Sequence and structural conservation reveal fingerprint residues in TRP channels.

Deny Cabezas-Bratesco1, &, Francisco A. McGee2, &, Charlotte K. Colenso1,3, Kattina Zavala4, Danielle Granata2, Vincenzo Carnevale2, Juan C. Opazo4,5,6, Sebastian E. Brauchi1,6,7

1Instituto de Fisiologia, Facultad de Medicina, Universidad Austral de Chile, Valdivia, 511-0566, Chile.

2Institute for Computational Molecular Science and Department of Biology, Temple University, Philadelphia, USA.

3School of Cellular and Molecular Medicine, University of Bristol, Bristol, BS8 1TD, UK.

4Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile.

5Integrative Biology Group, Universidad Austral de Chile, Valdivia, Chile

6Millennium Nucleus of Ion Channel-associated Diseases (MiNICAD), Valdivia, Chile.

7Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, 20147, USA.

& D.C-B. and F.A.M. equally contributed to this work.

Correspondence should be addressed to
Juan C. Opazo: jopazo@gmail.com
Sebastian E. Brauchi: sbrauchi@uach.cl

Running title: A Structural Signature in TRP Channels.
ABSTRACT

TRP proteins are a large family of cation-selective channels, surpassed in variety only by voltage-gated potassium channels. Detailed molecular mechanisms governing how membrane voltage, ligand binding, or temperature can induce conformational changes promoting the open state in TRP channels are still a matter of debate. Aiming to unveil distinctive structural features common to the transmembrane domains within the TRP family, we performed phylogenetic reconstruction, sequence statistics, and structural analysis over a large set of TRP channel genes. Here we report an exceptionally conserved set of residues. This fingerprint is composed of twelve residues localized at equivalent three-dimensional positions in TRP channels from the different subtypes. Moreover, these amino acids are arranged in three groups, connected by a set of aromatics located at the core of the transmembrane structure. We hypothesize that differences in the connectivity between these different groups of residues harbors the apparent differences in coupling strategies used by TRP subgroups.
INTRODUCTION

Transient receptor potential (TRP) proteins constitute a large family of cation-selective ion channels involved in a number of physiological functions (Clapham, 2003; Nilius et al., 2007). Changes in TRP channel function and expression are associated with a variety of metabolic, respiratory, cardiovascular, and neurological diseases (Nilius et al. 2007; Nelson, Beck, and Cheng 2011; Wang et al. 2020). Moreover, the abnormal expression of TRP channels has been related to cancer development and progression (Shapovalov et al. 2016; Yang and Kim 2020). TRP channels are therefore an attractive target for pharmacological development, and the understanding of their inner workings is critical for such endeavors.

TRPs have been shown to be evolutionarily related to all voltage-gated cation channels (VGCC), two-pore channels (TPC or TPCN), and CatSper channels (Yu et al., 2005). The TRP family is composed of two major groups (Groups I and II) and ten subfamilies, or subtypes: TRPA1, TRPV, TRPVL, TRPC, TRPM, TRPS, TRPN, TRPY/TRPF PKD2s, and TRPML (Ramsey et al., 2006). While Group I (GI) gathers members of TRPC, TRPM, TRPS, TRPV, TRPVL, TRPA, and TRPN subfamilies, Group II (GII) is composed of members of the TRPML, PKD2, and TRPY/TRPF channels (Venkatachalam 2007, Himmel 2020; Himmel and Cox, 2020).

In general, TRP channels share poor cation selectivity and a loose sequence similarity (Ramsey et al., 2006). Cataloged as polymodal ion channels, they have the ability to integrate multiple stimuli (e.g., chemical, mechanical, electrical, and thermal) to promote channel opening. Such polymodality observed in TRPs has been explained in terms of allosteric interactions (Brauchi et al., 2004; Latorre et al., 2009). Different lines of research have shown that the different sensor modules couple to each other and to the channel pore, modulating permeation (Hui et al., 2003; Castillo et al., 2018; Zubcevic et al., 2019; Zhao et al., 2020; Yang et al., 2018; Yang et al., 2020).

Structural data revealed that TRP channels share the general architecture of voltage-gated ion channels (VGIC) (Kasimova et al., 2016; Cheng, 2018; Cao, 2020). They assemble as domain-swapped tetramers, with monomers containing six transmembrane segments (i.e.,
TM1-TM6) flanked by cytoplasmic N- and C-terminal domains. The transmembrane helices TM5 and TM6, together with the section between them, give shape to the conductive pore (Ramsey et al., 2006; Liao et al., 2013). The first four transmembrane helices (TM1-TM4) and the intracellular domains have been described as regulatory regions as they provide binding sites for agonists and cofactors (Steinberg et al., 2014; Voolstra and Huber, 2014; Cao, 2020). The TM1 and TM4 transmembrane region shows fundamental differences with known VGICs, where the absence of charged residues within the transmembrane region should be underscored (Palovcak et al., 2015; Cao, 2020). It has been shown that TRPMs and TRPVs are about ten-fold less voltage dependent compared to voltage-gated potassium channels. The modest voltage dependence observed in TRP channels is likely supported by residues located in the pore region (Liu et al., 2009; Yang et al., 2020). The TM1 and TM4 transmembrane region hosts binding pockets for ligands in all the different TRP channel subtypes, serving as a ligand-binding domain (LBD) in TRPs (Steinberg et al. 2014; Huffer et al., 2020). Nevertheless, this LBD has historically been referred to as the voltage-sensing-like domain (VSLD).

Regardless of the considerable variability in sequence similarity and physiological function, it is clear that TRP channels are closely related, especially those within Group I (Kadowaki, 2015; Peng et al., 2015; Arias-Darraz et al., 2015). The structural understanding of TRP conformational and functional behavior has been strengthened by technical advances allowing the recent release of a large set of high-resolution cryo-electron microscopy (cryo-EM) structures (Cao, 2020; Samanta et al., 2019). Led by the structure of TRPV1, structures have recently become available for at least one member of each subfamily, including crTRP1 from the green algae *Chlamydomonas reinhardtii* (McGoldrick et al., 2019; Cao, 2020). Together with advances in structural biology, intense research during recent years has been focused into understanding the molecular mechanisms supporting TRP activation and regulation (Hofmann et al., 2017; Yang and Zheng, 2017; Yang et al., 2018; Singh, et al., 2018; Hilton et al., 2019; Zhang et al., 2019; Zubcevic et al., 2019; Zhao et al., 2020, Nadezhdin et al, 2021; Nadezhdin et al., 2021a). Although a consistent picture accounting
for general principles governing TRP channels’ mechanics is still missing, several structural
features of great importance have been identified. Among these are the proximal N- and C-
terminal domains flanking the transmembrane region, namely the pre-TM1(also called pre-
S1) and the TRP domain helix (TDh) respectively. These elements, seemingly related to the
integration of molecular mechanics during activation, are present in all members of GI-
TRPs and absent in GII-TRPs (McGoldrick et al., 2019; Zhao et al., 2020; Nadezhdin et al.,
2021).

Here we studied TRP channel proteins from a large variety of organisms aiming to
understand how they are evolutionarily related, and to find conserved structural elements.
We obtained a well-resolved phylogeny, providing a snapshot of the TRP gene family
duplicative history, and a phylogenetic framework to understand the evolution of
structural and functional attributes present in the TRP gene family. We also found 12
conserved and non-contiguous amino acids (W F Φ G Φ Φ Φ N L I A W) present in all TRP
channels from GI-TRPs that we interpreted as a fingerprint.

Moreover, we discovered strong conservation unique to each TRP subtype. The amino acid
conservation can be traced down to TRP channels from unicellular organisms, suggesting a
robust architectural design. In addition, we identified a group of aromatic residues facing
the core of the LBD (i.e., TM 1-4) in all TRP subtypes. In agreement with our phylogenetic
reconstruction, this aromatic core (AC) can be observed in the unicellular crTRP1, it is
absent in VGICs, and present in a rudimentary form in TPCs. Our structural analyses
suggest that TRP channel specialization could be oriented around inter-subunit
interactions between aromatic core residues of one subunit with fingerprint residues at the
selectivity filter of a neighboring subunit. Overall, our results suggest that TRP channel
specialization has been built around the connectivity of heavily conserved distant residues
that are located at critical sites within the structure.
RESULTS

Phylogenetic diversification of the TRP gene family and the phylogenetic position of unicellular TRPs.

Over millennia, the diversification of TRP channels produced a variety of lineages that we currently identify as subfamilies in an established nomenclature. The understanding of phylogenetic relationships among TRPs has been subject to an intense debate, and different phylogenetic hypotheses have been proposed (Clapham et al. 2001; Clapham 2003; Sidi et al. 2003; Yu and Catterall 2004; Yu et al. 2005; Montell 2005; Yu et al. 2005; Venkatachalam and Montell 2007; Latorre et al. 2009; Nilius and Owsianik 2011; Arias-Darraz et al. 2015; Ferreira et al. 2015; Peng et al. 2015; Eriksson et al. 2018; Kozma et al. 2018; Himmel et al. 2020; Himmel and Cox, 2020; Hsiao et al. 2021). The diversity reported in phylogenetic topologies so far can be explained by the differences in taxonomic sampling, and to the fact that not all studies include all subfamilies and/or outgroups. To advance our understanding of such a process of diversification, we first performed a phylogenetic analysis, overcoming the caveats mentioned above by including representative members of each reported TRP subfamily in addition to outgroups (i.e., Kv and Nav channels).

Our gene phylogeny is well resolved and provides a snapshot of the TRP gene duplicative history (Figure 1; Figure 1 – figure supplement 1). To validate this phylogeny, we repeated the phylogenetic analysis ten times, observing that the evolutionary relationships among the main TRP lineages were consistent, with only negligible variation in the likelihood values. We not only recovered the monophyly (i.e., the single evolutionary origin of the TRP gene family) with strong support of 98%, but also the diversity of TRP channels into two main groups: 1) GI-TRPs, a clade containing TRPA1, TRPV, TRPV1, TRPC, TRPGamma, TRPN1, TRPY/TRPF, TRPM and TRPS channels, and 2) GII-TRPs, a clade containing PKD2 and MCLN channels (Figure 1). We recovered four well-supported clades for GI-TRPs: A) TRPV/TRPV1/TRA1, B) TRPC/TRPGamma/TRPN1, C) TRPY/TRPF and D) TRPM/TRPS. Among these, we recovered the sister group relationship between the
TRPV/TRPV2/TRPA1 and TRPC/TRPGamma/TRPN1 clades with strong support (Figure 1). Moreover, the TRPY/TRPF group was recovered sister to the latter (Figure 1). The TRPM/TRPS clade was recovered sister to all other members of GI-TRPs (Figure 1).

Next, we performed phylogenetic analyses to investigate the phylogenetic position of TRP channels from unicellular organisms, as they are most divergent. According to our analyses, a TRP sequence from *Coccomyxa subellipsoidea* was recovered sister to the TRPV/TRPV2/TRPA1 clade, while a TRP sequence from *Chlamydomonas reinhardtii* was recovered sister to the TRPM/TRPS clade (Figure 1 - figure supplement 2 and 3). As reported before (Arias-Darraz et al. 2015), all other TRP sequences from unicellular organisms considered in this work are more evolutionarily related to the GII-TRP clade (Figure 1- figure supplement 2).

