrane regions, or of species that were not represented by all four subunits, were removed. All four subunits of each of the remaining 41 species were concatenated in numerical order and the N- and C-terminal sequences were removed for further analysis. Sequences were aligned in UGENE (45). Sequences were removed if they exceeded 92% sequence identity to any other sequence, leaving 24 species. Poorly aligned regions were omitted before calculating phylogenetic trees. Phylogenetic trees were computed in UGENE using the PhyML Maximum Likelihood (47) method, using the Blosum62 substitution matrix. Regions around conserved cysteines were dissected out using in-house fortran programs and presented along with the phylogram using the ETE python objects (48). Modified "Zappo" coloring is used throughout (49).

For multiple sequence alignments of individual CatSper subunits, all species passing the above filters were included after removing those with 90% identity or greater, where the cutoff was based on all aligned positions. Any N- or C-terminal sequences that aligned poorly were left unaligned in Figure 1.

The significance of finding four disulfide isomerase motif sequences CXC or CXXC within three out of five CatSper-associated sequences was calculated as the empirical probability of being conserved across mammals in CatSper-associated proteins (β:0.59,γ:0.51,δ:0.59,ε:0.46,ζ:0.45) times the frequency of cysteine respectively in those sequences (β:0.019,γ:0.019,δ:0.024,ε:0.022,ζ:0.0089) squared because there are two cysteines in each motif, times the respective length of the protein (β:994,γ:997,δ:781,ε:842,ζ:197). The expected number of CxC or CxxC motifs by chance in the five proteins is 0.12. The chance likelihood of four motifs was found by randomly sampling this distribution 10000 times and summing the number of times that four motifs or more was found across three or more proteins, p-value=0.0009. Therefore chance occurrence is rejected with 99.9% confidence.
Transmembrane helix prediction was done using TMHMM (50). Secondary structure prediction for globular proteins was done using PsiPred (51). Disordered regions were predicted using DISOPRED. Glycosylation sites were predicted using NetNGlyc (52). Compositional analysis, tree pruning, pH and charge calculations were carried out using Microsoft Excel 2016 and in-house Fortran programs. Sequence motifs were identified using PrositeScan (53). Charge calculations were done by summing the formal changes of the sidechains times the pH-dependent fraction charged using the Henderson-Hasselbach equation.

Molecular modeling

All molecular modeling was carried out using MOE (Chemical Computing Group, Montreal). All six possible subunit arrangements for the P-loop segment were built using MOE’s homology modeling tools. Side chains were repacked using MOE’s Gizmin function while holding backbone atoms fixed. Multiple tetramer models were docked manually to form dimers and higher-order oligomers by searching in the space of two translational and one rotational degree of freedom. Rotations and translations were constrained to the plane of the membrane. Conserved intramembranal cysteines that could come into contact were assumed to form disulfide bonds (54).

Template-based remote homolog three-dimensional structure predictions for cytoplasmic CatSper domains and CatSper associated proteins were carried out using RaptorX (55). In these cases, several mammalian homolog sequences were predicted independently and the predictions were compared to extract consensus features. No consensus structures were found.

Disulfide prediction

Evolutionary conservation of cysteines in structural context predicts disulfide formation (56, 57). Specifically, proteins that contain cysteines which find themselves within 6Å of each other in the folded state readily crosslink when expressed in the lumen of the ER, depending on the presence of protein disulfide isomerases (58). Disulfides do not dictate the protein topology, but they serve to lock the structure into place once formed. Moreover, insertion of cysteines into a sequence is a technique for the determination the transmembrane structure by identifying positions that support spontaneous crosslinking (59, 54).