**Multiple sequence alignments identify a discrete set of highly conserved residues.**

We then focused our attention on the transmembrane regions and the immediate flanking segments that are common to all GI-TRPs. In particular, we analyzed the protein segment containing the pre-TM1 region, the transmembrane region, and the TRP domain helix (TDh). We enlarged the set of sequences used for phylogenetic analysis and constructed a multiple sequence alignment (MSA) using a set of bona fide TRP orthologs. The final set contains 861 sequences from vertebrates pulled from the orthologous matrix project (Altenhoff et al., 2021), plus 620 non-redundant TRP channel sequences gathered from the UniProt database, containing representatives of invertebrates and unicellular organisms (Figure 1 - figure supplement 4). Amino acid frequency histograms were used to define conserved regions and structural features were mapped according to the structural data available (Figure 2a). In agreement with previous structural alignment studies (Huffer et al., 2020), we observed that TM regions are well defined and conserved (Figure 2a). Moreover, we observed that gaps detected in the alignment are mostly confined to linkers in between TM segments. This includes linkers, pre-TM1 to TM1, TM1-TM2, TM3-TM4, and
the outer pore regions surrounding the selectivity filter. In contrast, gaps are almost non-existent at the intracellular linker between TM4-TM5, the elbow connecting TM6 with the TDh, and the TDh itself (Figure 2a). The pattern of these gaps in the loops connecting transmembrane segments seems to be a good predictor for the subfamily grouping (Figure 2- figure supplement 1). This is the case of the longer linker between preTM1 and TM1 that is characteristic of TRPM channels, or the extended TM3-TM4 loop seen in channels from the TRPC family (Figure 2- figure supplement 1). Moreover, the region surrounding the selectivity filter shows diversification. While TRPV, TRPC, and TRPN channels display a larger linker between TM5 and the selectivity filter/pore helix, TRPMs show a larger linker region between the selectivity filter/pore helix and TM6 (Figure 2- figure supplement 1).

A thorough analysis of the distribution of amino acids identified a discrete set of highly conserved residues (identity >90%) common to the TRP superfamily (Figure 2 a and b). 69% of TRP channels analyzed have 12 conserved and non-contiguous amino acids we interpreted as an amino acid signature or fingerprint (W Φ G Φ Φ Φ Φ N L I A W; where Φ could be either Phe or Tyr; Figure 2c). We observed that 95% of the TRP channels contained at least 9 of these conserved side chains (Figure 2c; Table 1). In contrast, the bacterial potassium channel KvAP displays five conserved residues in relatively similar positions while Kv1.2 exhibits only three (Table 1). Given the fact that the overall dynamics and mechanism of action of the S1-S4 domain in voltage-gated channels is different from that of TRP channels it is unclear for us whether the apparent conservation of TRP fingerprint residues is valid for the case of voltage-gated channels. Further work would be needed to trace the evolutionary importance of these particular residues in the context of the extended family of 6TM ion channels.

This fingerprint was used to compare the conservation between TRP channel subtypes (Figure 2c and Figure 2- figure supplement 3). While TRPVs, TRPCs and TRPMs conserve the complete set of 12 residues (W F Φ G Φ F Φ N L I A W), TRPNs show 11 (W F Y G Φ Y Φ N L V A W), TRPA1 10 (W F Y G F F N L I G R), and the more distantly related TRPY only 8 residues. The differences observed in TRPA1 can be mapped to the TDh. The latter is
considered an important modulator of TRP gating and although different in sequence in TRPA1 channels when compared to other GI-TRPs, it is structurally equivalent (Paulsen et al., 2015). Thus, our results suggest that such specific specialization in TRPA originated after the divergence from the common ancestor shared with TRPVs and is unlikely to be found outside this group. Within the unicellular algae set crTRP1, which shares a common ancestor with the TRPM family, shows only 7 fingerprint residues that increases up to 9 for the case of csTRP2 that shares a common ancestor with the clade containing TRPV, TRPVL, and TRPA1 channels (Figure 2c; Table 1).

There are some highly conserved residues that were not considered in our analysis because they score just below the 90% threshold. These include a highly conserved glycine (86.3%) commonly found at the selectivity filter, a phenylalanine (85.5%) at the beginning segment of TM5, and an aspartic acid (85.2%) localized at the end of TM4-TM5 linker. The latter two residues are in close proximity to L678 (according to rTRPV1 numbering) a sidechain associated to channel response to both agonist and pH in different TRPs channels (Boukalova et al., 2010; Du et al., 2009; Kasimova et al., 2017; Klausen et al., 2014). These results were corroborated by a hidden markov model (HMM) analysis using the same dataset (Figure 2 - figure supplement 2).

Several groups have suggested an evolutionary relationship between Two-Pore Channels (TPC) and TRP channels (Clapham and Garbers, 2005; Galione, 2010). All three identified TPCs are thought to be asymmetric (Penny et al., 2016; Kintzer and Stroud, 2017). In support of the argument of asymmetry, while one domain (D1) shows 6 coincidences with the TRP fingerprint, the other domain (D2) is more similar to eukaryotic voltage-gated sodium channels (Nav) with only 5 coincidences (Table 1). Interestingly, the monomeric bacterial channel NabAB (Payande et al., 2011) shares seven of these eleven signature residues, and the mammalian Nav1.4 exhibits only five of these residues in domain III, which holds the highest number of hits compared to other domains (Table 1). Thus, in our analysis, TPC and bacterial Nav channels exhibit the larger score of conservation at the fingerprint outside the TRP family. Overall, phylogenetic and primary sequence analyses
provide strong support for a fingerprint in GI-TRP channels that is composed of 7 to 11 non-contiguous residues (W F Φ G Φ Φ N L I A W).

Sequence conservation highlights structural features.

To visualize the position of these fingerprint residues, TRP channel structures from the different families were compared. We first display the frequency of amino acid coincidences and highlight the fingerprint in the context of TRPV1 (Figure 3a and b). By doing this structural mapping, we identified three well defined clusters (hereafter referred to as patches) of fingerprint residues that are present in representative channels of the different families, including crTRP1 from green algae (Figure 3c). By extending our structural alignment to match with recently published results (Huffer et al. 2020), we not only confirmed that our sequence alignments are in full agreement with reported structural alignments, but we observed that the three-dimensional arrangement of the fingerprint is a robust feature among TRPs (Figure 3 - figure supplement 1). To reconcile the different sequence datasets used in this work (Figure 1 - figure supplement 2), rTRPV1 numbering was used throughout the manuscript to identify amino acid positions.

The first patch (P1) gathers several residues from a hotspot that has been historically linked to channel modulation. It is composed of side chains from pre-TM1 (Trp426 [95.3%]), the TM1 (Phe434 [95.5%]), the TM4-TM5 linker (Gly563 [98.0%]) and the TDh (Trp697 [93.1%]) (Figure 3b; Table 1). Initially proposed as critical for TRPV1 channel activation (Gregorio-Teruel et al., 2014), Glycine 563 at the linker and Tryptophane 697 at the TDh have been reported as a common theme in TRPs (Table 2). In this context, the TDh seems to operate as an integrator between various functional elements, receiving information from lipids (such as PIP2), the TM4-TM5 linker reporting changes occurring at the transmembrane region, and the coupling domain (CD) composed of the pre-TM1 helix and a helix-loop-helix (HLH) motif. This CD in TRPs is thought to participate in the functional association between the cytosolic and the transmembrane domain. The high conservation of Phe434 is underscored by the importance of the CD array that integrates with critical cytoplasmic features (Garcia-Elias et al., 2015; Romero-Romero et al., 2017;
Hofmann et al., 2017; Yang et al., 2018; Hilton et al., 2019; Yuan, 2019; Zubcevic et al., 2019; Cao, 2020).

The second patch (P2) is located at the selectivity filter region and is composed of three phenyl group residues, i.e., Phe591 [Φ 96.6%], Phe638 [Φ 96.8%], and Tyr666 [Φ 98.6%], located at both the TM6 helix and the pore helix (Figure 3b; Table 1). A glycine residue that forms part of the selectivity filter/upper constriction in TRPs is also highly conserved in all subtypes (85.2%) but slightly off to the 90% threshold (Figure 2). The high conservation of these three phenyl residues located within the selectivity filter and pore helix contrasts with the otherwise high variability observed in this region (i.e. turret and re-entry linker) (Figure 2 - figure supplement 1).

A third patch (P3) is localized at the lower portion of the pore and is composed of a well-studied set of residues forming the lower gate (Asn676 [93.7%], Ile679 [89.7%], Ala680 [95.3%]) as well as Leu678 [95.3%] that is facing the interface between TM5 and TM6 (Palovcak et al., 2015) (Figure 3b; Table 1).

Finally, we identified a prevalent aromatic side chain (Tyr441 [76.5% Tyr + 14.4% Phe]), located at the middle of the TM1 helix (Figure 3b; Table 1). Mutations in that position have been reported deleterious for TRPV1 channel function (Boukalova et al., 2013). This residue was not observed in close proximity to any other fingerprint residue (Figure 3 b and c).

From the structural data, the most conserved interaction amongst fingerprint residues is between the Gly563 at the TM4-TM5 linker and Trp697 at the TDh. This observation is further supported by our evolutionary coupling analysis (ECA), showing that the highest score for putative interactions are precisely those established between the TM4-TM5 linker and the TDh (Figure 3 - figure supplement 2). ECA also links the lower portion of TM2 to the end of the TDh. Structurally, such interaction cannot be easily explained by direct contact between residues from TM2 and TDh. Thus, it follows that the high covariance score between TM2 and the TDh could involve an additional linker molecule such as PIP2.
or other lipid binding to this region (Poblete et al., 2014; Yin et al., 2017; Yazici et al., 2021; Hughes et al., 2018). Interactions between the channel and membrane lipids at the cytosol-membrane interface are emerging as a common theme in TRPs. Under this view, different parts of the CD/TDh coupling mechanism are tuned by the differences in binding of membrane lipids and/or canonical ligands (reviewed in Zubcevic, 2020).

Connectivity between the signature residues.

Series of independent studies had reported that mutations of residues that form part of the fingerprint -or the connecting side chains- modify or impairs channel activity, underscoring their importance in maintaining proper channel activity (Table 2). Therefore, we analyzed the connectivity between the signature residues at the different patches using simple MSA statistics and DCA, in addition to a structural analysis. To this end, we first parsed the whole alignment by selecting only positions of high frequency (Figure 3 – figure supplement 3) to build distance matrices for the different structures that were included in our set of sequences (138 individual structures). By averaging these individual matrices, we obtained the mean distance between the highly conserved residues (i.e., residues of high frequency that are present in all structures analyzed). Mean distance between fingerprint residues of the same patch is low, suggesting proximity as observed in the exemplary structures (Figure 3 c and d; Figure 3 – figure supplement 4). Moreover, the normalized variance of these pairwise distances is also low (Figure 3e, purple squares). This suggests that fingerprint residues on each patch remain in close proximity with each other, regardless of the variation in conformations rendered during experimental derivation. In contrast, sections of high pairwise variability are also detected. This is the case of pre-TM1, where the variance is higher for the distances with the linkers TM2-TM3 and TM4-TM5 (Figure 3e, darker regions). As reported abundantly in literature, the lower gate transition to the open state involves the motion of the lower portion of the TM6 (Susankova et al. 2007; García, Oseguera, and Emir 2009; Cao et al. 2013). Accordingly, we can observe the large variability of distances between TM5 and TM6, an indication of the multiple conformations we included in our statistical analysis (Figure 3e, darker regions).
Moreover, the TM1-TM4 region displays larger variability among close residues (Figure 3e). Notably, the amino acids belonging to the different patches are consistently close to regions of larger variability (Figure 3e; dark spots next to purple squares). It is important to note that our analysis surveys the landscape of possibilities observed out of a diverse collection of structural arrangements, therefore should not be over-interpreted as a direct proxy for mobility.

A conserved aromatic core at the transmembrane region of TRPs.

Four phenyl groups (i.e., Phe/Tyr) were identified in specific positions within the transmembrane domain in more than 90% of the sequences (Figure 4a and b). Further inspection of structural data showed that these residues are part of a larger cluster of aromatic residues facing the center of the four-helix bundle formed by TM1-TM4, that are common to all TRP channel structures. A group of five to eight inward-facing aromatic residues belonging to the LBD domain (including at least one signature residue from P1), appear to form an aromatic core (AC) (Figure 4; Figure 4 - figure supplement 1). In contrast, voltage-gated potassium and sodium channels occupy these positions with charged amino acids forming salt bridges, short aliphatic side chains, or aromatics located at the membrane-water interface (Figure 4a; Figure 4 - figure supplement 1). Although the overall three-dimensional shape of the AC varies (Figure 4c), it is present in all TRP subtypes, connecting three to four helices from within the LBD, suggesting they could serve as a scaffold that stitches together the whole domain.

This putative AC emerges as a common theme present from crTRP1 to TRPV1-4 (Figure 4b; Figure 4 - figure supplement 1). The most dramatic case is observed in the TRPV1-4 clade, exhibiting both the highest number of aromatics (8) and the most ordered stack (Figures 4b and 4c). On the other hand, TRPA1 shows only 5 aromatics forming the AC without a clear stacking, as observed also in the other subfamilies (Figure 4b). Although the residues in TRPA1 are still in close contact, they do not form the compact stacking observed in their
sister group, indicating that this might correspond to a specialization exclusive for TRPV
channels (Figure 4b and 4c). In contrast to these examples, crTRP1 does not clearly form a compact-extended AC.

The presence of these conserved aromatics next to ligand binding sites, or even forming part of them, suggests a mechanism in which the AC acts by imparting rigidity to the region and functionally linking the transmembrane helices 1-4, facilitating the translation of mechanical force from the ligand binding sites to the coupling domains that connect multiple parts of the channel including the TDh, cytoplasmic CD, and pre-TM1.

Consistent with our hypothesis, we observed a large variability in the distances between residues located within the TM1-TM4 region (Figure 4 - figure supplement 2a). However, minimal or no change in mean distance was observed among the conserved aromatics (Figure 4 - figure supplement 2a and b). This suggests to us that their role might not be related to being part of a switch but rather stitching channel machinery in place.

The analysis of the most represented subfamilies (TRPV, TRPM, and TRPC) shows that the connectivity and association between amino acids in this region has been shaped differently (Figure 4 - figure supplement 2c-h). Considering the role of the region in ligand binding, our sequence and structure-based analyses support the notion of progressive structural transitions enabling the functional specialization of the LBD within each different subfamily.

Finally, consistent with the notion that TPC channels are close relatives of the TRP family, the presence of a group of 3 aromatics facing inside the VSLD is conserved in TPC’s domain 1 and absent in domain 2 that is devoid of an extended aromatic network, resembling Nav channels (analyzed on mTPC1, PDB:6C96; NaV1.4, PDB:6AGF). Unlike for TRPs, the AC in TPC channels looks less cohesive, or “disconnected” (Figure 4c).

**Conserved residues at the interaction between subunits.**
TRP channels show a domain-swapped configuration (Liao et al, 2013). That is, the VSLD/LBD of one subunit appears in close contact with its pore domain (PD) through the protein backbone and, at the same time, in close contact with the PD of the neighboring subunit. The coupling between the LBD and the pore domain from different subunits of the tetramer is one topic that has been poorly studied in TRP channels and not well understood in VGIC (Carvalho-de-Souza and Bezanilla, 2019; Shem-Ad et al., 2013; Neale et al., 2003).

By inspecting the structural data, we found a conserved interaction formed by a residue at the middle of TM4 and a residue that is consistently preceding a signature residue in TM5 (Phe591 in rTRPV1; Figure 5a). Although the nature of the interaction varies, it is present in all surveyed structures (Figure 5b; Figure 5- figure supplement 1a). Such interaction would put in direct contact TM helices 4 and 5 from different subunits (Figure 5- figure supplement 1). For the case of TRPV channels, this interaction might support long-range communication between patches P1 and P2 from different subunits via the aromatics running alongside the LBD. Such putative connection would be absent in subfamilies not presenting a stacked arrangement such as TRPC (Figure 5- figure supplement 1b). The high conservation of these residues and their relative positions within the structure suggests a common mechanism in TRPs where the selectivity filter (P2) would be functionally connected to the ligand binding site located on a neighboring transmembrane region.

Discussion

Here we studied the process of diversification of TRP channels and the conservation arising from the residue frequency distribution in the transmembrane region and put them in a structural context. A highly conserved set of residues are located and grouped at strategic positions within the channel’s transmembrane region, defining a fingerprint for CI-TRPs.

Phylogenetic relationships among transient receptor potential channels.

To the best of our knowledge, the more complete TRP phylogeny was recently proposed by Himmel and Cox (2020). However, their phylogenetic hypothesis is not fully resolved as it
presents two trichotomies: the first is among TRPY/TRPF, MCLN/PKD2, and all other TRP lineages, whereas the second is among TRPA, TRPV/TRPVL and TRPC/TRPN/TRPM/TRPS clades. Also recently, Hsiao et al. (2021) performed an effort to define the sister group relationship among TRPs in animals, testing different methodological approaches and taxonomic samplings. However, most of their phylogenetic analyses recovered unresolved trees in addition to not including all TRP lineages (e.g. TRPY/TRPF) (Hsiao et al. 2021). In our case, the proposed phylogenetic hypothesis is well supported and resolves the sister group relationships among the main groups of TRP channels. Similar to other studies, we recovered the PKD2/MCLN clade sister to all other TRP lineages (Montell 2005; Venkatachalam and Montell 2007; Kozma et al. 2018), however, it is different from the proposed scenario by Himmel and Cox (2020) in which they suggest that the TRPY/TRPF clade could belong to the group that includes PKD2 and MCLN.

Most studies do not incorporate all of the TRP lineages and outgroups, making it difficult to perform direct comparisons. Regardless, our phylogenetic arrangement shows some differences with the results reported in the literature. For example, in some studies the clade containing TRPM sequences has been recovered sister to the TRPC/TRPN1 clade (Montell 2005; Venkatachalam and Montell 2007; Hsiao et al. 2021). Other studies suggest that TRPC is sister to TRPM, and the clade containing TRPN1 sequences is the sister group of the TRPC/TRPM clade (Arias-Darraz et al. 2015; Ferreira et al. 2015; Kozma et al. 2018). There are also studies that show the sister group relationship between TRPV and TRPC and this clade sister to TRPM (Clapham et al. 2001). The TRPA1 gene lineage has been recovered sister to the TRPM/TRPC clade (Clapham et al. 2005), to the TRPN1 clade (Latorre et al. 2009; Nilius and Owsianik 2011; Eriksson et al. 2018), to the TRPV clade (Ferreira et al. 2015; Peng et al. 2015; Kozma et al. 2018), and to the TRPV/TRPC/TRPM/TRPN1 clade (Montell 2005; Venkatachalam and Montell 2007; Hsiao et al. 2021).

Here we want to sound a note of caution about the use of TRP names in a taxon-specific manner because represent the common ground of this debate. The way we should "assign
names” to genes in different taxonomic groups must be based on our understanding of the
duplicate history of the group of genes we are interested in (Gabaldon 2008; Altenhoff et al.
2018). To do this, we need to perform studies including a broad and balanced taxonomic
sampling and appropriate outgroups, where the reconciliation of the gene tree with the
species tree plays a fundamental role (Goodman et al. 1979). In summary, we present a
phylogenetic hypothesis for the sister group relationships of TRP channels that was
inferred including all members of the gene family and outgroups. We believe that our tree
topology can serve to understand the diversity and speciation of structural attributes
present in the different subfamilies of TRP channels (Figure 6).

A robust signature for TRP channels.

A consistent architectural picture is needed to extend and generalize the multiple
observations provided by structural biology. In this context, we report here a highly
conserved set of residues are located at strategic positions. From the identified signature
residues, those that belong to P1 and P3 have been recurrently studied in the literature. In
particular, residues from P1 have been proposed critical to support the interaction
between the TDh and the TM4-TM5 linker acting as an allosteric integrator (Taberner et al.,
2014; Sierra-Valdez et al., 2018; Romero-Romero et al., 2017; Gregorio-Teruel et al., 2014;
Zhao et al 2020). On the other hand, residues from P3 have been associated with the
function of the lower gate of TRP channels. In particular, to participate in stabilizing the
transition between α- and π-helix types during opening (Palovcak et al., 2015; Kasimova et
al., 2017; Kasimova et al., 2018). This observation -the critical role of amino acids in P1 and
P2- holds true under the light of the structural analysis presented in here.

In contrast, the set of aromatics forming P2 are not described in the literature as such.
Nevertheless, previous studies showed the importance of the residues composing this
patch. Phe591 is described as forming part of the higher, wider region of the capsaicin
pocket binding in TRPV1 channels (Elokely et al., 2015) and mutations at Tyr638 and
Tyr666 have obvious effects on channel activity. Specifically, Tyr666Ala is not functional in TRPV1 (Susankova et al., 2007), the equivalent mutant Tyr661Cys in TRPV3 is not activated by temperature, but still responds to agonist (Grandl et al., 2008), and the T702L equivalent mutant in TRPV4 has significantly reduced responses to agonists, temperature, and mechanical stimulation (Klausen et al., 2014). Moreover, mutations at equivalent positions to rTRPV1 Tyr638 in channels from different subfamilies elicit a wide range of effects from gain-of-function to dominant-negative phenotypes (Munns et al., 2015; Kitaguchi and Swartz, 2005; Bidaux et al., 2015; Strübing et al., 2003; Chandrabalan et al., 2019; Vinayagam et al., 2018; Paulsen et al., 2015). Noteworthy, Bidaux et al., (Bidaux et al., 2015) demonstrated that mutations at this position (Tyr908 in TRPM8) to Ala or Trp abolish the response to temperature and menthol but not to icilin, yet mutation to Phe keeps the channel fully functional. We observed that the three aromatics forming P2 always need a non-conserved fourth hydrophobic residue to connect them all. By comparing apo and ligand-bound structures, P2 consistently seems to translate as if it were a near-rigid-body, moving as a whole compact motif. The function of this patch might be associated with the communication between the ligand binding domain and the state of selectivity filter in an inter-subunit fashion.