RESULTS

Terminology and notation

Sequence positions denoted (Mm<n>:<residue><number>) refer to the Mus musculus sequences for Catsper n (n=1-4, β, γ, δ, ε, ζ). Positions in sequences from other species are referenced using the aligned mouse sequence numbering. Transmembrane (TM) helical segments are denoted s1 through s6, where s1-s4 constitute the VS-domain and s5-s6 constitute the P-domain. Subunit arrangements for the CatSper tetramer around the pseudo-four-fold axis are denoted in clockwise order as viewed from the extracellular side. Thus, arrangements 1234, 2341, 3412 and 4123 all reference the same tetramer. We introduce the term "super-quaternary" to mean the arrangement of tetramers into higher-order macromolecular structures.

Homology modeling of CatSper monomers
The closest homolog of known structure to CatSper is an archael voltage-gated calcium channel, (CaV) from *Arcobacter butzleri* (8). The archael and mammalian sequences share only “twilight zone” (9) similarity and then only in the transmembrane regions. But ancillary data makes the assignment of homology much more solid. Specifically, all mammalian CatSpers have six TM segments, a pore-forming motif between s5 and s6, and the presence of periodic arginines in s4 that are characteristic of a VS-domain (10).

The RefSeq database (11) contains between 100 and 130 mammalian CatSper sequences for each subunit. As a group the four paralogs are characterized by conservation only in the TM region (Figure 1). The N- and C-terminal domains of the individual subunits are conserved across mammals but no detectable homology was found between any of the terminal domains and any protein of known structure, therefore the homology models discussed here cover only the TM region. Interesting features of the terminal domains will be discussed later. The TM regions show intersubunit conservation in the TM helical segments, in the internal loops and in the P-loop. CatSpers also align confidently to the archael CaV template within these segments (*Supplemental Figure S1*). The spacial positioning of amino acids in sequences that are confidently predicted to be homologs is generally regarded as reliable for the aligned regions, sufficiently so that homology models are routinely used to solve crystal structures by molecular replacement. Expectation values for alignments of mouse CatSpers 1-3 to archocobacter butzleri CaV (PDBID: 4MS2) were highly significant (2e-18, 6e-19, 1e-7). Although CatSper 4 falls below BLAST-detectable similarity of 4MS2, it does align significantly (3e-9) to the rabbit CaV1.1 structure (PDBID:5GJW) (12) which superposes with 4MS2 with RMSD 3.44Å. The two potential templates are remarkably similar in their TM structure despite having only 27.3% sequence identity (*Supplemental Figure S5*). As expected, most of the external loops are highly variable and cannot be confidently modeled. In the discussions below, important inferences are confined to the TM helices and one of the internal loops, which are confidently aligned.

**Tetrameric subunit arrangement**

*Disulfide 1 -- Mm4:C170 to Mm1:C499*

The template archael CaV is homomeric but CatSper is heteromeric. There are six ways to organize four subunits clockwise around a pseudo-four-fold axis, abbreviated 1234, 1324, 1243, 1342, 1423, and 1432. We may assume that the energetic interactions between monomers dictate the selection of binding partners, but the prediction of the subunit arrangement is a prohibitively difficult computational docking problem given the uncertainties in energy calculations and the expected errors in the homology-based models. Conserved cysteines within the transmembrane segments provide the most compelling evidence for the correct arrangement of the subunits (Figure 2). Mm1:C499 sits on s5 near the internal face, while Mm4:C170 sits on s4, also near the internal face. These two would form a disulfide if subunit 4 packed clockwise against subunit 1. The nature of the interaction is best described as the wrapping of subunit 4 P-domain and VS-domain around the P-domain of subunit 1 (Figure 3). Both Mm4:C170 and Mm1:C499 are strictly conserved in mammals. Moreover, no other cysteine partners are available for Mm4:C170 on s5 of any of the other three CatSper subunits in any mammalian species, and no other cysteine partners are available for Mm1:C499 on s4 of any of the other three subunits in any mammalian species. Their strict conservation and the conspicuous absence of any alternative pairings for either subunit 1 or 4 in any species leads us to conclude that CatSper 4 (Mm4:C170) crosslinks with CatSper 1 (Mm1:C499) in a clockwise 41 arrangement. Crosslinking CatSper 4 and 1 eliminates four alternative arrangements, leaving just 1234 and 1324.
Disulfide 2 -- Mm3:C197 to Mm2:C210

The only other conserved, outward-facing cysteine in the P-domain of any CatSper subunit is Mm3:C197, which sits on s5 near the extracellular surface. An approximate 10Å difference in depth within the membrane makes Mm3:C197 unreachable by either of the Disulfide 1 cysteines Mm4:C170 or Mm1:C499. The only cysteine on the P-domain-facing surface of the VS domain on any subunit is Mm2:C210, which sits on s4 near the extracellular surface. Mm3:C197 and Mm2:C210 would come within 6Å and crosslink if CatSper2 binds CatSper3 in a clockwise fashion. With 6 exceptions, no other cysteines are present at a position to pair with Mm3:C197 in any mammalian species. Those six exceptions (2 mole rats, guinea pig, common shrew, chinchilla, and degu) have a pairable cysteine at or near position Mm4:P160. However, if CatSper4 is already bound to CatSper1, then position Mm4:P160 is buried by the 41 dimerization interaction and unavailable for pairing with CatSper3. If we assume that the prior crosslinking of CatSpers 4 and 1 blocks access to Mm4:P160, then there are no alternative pairings for either Mm3:C197 or Mm2:C210 in any mammalian species. We conclude that Mm2:C210 crosslinks with Mm3:C197 in all mammals, and that therefore the subunit arrangement is 1234, not 1324. CatSpers 1 and 4 pair to form a 41 heterodimer, and CatSpers 2 and 3 simultaneously form a 23 dimer.

The final assembly of the channel consists of the specific association of the CatSper dimer 41 to the dimer 23. Since there are no additional conserved cysteines in positions to define the specificity of this tetramerization event, other reasons must be cited to account for selectivity of the 41 + 23 pairing over the 41 + 41 or 23 + 23 mispairings, which are also possible. We can safely say that both mispaired tetramers, 1414 and 2323 either do not form or are sub-functional, based on the CatSper-null mouse studies in which the knock-out of any subunit leads to hyperactive motility-impaired sperm (13). Specific shape complimentarity interactions must account for the selective tetramerization of the correct and functional form. However, it cannot be ruled out that one of the CatSper associated proteins acts as a chaperone to block the formation of the mispaired tetramers.

Superquaternary structure

The strong selective pressure for a heterotetrameric ion channel may be explained by considering the known organization of CatSper into four quadrilaterally-arranged rows along the principal piece of the flagellum (14). If dictated by the CatSper structure itself, this organization would require asymmetry in the tetramer since specific interactions of a homomeric complex would be equally distributed in four directions, not just in two. Specific interactions between subunits of a heterotetrameric arrangement can form straight line polymer.

Disulfide 3 -- Mm1:C362 to Mm4:C114

Upon inspection of the multiple sequence alignments in the context of the 1234 tetramer model, we find that cysteines Mm1:S393 and Mm3:C96 sit on the external, intramembranal surfaces of the VS-domains. Interestingly, these conserved cysteines vary across mammals in their sequence positions within the VS-domain. Another pair of cysteines, Mm1:C362 and Mm4:C114, are also externally-facing, intramembranal, and conserved across mammals, but these two occupy a fixed position throughout their evolutionary history (Figure 2). A cysteine that pairs with itself on another tetramer (i.e. self-reacting) always sits at the same depth in the membrane as
its crosslinking partner, which is itself, regardless of whether it has mutated to a new depth. However, if two different cysteines pair (i.e. cross-reacting) then there is no simple mutational pathway to a shifted depth within the membrane. Two cysteines would have to simultaneously appear in crosslinkable orientations, which is highly unlikely. Therefore, we expect that cysteines that vary in position over evolutionary time are probably self-reacting, while cysteines that remain fixed in their position are probably cross-reacting.