The interaction between signature residues and the aromatic core.

Two of the residues in the P1 patch have been proposed as fundamental to the interaction between the TDh and the TM4-TM5 linker (i.e., Gly 563 and Trp 697 in TRPV1), acting as an allosteric integrator (Taberner et al., 2014; Sierra-Valdez et al., 2018; Romero-Romero et al., 2017; Gregorio-Teruel et al., 2014; Zhao et al, 2020) (Table 2). A third residue in P1 corresponds to a conserved aromatic at the lower end of TM1 (i.e, Phe 441 in TRPV1) that is always connected to the other elements of the patch. Separated by 2 intermediate interactions, TRPM8 displays the largest distance between these two components of patch P1 (i.e., TM1 and TM4-TM5 linker/TDh). This larger distance is consistent among the TRPM family as depicted in the distance histograms. Notably, the gap observed between these
two components has been associated to the binding site of menthol in TRPM8 channels (Bandell et al., 2006; Malkia et al., 2009). Moreover, the menthol analog WS-12 binds to Tyr745 (TM1), and sits close to Tyr1004 (TDh), seemingly reinforcing the connection between the LBD and the end of the TDh in TRPM8 (Yin et al., 2019).

There is only one signature residue that is consistently forming part of the aromatic core (i.e., Tyr441 in TRPV1). Mutations to Ser of Tyr441 in TRPV1 generates non-functional channels (Boukalova et al., 2013). The equivalent residue in TRPM3 channels presents impairments in response to agonists when mutated from Tyr to Thr, and a Tyr to His mutation in TRPM8 channels has been shown critical to both menthol activation and the inhibitory effect of the small molecule SKF96365 (Bandell et al., 2006; Malkia et al., 2009). This coincided with similar phenotypes observed by mutating other residues belonging to the AC (Table 3). This is the case of Tyr444 that in TRPV1 generates non-functional channels (Boukalova et al., 2013) and the double mutation Y885T / W982R in TRPM8 that presented altered responses to agonists and temperature (Held et al. 2018).

When comparing the disposition of the aromatic core in detail, it appears obvious they have different distributions in the different TRP channel families. However, certain generalizations can be drawn. The AC appears associated with P2 at the selectivity filter and disconnected from P1 in most families. Notably, in TRPVs the AC extends down to reach the lower aromatic residues of P1 making a direct connection between the AC and the TDh/TM4-TM5 linker. Hence, a P2-AC-P1 “continuum” can be observed in the TRPV subgroup. Interestingly, from the structures obtained in the presence of agonists we noticed that these agonists bridge the AC and P1 (e.g., WS-12 in TRPM8 and GFB-9289 in TRPC4; PDBIDs 6NR2 and 7B16, respectively; Yin et al., 2019; Vinayagam et al., 2020). The disposition of the components of P1 (i.e., preTM1; TM1; TM4-TM5 linker; TDh) and differences in the AC suggest similar but non-identical coupling strategies among GI-TRPs. This is reinforced by the diversity of mean distance variability obtained for the different families within the TM1-TM4 region.
The presence of an aromatic core, connecting the transmembrane helices of the ligand binding pocket and communicating critical modulatory regions that are far apart in the structure (e.g., the selectivity filter and the TDh) somewhat remind us of the case of Cys-loop receptors where an intra-membrane aromatic network contributes to the assembly and function of the receptor via interactions of both nearby and far apart residues (Haeger et al. 2009; Tang and Lummis 2018). The AC in TRPs appears as a modular connector that has been subject to variations throughout evolution and suggests a common aspect of TRP channel mechanics that remains largely unexplored. Our line of reasoning also suggests that the observed rearrangements of the selectivity filter during activation are somewhat linked to inter-subunit interactions, likely modulated by the AC network. At the same time, it implies that certain ligands might have the ability to modulate the conformation of the selectivity filter, and by extension the extracellular linkers, without the need of an open gate conformation.

Temperature-dependent gating from a phylogenetic and structural perspective.

TRP channel’s ability to respond to temperature is a property that has been reported in 11 TRP channels (TRPV1-4, TRPA1, TRPM2, TRPM3, TRPM4, TRPM5, TRPM8, and TRPC5 (Caterina et al. 1997; Saito and Shingai, 2006; Saito et al. 2011; Ferreira et al. 2015; Saito and Tominaga, 2015; Castillo et al. 2018). Despite the strong conservation of the phenotype across GI-TRPs, we have failed to find a conserved domain common to all temperature-sensitive TRP channels within the region analyzed in the present study. Considering the phylogenetic relationships among TRPs, the phenotype distribution suggests a scenario involving multiple gain and/or losses of the ability to respond to temperature. The temperature-sensitive crTRP1 (Arias et al. 2015; McGoldrick et al. 2019), from the unicellular algae Chlamydomonas reinhardtii, shares a common ancestor with the TRPM subfamily, that includes at least 5 out of 8 members displaying different degrees of temperature sensitivity. The feature is apparently lost in TRPCs, with the notable exception of TRPC5 (Zimmermann et al. 2011). Would be hard at this time to elaborate whether the
origin of temperature sensitivity in TRPC5 channels represent a gain or loss of a functional
trait. In contrast, temperature-dependent gating is strongly conserved in the TRPV group,
with the exception of calcium-selective TRPV5 and TRPV6. The phenotype was apparently
preserved in TRPA1 during a process of specialization that involved multiple sequence
changes in the so called “allosteric nexus” region (Paulsen et al., 2015), while preserving the
core structural features defining GI-TRP channels presented in here.

Recent structural work (Nadezhdin et al. 2021; Kwon et al. 2021), anisotropic thermal diffusion
calculations (Diaz-Franulic et al., 2016), and a combination of patch clamp fluorometry,
mutagenesis and molecular modeling (Yang et al., 2018) suggest that the temperature-
dependent transition may initiate close to the intracellular water-lipid boundary and propagate
through a conformational wave. Interestingly, this collective molecular motion is enabled by a
network of interactions involving fingerprint residues Trp426 and Phe434, which connect to the
transmembrane helices via the AC and the TM4-TM5 linker/TDh interface. The conserved
residues and interactions highlighted in the present work suggest new and exciting mutagenesis
experiments that could potentially dissect the temperature-driven conformational wave and
thus shed light on the microscopic mechanism of heat activation.

METHODS

Amino Acid sequences, alignments, and phylogenetic analyses. To advance our
understanding of the sister group relationships among TRPs, we retrieve amino acid
sequences from the National Center for Biotechnology Information (NCBI) (PMID: 29140470) corresponding to TRPVs, TRPVL, TRPA1, TRPC, TRPgamma, TRPN1, TRPY/TRPF, TRPM, TRPS, PKD2, and MCLNs lineages. In most cases, we included
representative species of vertebrates (TRPA1, TRPVs, TRPCs, TRPN1, TRPMs, PKD2s, and
MCLNs). However, in the case in which the TRP lineages are not present in vertebrates, we
included representative species of the groups in which the TRP channel is present. Thus,
TRPVL included cnidarians and annelids, TRPgamma included insects, arachnids, and
merostomes, TRPS included nematodes, chordates, arachnids, chilopods, priapulids,
cephalopods, bivalves, and tardigrades, TRPY/TRPF had several species of fungi. Accession numbers and details about the taxonomic sampling are available in Supplementary Table S1. Amino acid sequences were aligned using MAFFT v.7 (Katoh et al., 2017) allowing the program to choose the alignment strategy (FFT-NS-i). We used the proposed model tool of IQ-Tree v.1.6.12 (Minh et al., 2020) to select the best-fitting model of amino acid substitution (JTT+F+I+G4). We used the maximum likelihood method to obtain the best tree using the program IQ-Tree v1.6.12 (Minh et al., 2020). We performed five tree searches in which the initial gene tree was provided by ourselves, which was previously estimated using IQ-Tree v.1.6.12 (Minh et al., 2020). We also carried out five additional analyses in which we performed more exhaustive tree searches by modifying the strength of the perturbation (-pers) from 0.5 (default value) to 0.9 and the number of unsuccessful iterations to stop (-nstop) from 100 (default value) to 500. The tree with the highest likelihood score was chosen. We assessed support for the nodes using the ultrafast bootstrap routine as implemented in IQ-Tree v1.6.12 (PMID: 29077904). Potassium voltage-gated channel subfamily A member 2 (KCNA2) and Sodium Voltage-Gated Channel Alpha Subunit 8 (SCN8A) amino acid sequences from mammals were included as an outgroup (Supplementary Table S1). Our next step was to retrieve TRP amino acid sequences from *Chlamydomonas reinhardtii*, *Volvox carteri*, *Coccomyxa subellipsoidea*, *Micromonas pusilla*, *Dictyostelium discoideum*, *Dictyostelium purpureum*, *Leishmania infantum*, *Leishmaniamajor*, *Leishmania mexicana*, *Paramecium tetraurelia*, and *Trypnosoma cruzi*. To do so, the transmembrane regions corresponding to TRPA1, TRPC1, TRPC3-7, TRPM1-8, TRPML1-3, TRPP1-3, and TRPV1-6 from human (*Homo sapiens*), TRPC2 from the house mouse (*Mus musculus*), NompC from the fruit fly (*Drosophila melanogaster*) and TRPY1 from the brewing yeast (*Saccharomyces cerevisiae*) were used as queries in blastp searches (Altschul et al., 1990) against the proteomes of above-mentioned species. Putative channels were selected based on the frequency of hits to the query sequences relative to human, fruit fly, and yeast. This was followed by reciprocal blastp searches (E-value < 1e), and a final inspection for the presence of the TRP domain. To investigate the phylogenetic position of these candidate TRP channels we
aligned them with the sequences previously sampled which includes all main groups of TRP channels. Amino acid sequences were aligned using MAFFT v.7 (Katoh et al., 2017) allowing the program to choose the alignment strategy (FFT-NS-i). We used the proposed model tool of IQ-Tree v.1.6.12 (Minh et al., 2020) to select the best-fitting model of amino acid substitution (LG+G4). We used the maximum likelihood method to obtain the best tree using the program IQ-Tree v1.6.12 (Minh et al., 2020). We performed five tree searches in which the initial gene tree was provided by ourselves, which was previously estimated using IQ-Tree v.1.6.12 (Minh et al., 2020). We also carried out five additional analyses in which we performed more exhaustive tree searches by modifying the strength of the perturbation (-pers) from 0.5 (default value) to 0.9 and the number of unsuccessful iterations to stop (-nstop) from 100 (default value) to 500. The tree with the highest likelihood score was chosen. We assessed support for the nodes using the ultrafast bootstrap routine as implemented in IQ-Tree v1.6.12 (PMID: 29077904). Potassium voltage-gated channel subfamily A member 2 (KCNA2) and Sodium Voltage-Gated Channel Alpha Subunit 8 (SCN8A) amino acid sequences from mammals were included as an outgroup (Supplementary Table S1).