By assuming that the two fixed-position cysteines, Mm1:C362 and Mm4:C114, crosslink we create a linear polymer (Figure 4). Although it is sterically possible that one of the variable-position cysteines, either Mm1:S393 or Mm3:C96 could pair with one of the fixed-position cysteines, such pairing would create a 90° rotation and further cross-linking would generate a tetramer of tetramers, not a linear polymer. Both the linear polymer and the tetramer of tetramers are possible, but only the linear polymer agrees with what we know about the macromolecular architecture of CatSper. We assume once again that energetic packing interactions direct and determine the true super-quaternary structure, which is then locked into place by the actions of disulfide isomerases.

Disulfide 4 -- Mm3:C96 to Mm3:C96

If we accept the prediction of a linear polymer created by Disulfide 3, then the CatSper 4 VS-domain is docked to a notch between the VS-domains of CatSpers 1 and 2, and the VS-domains of CatSpers 1 and 3 remain exposed. The two variable-position cysteines Mm1:S393 and Mm3:C96 are positioned to self-react if two linear polymers are arranged antiparallel. Either the CatSper 1 side or the CatSper 3 side of the polymer may self-react. There is a no clear energetic difference between these two possibilities except the energetic ramifications of the positioning of the highly charged N-terminal domain of CatSper 1. Electrostatic repulsion would disfavor the self-association of CatSper 1's. We tentatively propose a zipper structure in which CatSper polymers are cross-linked at Mm3:C96 (Figure 4). It is not likely that the CatSper polymer rows associate in a parallel fashion by cross-reacting disulfides between Mm1:S393 and Mm3:C96 for the reasons described above regarding the evolutionary variability of the depth of these cysteines in the membrane, and because we do not see a correspondence across species between the depths of these cysteines in the membrane. In some mammals this cross-link is possible but not in all.

DISCUSSION

CatSper role in sperm competition.

Swim speed is one of many phenotypic effector mechanisms selected by sperm competition (15), which also include sperm size, testes size, ejaculate volume, and seminal fluid proteins (16). Additionally, many factors, not just CatSper function, influence swim speed in vertebrate sperm, including cell length and morphology (17) and the size of the mid-piece (18). Thus no one factor explains success in sperm competition. But genetic adaptations in CatSper offer a uniquely adaptive advantage that they are expressed exclusively in the male and therefore suffer no conflicting manifestations in the female (19). Combining strong selective pressure for motility with the absence of genetic interference may have focused evolution on the CatSper genes.
Three of the four cysteines proposed to form Disulfide 1 (Mm1:C499, Mm4:C170) and Disulfide 2 (Mm2:C210, Mm3:C197) are 100% conserved in mammals. Mm3:C197 is missing only in naked mole rat (*Heterocephalus glaber*) and Chinese horseshoe bat (*Rhinolophus sinicus*). Naked mole rat is also missing Disulfide 3 cysteine Mm1:C362, and Disulfide 4 cysteine Mm3:C96, implying defects in subunit arrangement, polymerization, and cross-linking of rows in that species. Indeed, the swim speed for naked mole rat sperm has been quantified and characterized as "very slow", and the slowness was attributed to the absence of sperm competition in that eusocial species (20). Horseshoe bat has lost the proposed tetramer determinant Mm3:C197 from Disulfide 2, suggesting an impact on sperm motility. Unfortunately there is no quantitative data on sperm motility in this or any bat. We can assert, however, a well-supported presence of sperm competition in bats, shown using comparative anatomy (21), and if taken as proof of sperm motility this suggests that loss of just one of the conserved disulfides is not enough to adversely affect hyperactive motility.

Disulfide 3 (Mm1:C362, Mm4:C114) is proposed to polymerize the CatSper tetramers into a row, but one or the other cysteine is absent in naked mole rat (*Heterocephalus glaber*), star-nosed mole (*Condylura cristata*), leopard (*Panthera pardinus*), alpaca (*Vicugna pacos*) and elephant shrew (*Elephantulus edwardii*). Elephant shrews lack Mm1:C362, and also lack Disulfide 4 (Mm1:C393) which is hypothesized crosslink CatSper rows into a zipper. No data is available on the motility of elephant shrew sperm; however it is known that elephant shrews are socially monogamous (22, 23), which would appear to reduce the likelihood of sperm competition as a selective pressure. Some elephant shrews have sperm that are much shorter than those of such polyandrous species as rodents and primates (24). Our model predicts that elephant shrew sperm motility, once it is measured, will be found to be slow.

The leopard (*Panthera pardus*) and other cats lack one cysteine of Disulfide 3 (Mm1:C362). But these animals are both promiscuous in their sexual behavior and show clear anatomical signs of sperm competition (25). A closer look revealed that two partially-conserved cysteines (Mm2:C71, Mm3:C39) exist at positions just N-terminal to the start of the first TM helix of CatSpers 2 and 3 in cats and many other genera. This pair can form a disulfide in our model, which may serve as a "backup" in case Disulfide 3 is lost by mutation. Most mammals have both Disulfide 3 (Mm1:C362, Mm4:C114) and the "backup" (Mm2:C71, Mm3:C39) and both are compatible with the same polymer model. The "backup" crosslink is absent in star-nosed mole, alpaca, and elephant shrew. Alpaca and all other camelids also have the mutation Mm4:C114F, which would imply the absence of both Disulfide 3 and its "backup". Although new world camelids (alpaca) have a cysteine at Mm4:Q48 which is spacially positionable to link to Mm1:C362 and rescue the Mm4:C114F mutation, old world camelids do not have this mutation. Thus, our model predicts that old world camelids may have slow swimming sperm. There are no studies to date on sperm motility in these species. Camelid's ovulation is induced by seminal plasma (26, 27) but induced ovulation has never been tied to the absence of sperm competition.

Sites on the outward facing edge of either CatSper 1 or CatSper 3 VS-domains are proposed to cross-link antiparallel CatSper polymers, but both Disulfide 4 (Mm3:C96) and its alternative self-reacting cysteine (Mm1:C393) are absent in two species: elephant shrew and lemur. Elephant shrew has been discussed above as probably lacking sperm competition. However, since lemur species probably do exhibit sperm competition (28, 29), it would be premature to conclude that these zipper cross-bridges are required for sperm hyperactive motility.

**The case for disulfide isomerase activity in CatSper associated proteins β, δ and ε.**
Five CatSper-associated proteins have been identified, CatSper β (30), γ (31), δ (32), ε (7) and ζ (7). CatSper β, δ and ε each contain at least one strongly conserved instance of the CXC or CXXC motif, the presence of which is associated with disulfide isomerase activity (33). The two cysteines in Mmβ:920-CNCT are 100% conserved and those in a second instance at Mmβ:187-CPC-189 are 80% conserved in a representative set of 27 mammalian sequences (Figure S2). CatSper δ has a C-terminal region that is rich in cysteine. In this region, a CxC or CxxC motif occurs in 74% of the representative set of 38 mammalian sequences. Interestingly, African elephant is missing the cys-rich region but has a CXXC motif at another location, Mmδ:305-CLAC, possibly suggestive of a rescue-of-function mutation. CatSper δ also contains a 100% conserved CCxxxxxC motif of unknown function, but which is reminiscent of the SCCS motif that has been associated with mammalian selenoprotein thioredoxin reductase (32). Selenium is known to up-regulate CatSper expression in mice (61). CatSper ε has a CXXC motif at Mmε:224 which is present in 27 out of 30 mammalian species. Two of the three that do not contain CXXC (ground squirrel and star-nosed mole) do contain CXXXC. Only mongolian gerbil is missing any disulfide isomerase motifs in CatSper. Chance occurrence of four CxC or CxxC motifs occurring in three out of five CatSper associated proteins was rejected with 99.9% confidence using the resampling method.