**MSA database.** We retrieved 969 protein sequences corresponding to the subgroups TRPV, TRPA1, TRPM, TRPN, and TRPC in representative species of all main lineages of amniotes from the Orthologous Matrix project (OMA) (Altenhoff et al., 2021). 646 extra sequences (including those of unicellular TRP) from the Uniprot protein database were added to the pool of sequences rescued from the OMA server and then aligned using MAFFT (FFTNS1 strategy). The region corresponding to the transmembrane was identified, and the rest of the sequence was removed, leaving the section from the last portion of the pre-TM1 region to the last residue of the TRP helix (residues 396 to 718, using rTRPV1 as reference). A second round of sequence alignment in MAFFT (L-INS-I strategy) was performed and manually refined to minimize gaps. The resulting MSA –Primary MSA– database contains 1481 monophyletic sequences, 861 selected from OMA database and 620 from Uniprot, including sequences annotated as TRP or TRP-like channels and several uncharacterized
protein sequences (Figure 1 - figure supplement 2). In order to facilitate the visualization by minimizing even more the gaps, TRPS and TRPVL sequences were taken out for the generation of figures but remained for the statistical processing.

**HMMs.** The fingerprint TRP residues in Table 1 were identified from sequence and structural alignments. To further corroborate this conservation analysis, we defined a profile hidden Markov model (HMM) and analyzed the emission probabilities at each position. For each matching position of the HMM, emission probabilities describe a generalized Bernoulli distribution: conserved positions are characterized by a distribution peaked around the invariant amino acid. We use Shannon entropy to quantify the “peakedness” of the distribution. We create separate HMMs from the MSAs of the major TRP protein subfamilies, and also a cumulative HMM that includes all of these MSAs together. To define the matching positions, we selected only the MSA columns with less than 50% of gaps. To train the HMM, we used hmmbuild from the HMMER suite. The HMM predictions for the fingerprint residues in each protein subfamily match those in Table 1, which confirms the identification of the fingerprint residues in that Table. The 12 signature residues are among the low entropy positions in the MSA, indicating that the HMM has determined they are indeed well-conserved (Figure 2 - figure supplement 2b).

**Structural alignment.** Structural alignment in figure 3C was performed in the inbuilt extension of VMD, aligning the first and sixth transmembrane helices, following the numbering provided by the MSA.

**Coevolution analysis.** Coevolution scores were calculated using asymmetric pseudo-likelihood maximization direct coupling analysis algorithm (aplmDCA), (Ekeberg et al., 2013). This algorithm finds the approximate parameters of the maximum entropy probabilistic model consistent with selected MSA statistics (univariate and bivariate
frequency distributions). Default parameters were used for field and coupling
regularization and sequence reweighting (lambda_h = 0.01, lambda_j = 0.01, theta = 0.1).

**Distance matrices.** Data PFAM provides deletion-free (insert-only) MSAs whose
sequences they have aligned to PDBs on a residue-by-residue basis. In order to align
sequences in this way, the resulting MSA has a high number of gaps. Our structural analysis
includes 140 TRP sequence-structure pairs across multiple TRP subfamilies, which were
already pre-indexed and mapped by PFAM, and then subsequently verified by us. The
original MSA was 958 long, and after feature selection and verification, 208 positions
remained from 91 sequences. **Feature Selection.** To narrow down which positions to use,
we removed positions with high gap frequency, which are potentially not shared across the
TRP subfamilies. Accordingly, we considered only MSA positions present in more than 96%
of the sequences. This means that the tolerable gap frequency across the entire alignment
for a position was <4%, ensuring that the positions selected in this way were present
across all subfamilies. We created a frequency histogram of the full MSA, which showed
which positions of the MSA have a high frequency of gaps. We also used knowledge-based
feature selection to isolate positions that fall within key regions of functional importance:
pre-TM1, TM1, TM2, TM3, TM4, Linker, TM5, Pore, TM6, and TDH. **Distograms.** From the
feature selected sequence-structure maps, pairwise distance matrices were computed from
the PDB structures using Cβ-Cβ residue distances (or Cα for glycine). Distograms for mean,
variance, and normalized variance were computed across the distance matrices and
visualized. Not only do they show which pairwise positions vary by distance, but they also
show which pairwise distances are conserved, across the subfamilies. **Interpretation.** The
mean distograms are the easiest to interpret because they simply give an overall picture of
which residue pairs are nearby, or distant, to others across the subfamilies. Mean
distograms showed expected α-helical structure and reasonable separation between the
different helical elements. For example, whereas TM1-TM4 are observed to be close to each
other as a group, they appear to be further away from TM5-TM6, which are close to each
other as a group. This is consistent with the canonical structure of 6TM channel proteins.
On the other hand, variance distograms give an overall picture of variation in residue-residue distances. This information is useful to characterize specific sequence patterns within a large collection of structural states. Given the wide range of distances, thermal fluctuations and other noise sources are expected to produce the largest variances for residue-pairs that are separated by large distances. To remove this bias, variances are normalized, i.e. they are divided by the square of the average distance (from the distance matrix). By this procedure we effectively highlight variance of residues that are at short distances.

Identification of residues at the Aromatic Core (AC). A python script was used to identify the aromatic residues facing the internal cavity formed by the TM1 to TM4 helixes in channel structures extracted from PDB files. For this, the distances between all alpha carbons in opposing helixes (TM1-TM3 and TM2-TM4) were calculated, choosing the nearest residue pairs. The same procedure was followed for gamma/alpha pairs. Those residues with \((C\alpha-C\alpha)\)distance > \((C\alpha-C\gamma)\)distance were recorded in a list. After, we calculated distances between all atoms in pairs of aromatic residues in this list, and kept the minimal distances between pairs. Then, a third code grouped in clusters all the pairs with a distance below a threshold of 5Å. Scripts and files used are available in https://github.com/brauchilab/ProteinCoreCluster.

Figure preparation. To visualize the identities and gap patterns on the MSA, images were exported using Jalview 2.11.1.3 (Waterhouse et al., 2009). The sequence logos show the distribution of amino acid residues at each position in the regions of interest, and were generated using WebLogo version 3 (Crooks, 2004). The structural figures were generated using VMD 1.9.2 (Humphrey et al., 1996), using the files next PDB files: 7LP9 for rTRPV1; 6O6A for paTRPM8; 6AEI for hTRPC5; 3J9P for hTRPA; 5VKQ for dmTRPN1; 6PW4 for crTRP1; 6C96 for mTPCN1; and 2R9R for rKv1.2. Direct interactions between residues were identified using a distance threshold of 4 Å, except for the pi-pi interactions, where a 5 Å threshold was used (Piovesan et al., 2016).
Keywords: MSA, evolution, allosterism, structure.

Acknowledgements: We acknowledge Dr. Gonzalo Riadi from Universidad de Talca for providing computer support. JCO acknowledges the Integrative Biology Group members, Universidad Austral de Chile, for their constant support, scientific enthusiasm, and creative feedback. This work was supported by Fondo Nacional de Desarrollo Científico y Tecnológico from Chile (FONDECYT 3140233) to CKC, (FONDECYT 1191868) to SB, (FONDECYT 1210471) to JCO, and ANID-Millennium Science Initiative Program #NC160011 (SB and JCO). This work was funded in part by the National Institutes of Health (R01GM093290, S10OD020095, and R01GM131048; VC), and the National Science Foundation through Grant no. IOS-1934848 (VC). This research includes calculations carried out on HPC resources supported in part by the National Science Foundation through major research instrumentation grant number 1625061 as well as by the US Army Research Laboratory under contract number W911NF-16-2-0189 (VC).

Author contribution: SEB, CKC, DCB, FAM, and VC designed experimental strategy and analyzed data. DCB, VC, KZ and JCO designed the taxonomic sampling, retrieved amino acid sequences and performed phylogenetic analyses. DCB, FAM, JCO, and SEB prepared the figures. DCB, JCO, and SEB wrote the manuscript.
Table 1. Table summarizing the percentage of identity of highly conserved positions in the alignment and in the corresponding positions in rTRPV1, pm TRPM8, mTRPC5, hTRPA1, and dmTRPN1. Residues in the corresponding position of the two unicellular GI-TRPs identified (i.e., CrTRP1 and CsTRP1) are indicated. Corresponding residues in GII-TRPs and non-TRP channels are also indicated. The last column corresponds to the total number of fingerprint residues for consensus. Green residues correspond to identities while black represents homology. Red shades denote non-conserved residues.

<table>
<thead>
<tr>
<th></th>
<th>W</th>
<th>F</th>
<th>Φ</th>
<th>G</th>
<th>Φ</th>
<th>Φ</th>
<th>N</th>
<th>L</th>
<th>I</th>
<th>A</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Ident</td>
<td>95.3</td>
<td>95.5</td>
<td>91.0</td>
<td>98.0</td>
<td>96.6</td>
<td>98.6</td>
<td>93.7</td>
<td>95.3</td>
<td>89.7</td>
<td>95.3</td>
<td>93.1</td>
</tr>
<tr>
<td>Pos Align</td>
<td>81</td>
<td>320</td>
<td>331</td>
<td>633</td>
<td>662</td>
<td>755</td>
<td>852</td>
<td>862</td>
<td>864</td>
<td>865</td>
<td>866</td>
</tr>
<tr>
<td>Pos rTRPV1</td>
<td>426</td>
<td>434</td>
<td>441</td>
<td>563</td>
<td>591</td>
<td>638</td>
<td>666</td>
<td>676</td>
<td>678</td>
<td>679</td>
<td>680</td>
</tr>
<tr>
<td>Pos pmTRPM8</td>
<td>677</td>
<td>733</td>
<td>740</td>
<td>848</td>
<td>875</td>
<td>902</td>
<td>957</td>
<td>967</td>
<td>969</td>
<td>970</td>
<td>971</td>
</tr>
<tr>
<td>Pos mTRPC5</td>
<td>315</td>
<td>367</td>
<td>374</td>
<td>504</td>
<td>531</td>
<td>576</td>
<td>608</td>
<td>618</td>
<td>620</td>
<td>621</td>
<td>622</td>
</tr>
<tr>
<td>Pos hTRPA1</td>
<td>711</td>
<td>716</td>
<td>726</td>
<td>857</td>
<td>884</td>
<td>909</td>
<td>944</td>
<td>954</td>
<td>956</td>
<td>957</td>
<td>958</td>
</tr>
<tr>
<td>Pos dmTRPN1</td>
<td>1260</td>
<td>1304</td>
<td>1311</td>
<td>1427</td>
<td>1454</td>
<td>1501</td>
<td>1541</td>
<td>1551</td>
<td>1553</td>
<td>1554</td>
<td>1555</td>
</tr>
<tr>
<td>TRPVs</td>
<td>W</td>
<td>F</td>
<td>Y</td>
<td>G</td>
<td>Φ</td>
<td>F</td>
<td>Φ</td>
<td>N</td>
<td>L</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>TRPMs</td>
<td>W</td>
<td>F</td>
<td>Y</td>
<td>G</td>
<td>Φ</td>
<td>Y</td>
<td>Φ</td>
<td>N</td>
<td>L</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>TRPCs</td>
<td>W</td>
<td>F</td>
<td>Φ</td>
<td>G</td>
<td>F</td>
<td>F</td>
<td>Φ</td>
<td>N</td>
<td>L</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>TRPA1</td>
<td>W</td>
<td>F</td>
<td>Y</td>
<td>G</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>N</td>
<td>L</td>
<td>I</td>
<td>G</td>
</tr>
<tr>
<td>TRPN1</td>
<td>W</td>
<td>F</td>
<td>H</td>
<td>G</td>
<td>F</td>
<td>F</td>
<td>Y</td>
<td>N</td>
<td>L</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>TRPY</td>
<td>W</td>
<td>N</td>
<td>S</td>
<td>G</td>
<td>F</td>
<td>T</td>
<td>Φ</td>
<td>N</td>
<td>L</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>TRPS</td>
<td>W</td>
<td>Φ</td>
<td>Y</td>
<td>G</td>
<td>W</td>
<td>Y</td>
<td>T</td>
<td>L</td>
<td>F</td>
<td>A</td>
<td>W</td>
</tr>
<tr>
<td>TRPVL</td>
<td>W</td>
<td>-</td>
<td>N</td>
<td>G</td>
<td>Φ</td>
<td>F</td>
<td>W</td>
<td>N</td>
<td>F</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>Unicellular</td>
<td>CrTRP1</td>
<td>W</td>
<td>W</td>
<td>L</td>
<td>G</td>
<td>F</td>
<td>Q</td>
<td>F</td>
<td>N</td>
<td>F</td>
<td>I</td>
</tr>
<tr>
<td>CsTRP2</td>
<td>W</td>
<td>W</td>
<td>Y</td>
<td>N</td>
<td>F</td>
<td>F</td>
<td>Y</td>
<td>N</td>
<td>L</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>TRP-GII</td>
<td>hPKD2</td>
<td>-</td>
<td>F</td>
<td>-</td>
<td>S</td>
<td>Y</td>
<td>F</td>
<td>F</td>
<td>N</td>
<td>F</td>
<td>L</td>
</tr>
<tr>
<td>mTRPML1</td>
<td>F</td>
<td>F</td>
<td>H</td>
<td>N</td>
<td>Y</td>
<td>F</td>
<td>F</td>
<td>S</td>
<td>F</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>Non-TRP</td>
<td>chTPCN1</td>
<td>W</td>
<td>Y</td>
<td>-</td>
<td>R</td>
<td>F</td>
<td>F</td>
<td>Y</td>
<td>N</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Nav1.4</td>
<td>W</td>
<td>F</td>
<td>L</td>
<td>N</td>
<td>F</td>
<td>F</td>
<td>V</td>
<td>N</td>
<td>F</td>
<td>L</td>
<td>A</td>
</tr>
<tr>
<td>hTPCN2</td>
<td>W</td>
<td>F</td>
<td>Y</td>
<td>A</td>
<td>F</td>
<td>W</td>
<td>W</td>
<td>N</td>
<td>F</td>
<td>L</td>
<td>A</td>
</tr>
<tr>
<td>P2X3</td>
<td>W</td>
<td>Y</td>
<td>Y</td>
<td>D</td>
<td>T</td>
<td>F</td>
<td>G</td>
<td>N</td>
<td>L</td>
<td>K</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>Cav1.2 III</td>
<td>Navab</td>
<td>KvAP</td>
<td>Kv11.1</td>
<td>Kv1.2</td>
<td>Shaker</td>
<td>Cav1.2 III</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>------------</td>
<td>------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