No disulfide isomerase or any other enzyme activity has yet been experimentally determined for CatSper associated proteins. Moreover, the fragmentation of the CatSper stripes that was cited here as evidence for a defect in the disulfide crosslinking was seen in CatSperζ-null mutant sperm (7) even though CatSperζ has no disulfide isomerase motifs.

The case for a mechanotransduction function of CatSper zipper architecture

The presence of rheotaxis in sperm requires an explanation for how the forces generated by fluid flow manifest in a rotation of the tail (35). Mechanotransduction of calcium channel opening provided by the proposed zipper architecture provides such an explanation. Specifically, forces exerted on the membrane may stretch the zipper, pulling on the VS-domains where they are crosslinked. The VS-domains pull on the s4s5 linker helix, whose positioning is known to change upon opening of the cytoplasmic gate (36). In superresolution images, CatSperζ-null mouse sperm labeled on CatSper1 show fragmentation (7) and are impaired in rheotaxis. Channel opening by mechanotransduction through disulfide cross-links would explain the coincidence a fragmented CatSper structure and the loss of efficient rheotaxis.

The case for pH sensing in the N terminal domain of CatSper 1

While all of the N-terminal domains of CatSper1-4 have a positive charge, the size and charge of CatSper1 N-terminal domain goes well beyond that required by the "positive-inside rule" of integral membrane proteins (37) and suggests that the domain has a role beyond simply dictating the topology of the TM helices. An abundance of histidines (Figure 1) suggests a pH sensing function. In the epididymis, at pH 6.5, 24% of these histidines would be charged whereas upon ejaculation and mixing with the prostatic fluids to pH 8.0 (38) only 1% of the histidine sidechains would remain protonated. This pH shift represents an average shift of -12 in the ionic state of the domain. This domain is highly variable across mammals and the pH-induced shift varies from -38 charge units for Pacific Walrus to as little at -1.4 for Little Brown Bat and -1.2 for Alpaca. The charge shift suggests a possible "priming" mechanism for gate opening. A loss of
positive charge, and in some species the acquisition of a negative charge, decreases or reverses the attractive force between the CatSper1 N-terminal domain and the negatively charged C-terminal domains, leading to a net repulsion of the C-terminal domains. The C-terminal domains are attached to the hydrophobic gate residues of s6.

The case for a flippase function in CatSper

Flippases, floppases and scramblases are transmembrane proteins that catalyze the translocation of membrane lipids from the inner to the outer leaf, from outer to inner, or both, respectively. Although the mechanism of translocation is still unknown, the crystal structure of PgiK flippase contains a salt bridge cluster that suggests a hypothetical translocation intermediate containing a charge-stabilized buried lipid head group. A large conformational change is required for flippase activity, as demonstrated by the inhibition of its function by a cross-binding nanobody (39). In other studies, it was experimentally demonstrated that same side positioning of multiple Arg or His on a transmembrane α-helix was sufficient to generate flippase activity, using only short peptides (40). A similar cluster of structurally-aligned arginine and glutamate sidechains is found in the interface between CatSper VS- and P-domains where it may be subject to voltage-induced conformational changes. Mammalian sperm undergo capacitation-induced membrane reorganizations by hitherto unknown flippases (or floppases or scramblases) (41), and CatSper's voltage-gated channel activity is capacitation-induced. This circumstantial evidence of a flippase-like structure, plus the experimentally observed membrane reorganization, does not together constitute proof that CatSper is a flippase, but it provides a plausible hypothesis. If so this may be an instance of "accidental" adaptation of a new function from an existing one (voltage gating), absent any obvious selective advantage. It is notable that the coincidence of ion channel and flipping activities has been previously documented in the case of the Ca2+-activated chloride channel TMEM16F in mouse, a voltage-gated channel having scramblase activity; specifically, the knock-out of TMEM16F produces a defect in phospholipid scrambling in platelets (42).