735

736
Table 2. Summary of structural-functional studies and the reported effects of site directed mutagenesis in signature residues. First column indicates the equivalent signature residue in the rTRPV1 sequence. Second column indicates the channel member studied. Third and fourth columns correspond to the type of study used to determine functional effects. MDS: Molecular dynamics simulations, SDM: Site-directed mutagenesis.
<table>
<thead>
<tr>
<th>TRPV1 Position</th>
<th>Channel Residue</th>
<th>Evidence Source</th>
<th>Effect</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>W426</strong></td>
<td>rTRPV1 W426A</td>
<td>SDM</td>
<td>Insensitive to Capsaicin</td>
<td>(Zheng et al. 2018)</td>
</tr>
<tr>
<td></td>
<td>hTRPV3 W433</td>
<td>Structure</td>
<td>Part of the 2-APB binding pocket</td>
<td>(Zubcevic et al. 2019)</td>
</tr>
<tr>
<td></td>
<td>rTRPV1 W426A</td>
<td>SDM</td>
<td>Impaired Voltage and Capsaicin response</td>
<td>(Zheng et al. 2018)</td>
</tr>
<tr>
<td></td>
<td>rTRPM8 W682A</td>
<td>SDM</td>
<td>Impaired Voltage and Menthol response</td>
<td>(Zheng et al. 2018)</td>
</tr>
<tr>
<td></td>
<td>hTRPA1 W711</td>
<td>Structure</td>
<td>Interaction site with phospholipids</td>
<td>(Suo et al. 2020)</td>
</tr>
<tr>
<td><strong>F434</strong></td>
<td>drTRPC4 F366</td>
<td>Structure</td>
<td>Part of cholesterol binding pocket</td>
<td>(Vinayagam et al. 2018)</td>
</tr>
<tr>
<td></td>
<td>faTRPM8 F738</td>
<td>Structure</td>
<td>Part of the Icilin and WS-12 binding pocket</td>
<td>(Izquierdo et al. 2021)</td>
</tr>
<tr>
<td><strong>F/Y441</strong></td>
<td>TRPV1 Y441S</td>
<td>SDM</td>
<td>Nonfunctional</td>
<td>(Boukalova, Teisinger, and Vlachova 2013)</td>
</tr>
<tr>
<td></td>
<td>rTRPM8 Y745H</td>
<td>SDM</td>
<td>Critical on Menthol Sensitivity.</td>
<td>(Bandell et al. 2006)</td>
</tr>
<tr>
<td></td>
<td>Y745H SDM</td>
<td></td>
<td>Low response to Mentol, but normal response to temperature. Critical on inhibition SKF96365-mediated of Cold- and voltage-activation, but just partially on other inhibitor</td>
<td>(Malkia et al. 2009)</td>
</tr>
<tr>
<td>Gene</td>
<td>Mutation</td>
<td>Function</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>-----------------------------------------------</td>
<td>-------------------------------</td>
<td></td>
</tr>
<tr>
<td>hTRPC3</td>
<td>Y745H</td>
<td>SDM Low response to Menthol, but normal response to temperature</td>
<td>(Nguyen et al. 2021)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y374</td>
<td>Structure Part of the inhibitor, clemizole, binding pocket</td>
<td>(Song et al. 2021)</td>
<td></td>
</tr>
<tr>
<td>rTRPV1</td>
<td>G563/C</td>
<td>SDM Gain of Function</td>
<td>(Boukalova et al. 2010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G563S/A</td>
<td>SDM Gain of Function, Inhibition by proton of Max current induced by capsaicin</td>
<td>(Boukalova, Teisinger, and Vlachova 2013)</td>
<td></td>
</tr>
<tr>
<td>mTRPV1</td>
<td>G564S</td>
<td>SDM Gain of Function</td>
<td>(Duo et al. 2018)</td>
<td></td>
</tr>
<tr>
<td>rTRPV3</td>
<td>G573S/C</td>
<td>SDM Gain of Function</td>
<td>(Xiao et al. 2008)</td>
<td></td>
</tr>
<tr>
<td>mTRPV3</td>
<td>G573S</td>
<td>SDM Gain of Function, Olmsted Syndrome</td>
<td>(Lin et al. 2012)</td>
<td></td>
</tr>
<tr>
<td>mTRPV3</td>
<td>G573S</td>
<td>SDM Non responsive to Menthol, Camphor and APB and mildly responsive to temperature</td>
<td>(Nguyen et al. 2021)</td>
<td></td>
</tr>
<tr>
<td>rTRPV1</td>
<td>G563S</td>
<td>SDM Non responsive to Camphor and APB and mildly responsive to temperature</td>
<td>(Nguyen et al. 2021)</td>
<td></td>
</tr>
<tr>
<td>mTRPC4/5</td>
<td>G503S/G504S</td>
<td>SDM Gain of Function</td>
<td>(Beck et al. 2013)</td>
<td></td>
</tr>
<tr>
<td>hTRPC3</td>
<td>G552</td>
<td>Structure Coupled W673 from TRP domain</td>
<td>(Fan et al. 2018)</td>
<td></td>
</tr>
<tr>
<td>hTRPC3</td>
<td>G552</td>
<td>Structure Coupled W673 from TRP domain</td>
<td>(Fan et al. 2018)</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Mutant</td>
<td>Type</td>
<td>Description</td>
<td>Reference</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>rTRPV1</td>
<td>F591</td>
<td>MDS</td>
<td>Part of the vanilloid binding pocket</td>
<td>(Elokely et al. 2015)</td>
</tr>
<tr>
<td></td>
<td>F591A</td>
<td>SDM</td>
<td>Low Capsaicin response, non response to pH and not RTX binding</td>
<td>(Ohbuchi et al. 2016)</td>
</tr>
<tr>
<td>hTRPM4</td>
<td>Y944</td>
<td>Structure</td>
<td>Forming face to face π-stack with F1027 on TM5</td>
<td>(Duan et al. 2018)</td>
</tr>
<tr>
<td>rTRPV1</td>
<td>F638A</td>
<td>SDM</td>
<td>Gain of Function, NMDG/Na selectivity raised</td>
<td>(Munns et al. 2015)</td>
</tr>
<tr>
<td></td>
<td>F638W</td>
<td>SDM</td>
<td>Enhanced the sensitivity to the acylpolyamine toxins AG489 and AG505</td>
<td>(Kitaguchi and Swartz 2005)</td>
</tr>
<tr>
<td>rTRPV2</td>
<td>F601</td>
<td>Structure</td>
<td>Part of the cannabidiol binding pocket</td>
<td>(Pumroy et al. 2019)</td>
</tr>
<tr>
<td>rTRPM8</td>
<td>Y908A/W</td>
<td>SDM</td>
<td>Not responsive to Cold and Menthol but responsive to Icilin</td>
<td>(Bidaux et al. 2015)</td>
</tr>
<tr>
<td></td>
<td>Y908F</td>
<td>SDM</td>
<td>Totally responsive to Cold and Menthol and Icilin</td>
<td>(Bidaux et al. 2015)</td>
</tr>
<tr>
<td>zfTRPC4</td>
<td>F572</td>
<td>Structure</td>
<td>Stabilizes the pore through an hydrophobic contact with neighbor protomer</td>
<td>(Vinayagam et al. 2018)</td>
</tr>
<tr>
<td>mTRPC5</td>
<td>F576A</td>
<td>SDM</td>
<td>Nonfunctional, dominant negative</td>
<td>(Strübing et al. 2003)</td>
</tr>
<tr>
<td>hTRPC5</td>
<td>F576A</td>
<td>SDM</td>
<td>Differential effect on agonists: Not responsive to AM237, but responsive elgerin</td>
<td>(Wright et al. 2020)</td>
</tr>
<tr>
<td>hTRPA1</td>
<td>F909A</td>
<td>SDM</td>
<td>Affect different agonists and antagonists responses</td>
<td>(Chandrabalan et al. 2019)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>--------</td>
<td>-----------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>F909T</strong></td>
<td><strong>SDM</strong></td>
<td><strong>Abolish the A-967079-inhibition of AITC-evoked response</strong></td>
<td>(Paulsen et al. 2015)</td>
<td></td>
</tr>
<tr>
<td><strong>Y/F666</strong></td>
<td><strong>rTRPV1</strong></td>
<td><strong>Y666A</strong></td>
<td><strong>SDM</strong></td>
<td><strong>Nonfunctional (present in membrane)</strong></td>
</tr>
<tr>
<td><strong>mTRPV3</strong></td>
<td><strong>Y661C</strong></td>
<td><strong>SDM</strong></td>
<td><strong>Not responsive to Temp, but responsive to agonist (2-APB and Camphor)</strong></td>
<td>(Grandl et al. 2008)</td>
</tr>
<tr>
<td><strong>hTRPV4</strong></td>
<td><strong>Y702L</strong></td>
<td><strong>SDM</strong></td>
<td><strong>Not responsiveness to Temp, Agonist and Swelling</strong></td>
<td>(Klausen et al. 2014)</td>
</tr>
<tr>
<td><strong>hTRPM6</strong></td>
<td><strong>Y1053C</strong></td>
<td><strong>SDM</strong></td>
<td><strong>Causes hypomagnesemia with secondary hypocalcemia, Decreased Current amplitude in heterologous expression in HEK293</strong></td>
<td>(Lainez et al. 2013)</td>
</tr>
<tr>
<td><strong>hTRPM4</strong></td>
<td><strong>F1027</strong></td>
<td><strong>Structure</strong></td>
<td><strong>Forming face to face π-stack with Y944 on TM5</strong></td>
<td>(Duan et al. 2018)</td>
</tr>
<tr>
<td><strong>hTRPA1</strong></td>
<td><strong>F909A</strong></td>
<td><strong>SDM</strong></td>
<td><strong>Affect different agonists responses</strong></td>
<td>(Chandrabalan et al. 2019)</td>
</tr>
<tr>
<td><strong>N676</strong></td>
<td><strong>rTRPV1</strong></td>
<td><strong>N676</strong></td>
<td><strong>MDS</strong></td>
<td><strong>Gating relies on the rotatory motion of N676</strong></td>
</tr>
<tr>
<td><strong>N676A</strong></td>
<td><strong>SDM</strong></td>
<td><strong>Nonfunctional (present in membrane)</strong></td>
<td>(Susankova et al. 2007)</td>
<td></td>
</tr>
<tr>
<td><strong>N676F</strong></td>
<td><strong>SDM</strong></td>
<td><strong>Not responsive to Temp and Agonist (Cap/RTX) and reduced response to pH</strong></td>
<td>(Kuzhikandathil et al. 2001)</td>
<td></td>
</tr>
<tr>
<td><strong>hTRPA1</strong></td>
<td><strong>N944A</strong></td>
<td><strong>SDM</strong></td>
<td><strong>Abolished inhibition by AZ868 and A-967079, but not by HC-030031</strong></td>
<td>(Klement et al. 2013)</td>
</tr>
<tr>
<td><strong>L678</strong></td>
<td>rTRPV1</td>
<td>L678A</td>
<td>SDM</td>
<td>Low response to Agonist (Cap) and Temp, but normal response to both at the same time</td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
<td>-------</td>
<td>-----</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td><strong>L678P</strong></td>
<td>SDM</td>
<td>Not responsive to Temp and Agonist (Cap/RTX) and reduced response to pH</td>
<td>(Kuzhikandathil et al. 2001)</td>
<td></td>
</tr>
<tr>
<td><strong>TRPV3</strong></td>
<td>L768F</td>
<td>SDM</td>
<td>Olmsted Syndrome and Erythromelalgia (gain of function)</td>
<td>(Duchatelet et al. 2014)</td>
</tr>
<tr>
<td><strong>TRPC3</strong></td>
<td>L654</td>
<td>Structure</td>
<td>Constriction site in the lower region of the pore</td>
<td>(Fan et al. 2018)</td>
</tr>
<tr>
<td><strong>I679</strong></td>
<td>rTRPV1</td>
<td>I697</td>
<td>Structure</td>
<td>Constriction site in the lower region of the pore</td>
</tr>
<tr>
<td>I697</td>
<td>Structure</td>
<td>Constriction site in the lower region of the pore</td>
<td>(Cao et al. 2013)</td>
<td></td>
</tr>
<tr>
<td>I697</td>
<td>Structure</td>
<td>Constriction site in the lower region of the pore</td>
<td>(Gao et al. 2016)</td>
<td></td>
</tr>
<tr>
<td>I697</td>
<td>Structure</td>
<td>Constriction site in the lower region of the pore</td>
<td>(Chugunov et al. 2016)</td>
<td></td>
</tr>
<tr>
<td>I697</td>
<td>Structure</td>
<td>Constriction site in the lower region of the pore</td>
<td>(Susankova et al. 2007)</td>
<td></td>
</tr>
<tr>
<td>mTRPM4</td>
<td>I1036</td>
<td>Structure</td>
<td>Constriction site in the lower region of the pore</td>
<td>(Guo et al. 2017)</td>
</tr>
<tr>
<td>hTRPM4</td>
<td>I1040</td>
<td>Structure</td>
<td>Constriction site in the lower region of the pore</td>
<td>(Autzen et al. 2018)</td>
</tr>
<tr>
<td>drTRPC4</td>
<td>I617</td>
<td>Structure</td>
<td>Constriction site in the lower region of the pore</td>
<td>(Vinayagam et al. 2018)</td>
</tr>
<tr>
<td>rTRPV4</td>
<td>I715</td>
<td>SDM</td>
<td>Hydrophobic single-residue</td>
<td>(Zheng et al. 2018)</td>
</tr>
<tr>
<td>Gene</td>
<td>Mutations</td>
<td>Domain</td>
<td>Effect</td>
<td>Reference</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>--------</td>
<td>------------------------------------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rTRPV4</td>
<td>A716S</td>
<td>SDM</td>
<td>Not responsive to agonists (4αPDD, Hypotonicity and AA), cause SMD Kozlowski type, and Metatropic Dysplasia</td>
<td>(Krakow et al. 2009)</td>
</tr>
<tr>
<td>hTRPV1A</td>
<td>G955A</td>
<td>SDM</td>
<td>Slower inactivation rate. Lower rectification rates</td>
<td>(Benedikt et al. 2009)</td>
</tr>
<tr>
<td></td>
<td>G958R</td>
<td>SDM</td>
<td>Inward-rectifier, constitutively active at resting potential, and impaired response to AITC</td>
<td>(Benedikt et al. 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W697</td>
<td>rTRPV1</td>
<td>Structure</td>
<td>It forms a hydrogen bond with the main chain carbonyl oxygen of F559 at the beginning of the S4–S5 linker</td>
<td>(Liao et al. 2013)</td>
</tr>
<tr>
<td></td>
<td>W697A</td>
<td>SDM</td>
<td>Low Response to Cap/Em</td>
<td>(Valente et al. 2008)</td>
</tr>
<tr>
<td></td>
<td>W697x</td>
<td>SDM</td>
<td>Low Response to Cap/Em, Affect allostéric activation</td>
<td>(Gregorio-Teruel et al. 2014)</td>
</tr>
<tr>
<td>TRPV3</td>
<td>W692G</td>
<td>SDM</td>
<td>Gain of Function, Olmsted Syndrome</td>
<td>(Lin et al. 2012)</td>
</tr>
<tr>
<td>Protein</td>
<td>Mutation</td>
<td>Domain</td>
<td>Description</td>
<td>Reference</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>--------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>TRPV4</td>
<td>W733R</td>
<td>SDM</td>
<td>Gain of Function, limited agonist response, and not inactivation to long depolarization</td>
<td>(Teng et al. 2015)</td>
</tr>
<tr>
<td>TRPC3</td>
<td>W673</td>
<td>Structure</td>
<td>It is extensively coupled with the S4-S5 linker through interactions with G552</td>
<td>(Fan et al. 2018)</td>
</tr>
<tr>
<td>TRPC4</td>
<td>W674</td>
<td>Structure</td>
<td>Coupled with the S4-S5 linker through interactions with G553 and P546 on TM4</td>
<td>(Fan et al. 2018)</td>
</tr>
</tbody>
</table>
Table 3. Summary of mutation effects reported in the literature for residues forming part of the AC and the conserved residue at TM4 connecting TM4 with TM5. First column indicates the equivalent signature residue in the rTRPV1 sequence. Second column indicates the channel studied. Third row corresponds to the effect of the mutation and/or proposed function.

<table>
<thead>
<tr>
<th>TRPV1 Position</th>
<th>Channel</th>
<th>Mutation</th>
<th>Effect</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>F/Y444</td>
<td>rTRPV1</td>
<td>Y444S</td>
<td>Nonfunctional</td>
<td>(Boukalova, Teisinger, and Vlachova 2013)</td>
</tr>
<tr>
<td></td>
<td>mTRPM3</td>
<td>Y885T</td>
<td>Impaired non-canonical current induced by pregnenolone sulfate + clotrimazol</td>
<td>(Held et al. 2018)</td>
</tr>
<tr>
<td>F448</td>
<td>rTRPV1</td>
<td>F448L</td>
<td>Decreased pH response but maintain all Cap responsiveness</td>
<td>(Boukalova, Teisinger, and Vlachova 2013)</td>
</tr>
<tr>
<td></td>
<td>mTRPM3</td>
<td>Y888T</td>
<td>Similar to wt response to pregnenolone sulfate + clotrimazol</td>
<td>(Held et al. 2018)</td>
</tr>
<tr>
<td>Y/F554</td>
<td>rTRPV1</td>
<td>Y554A</td>
<td>Nonfunctional (Cap 10µM, -70 to 200mV, 48°C)</td>
<td>(Boukalova et al. 2010)</td>
</tr>
<tr>
<td></td>
<td>Y554F</td>
<td>Normal responsiveness</td>
<td>(Boukalova et al. 2010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y554A</td>
<td>Not responsiveness to pH, Cap and RTX</td>
<td>(Elokely et al. 2015)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y554A</td>
<td>Increased sensitivity and affinity to 2-APB</td>
<td>(Singh et al. 2018)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rTRPV2</td>
<td>Y514A</td>
<td>Increased sensitivity and affinity</td>
<td>(Singh et al. 2018)</td>
</tr>
<tr>
<td>2-APB</td>
<td>rTRPV3</td>
<td>Y564A</td>
<td>Increased affinity to 2-APB</td>
<td>(Singh et al. 2018)</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>-------</td>
<td>-----------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Y/F555</td>
<td>rTRPV1</td>
<td>Y555S</td>
<td>Nonfunctional (Cap 10µM, -70 to 200mV, 48°C) 1</td>
<td>(Boukalova et al. 2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y555F</td>
<td>Normal responsiveness</td>
<td>(Boukalova et al. 2010)</td>
</tr>
<tr>
<td>W549</td>
<td>rTRPV1</td>
<td>W549A</td>
<td>Not responsive to Cap (and RTX and others) and pH</td>
<td>(Ohbuchi et al. 2016)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W549A</td>
<td>Interaction with vanillyl moiety of RTX or Capsaicin</td>
<td>(Gavva et al. 2004)</td>
</tr>
<tr>
<td></td>
<td>hTRPV4</td>
<td>W568A</td>
<td>Impaired responsiveness to heat and agonists (4α-PDD and BAA); but responsive to swelling and endogen lipids</td>
<td>(Vriens et al. 2007)</td>
</tr>
<tr>
<td></td>
<td>mTRPM3</td>
<td>W982R</td>
<td>Abolished non-canonical current induced by pregnenolone sulfate + clotrimazol</td>
<td>(Held et al. 2018)</td>
</tr>
<tr>
<td></td>
<td>mTRPM3</td>
<td>W982F</td>
<td>Similar to wt response to pregnenolone sulfate + clotrimazol</td>
<td>(Held 2018)</td>
</tr>
<tr>
<td></td>
<td>hTRPA1</td>
<td>Y840F</td>
<td>Reduced potency of ligand (GNE551)</td>
<td>(Liu et al. 2021)</td>
</tr>
<tr>
<td></td>
<td>hTRPA1</td>
<td>Y840W/H/L/A</td>
<td>Completely abolished potence of ligand (GNE551)</td>
<td>(Liu 2021)</td>
</tr>
</tbody>
</table>
**Figure legends**