ACKNOWLEDGEMENTS

This work was supported by the NIH grant R01 GM099827. Special thanks to Jean-Ju Chung (Cellular & Molecular Physiology, Yale University) for guidance and for critical reading of this manuscript.
REFERENCES


FIGURES

Figure 1. Multiple sequence alignment for mammalian CatSper1-4. Ten representative mammalian sequences were used for each of subunit in the TM region, comprising the VS and P-domains. N and C-terminal regions were aligned separately for each subunit using 55, 23, 39, and 36 mammalian sequences for CatSpers 1-4 respectively, based on a cutoff of 92% maximum sequence identity. Pie charts show amino acid composition of terminal domains color-coded as shown along the bottom. Names for transmembrane segments are indicated along the bottom.

Figure 2. Blocks of multiple sequence alignment for 24 representative mammals were extracted from around the conserved cysteine positions that are discussed in the text. *Mus musculus* position labeled with a dot. Lower panel shows the location of the cysteine (small yellow circle) in the cartoon tetramers (circles: VS domain, squares: P domain, 1:blue, 2:green, 3:yellow, 4:red). Connecting lines indicate predicted disulfide crosslinks 1 through 4 as numbered in the text. "5" marks an alternative to Disulfide 4, a self-reacting crosslink. Mm4:C233 is strictly conserved but is not outward facing and cannot form a disulfide.

Figure 3. CatSper tetramer model (1:blue, 2:green, 3:yellow, 4:red) showing Disulfides 1 (left) and 2 (right) in black boxes. Lower panel shows the difference in depth of the two disulfides in the membrane. Dotted lines show approximate location of membranes surfaces (top:external, bottom:internal).

Figure 4. Zipper model in ribbon and cartoon representations. Two linear rows of CatSper tetramers are polymerized via cross-reacting Disulfide 3 linking CatSpers 1 (blue) and 4 (red). Antiparallel rows are linked at CatSper 3 (yellow) by self-reacting Disulfide 4. In cartoon version disulfide linkages are shown as small yellow circles.
Figure 1.
Figure 2
Supplemental Information for "Intramembranal disulfide cross-linking illuminates the super-quaternary structure of mammalian CatSpers"

C. Bystroff

**Figure S1.** Mouse CatSper sequences (mCS1-4) in the transmembrane region aligned to the Arcobacter butzleri voltage-gated calcium channel crystal structure (4MS2). This alignment was used to assign coordinates for the TM segments and P-loop of each monomer (marked).
**Figure S2.** Aligned blocks from two regions of a multiple sequence alignment of 27 mammalian CatSper β’s. Mouse numbering is given above. The first CXC motif is 80% conserved, the second is 100% conserved in mammals. The CatSperβ multiple sequence alignment has a distribution of conserved and variable regions with very few indels, consistent with an ordered three-dimensional structure.
Figure S3. C-terminal sequences from 38 mammalian CatSper δ’s are poorly conserved, but consistently cysteine-rich. 74% of species have the CxxC or CxC motif. African elephant lacks the C-terminal extension but has the motif at another location. Mouse numbering.
Figure S4. C-terminal sequences from 32 mammalian CatSper ε’s. Arrows mark CXXC motif. Only three species are missing this motif. Mouse numbering.
Figure S5. Least-squares superposition of crystal structure of Ca\textsubscript{v} from arcobacter butzleri (4MS2, yellow) with cryoelectron microscopy structure of rabbit Ca\textsubscript{v} 1.1 (5GJW, N-to-C blue-red coloring). RMSD in superposed alpha carbons in 3.44Å. Sequence identity is 27.3%. 4MS2 is homotetrameric. 5GJW is a single chain fusion of four repeats arranged clockwise around the pore from N to C. (a) pore view. (b) membrane view.