**Figure 1. Maximum likelihood tree showing relationships among TRP channels.** The scale denotes substitutions per site and colors represent lineages. Potassium voltage-gated channel subfamily A member 2 (KCNA2), and Sodium Voltage-Gated Channel Alpha Subunit 8 (SCN8A) sequences were included as an outgroup.

**Figure 1 - Figure supplement 1. Maximum likelihood tree showing relationships among TRP channels with species indicated.** The scale denotes substitutions per site and colors represent lineages. Potassium voltage-gated channel subfamily A member 2 (KCNA2), and Sodium Voltage-Gated Channel Alpha Subunit 8 (SCN8A) sequences were included as an outgroup.

**Figure 1 - Figure supplement 2. Maximum likelihood tree showing relationships among TRP channels and the putative TRPs from unicellular organisms.** The scale denotes substitutions per site and colors represent lineages. Potassium voltage-gated channel subfamily A member 2 (KCNA2), and Sodium Voltage-Gated Channel Alpha Subunit 8 (SCN8A) sequences were included as an outgroup.

**Figure 1 - Figure supplement 3. Maximum likelihood tree showing relationships among TRP channels and the putative TRPs from unicellular organisms with all species indicated.** The scale denotes substitutions per site and colors represent lineages. Potassium voltage-gated channel subfamily A member 2 (KCNA2), and Sodium Voltage-Gated Channel Alpha Subunit 8 (SCN8A) sequences were included as an outgroup.

**Figure 1 - Figure supplement 4. Pipeline diagram.** Our analyses begin by pulling TRP sequences from the Uniprot and OMA databases (upper left). From this set of sequences (1615) we handpicked 58 individuals for phylogenetic analysis and produced an initial MSA with the whole set (blue) by using MAFFT (FFTNS1). Knowledge-based feature selection (purple) was implemented to retain only those positions between the pre-TM1 and TDh regions, and then performed another MAFFT (L-INS-I) to produce a new MSA, defined as Primary, containing 1481 sequences. From this Primary MSA, we identified the Fingerprint Residues (orange) using two separate analyses, a Fourier Analysis and an HMM Analysis (bottom-center). From the Primary MSA, we used feature selection to create a third and final MSA, the Structure MSA (138 structures). The knowledge-based feature selection for this MSA trimmed the positions to include only those within the borders of the individual helices, and nothing in between them. The statistical feature selection removed...
any positions with a gap frequency above 4%. The sequences in this MSA were all from Uniprot only and were mapped on a residue-by-residue basis to their corresponding PDB structures using a Uniprot-PDB index provided by PFAM. From this sequence-structure map, pairwise cβ-cβ (or cα in the case of glycine) distance matrices were computed, and from these the various distograms (mean, variance, normalized variance) were computed (orange). These distograms were used to corroborate the existence of the fingerprint residues identified by frequency and HMM Analyses.

Figure 2. Conserved residues in GI-TRPs. (a) stacked histogram showing the amino acidic probability in each position of the MAFFT alignment. Gray boxes depict the trans-membrane helices (TM1-TM6) and features such as pre-TM1, the TM4-TM5 linker (L), the Selectivity Filter and Pore Helix (SF&PH) and the TRP domain helix (TDH). Numbers over the arrows indicate the position in the alignment, and in brackets the corresponding position in the rat TRPV1 primary sequence. (b) Sequence logos for the TRP family, depicting highly conserved residues (>90% identity). (c) Upper: Cartoon of a TRP channel monomer depicting the location of conserved residues in the secondary structure. Φ denotes 6 carbon aromatic residues (i.e., Tyr or Phe). Bottom: Table summarizing the highly conserved positions in alignment and in the corresponding position in the rat TRPV1 primary sequence, along with the percentage of identity. Consensus residues for each subfamily are indicated. The last column corresponds to the total number of fingerprint residues for each subfamily. Green residues correspond to identities while black represents homology. Red shades denote non-conserved residues.

Figure 2 - Figure supplement 1. Stacked histograms showing the amino acid frequency in each position on the MAFFT alignment for the different TRP subgroups. Gray boxes depict the trans-membrane helices (TM1-TM6) and features like pre-TM1, the TM4-TM5 linker (L), the Selectivity Filter and Pore Helix (SF&PH) and the TRP domain helix (TDH). Numbers over the arrows localize the position in the alignment, and in brackets the corresponding position in the rat TRPV1 primary sequence.

Figure 2 - Figure supplement 2. Different strategies of alignment reveal the same highly conserved residues. (a) Stacked histograms showing the amino acidic probability in each position on the MAFFT alignment. Gray boxes depict the trans-membrane helices (TM1-TM6) and features such as pre-TM1, the TM4-TM5 linker (L), the Selectivity Filter and Pore Helix (SF & PH) and the TRP domain helix (TDH). Numbers over the arrows localize the position in the respective alignment, and in brackets the corresponding position in the rat TRPV1 primary sequence. (b) Shannon entropy of the amino acid distribution corresponding to each position in the alignment; the calculation is carried out using the emission probabilities from a Hidden Markov Model trained on the multiple sequence alignment. Low entropy values indicate conserved positions.

Figure 2 - Figure supplement 3. Sequence logos for the TRP family and the analyzed subfamilies, depicting highly conserved residues (>90% identity) for the MAFFT alignment.
Figure 3. Spatial distribution of TRP channel signature residues. (a) Conservation rates for each position in the alignment, calculated on Consurf (see methods), mapped on rTRPV1 structure (PDB: 7LP9) (b) Highly conserved (>90%) residues are arranged in three well-defined patches, highlighted as insets and dubbed P1, P2, and P3. The structural data and residue numbering corresponds to rat TRPV1 (PDB: 7LP9). For clarity only one protomer is shown. Backbone and residues follow the code color used in Figure 2b. 4-5L: TM4-TM5 linker; SF&PH: selectivity filter and pore helix; TRPh: TRP helix. (c) Structural alignment performed over representative channels (rTRPV1, PDB: 7LP9; mTRPC5, PDB: 6AEI; pmTRPM8, PDB: 606A; hTRPA1, PDB: 3J9P; dmTRPN1, PDB: 5VKQ; CrTRP1, PDB: 6PW4) reveals a consistency in the position of signature residues. (d,e) Distogram of mean distances (d) and normalized variance of mean distances (e) between pair of residues on transmembrane segments, revealing the proximity of signature residues of same patches (brighter areas in (d)) and the low variability on the distances of the same pairs (brighter areas in (e)). Blue, green, and red lines identify the P1, P2 and P3 residues, respectively, and squares locate the intersection between these residues.

Figure 3 - Figure supplement 1. Position of signature residues in Structural alignment. By mapping the signature residues in the structural alignment released in Huffer et al., 2021, we further confirmed the high level of conservation in the 3D position of the signature residues. Red letters identify residues with different identities compared with the signature. Red and blue shade boxes depict a shift to amino or carboxyl direction with respect to white boxes in the structural alignment. These shifts arouse from displacements throughout the helix axis or its rotation in each particular structure. In 87.8% of the structures it is necessary to use one-position shift to help coincide with the alignment of primary sequences. In yellow are the positions where there are no aligned residues (gap).

Figure 3 - Figure supplement 2. Coevolution analysis. Residue pairs with high coevolution scores (top 5%) are connected by red lines. Coevolution scores were calculated using an asymmetric pseudo-likelihood maximization direct coupling analysis algorithm (aplmDCA). Signature residues are drawn in blue licorice representation (rTRPV1, PDB: 7LP9; paTRPM8, PDB: 606A; mTRPC5, PDB: 6AEI).

Figure 3 - Figure supplement 3. Stacked histograms showing the amino acidic probability in each position on MAFFT alignment. (a) Complete sequence histogram. [b]. Parsed alignment used for building distance matrices. The parsed alignment contains highly conserved residues with a gap frequency <0.01. Gray boxes depict the transmembrane helices (TM1-TM6) and features like pre-TM1, the TM4-TM5 linker (L), the Selectivity Filter and Pore Helix (SF&PH) and the TRP domain helix (TDH). Numbers over the arrows localize the position in the respective dataset, and in brackets the corresponding position in the rat TRPV1 primary sequence.
Figure 3 - Figure supplement 4. Fingerprint residues remain at close distance. (a) frequency histogram depicting the distribution of pairwise distances in all analyzed structures for all the residues analyzed in the distograms (black) and also the fingerprint residues (red). (b) Individual frequency histograms for the mean distances depicted in the distogram presented in Figure 3d.

Figure 4. Aromatic residue distribution in LBD. (a) Aromatic residues facing the internal space shared by the 4 first transmembrane helices (core). The interacting aromatic (distance<5Å) in rTRV1 (PDB:7LP9) and rKv1.2 (PDB:2R9R) are depicted in blue. In violet residues with no other aromatic at <5Å. Right: surface representation of the sidechain of aromatic residues shown as licorice in the left. (b) Histogram of aromatic residues in the alignment, on the positions facing the core. At the bottom are depicted the positions in the alignment, and a (+1) or (-1) indicates that in one of the subfamilies the aromatic is immediately after or before the labeled position (shared for the rest of the subfamilies). (c) Comparison between AC volumes presented next to a schematic view of the topology obtained in our phylogenetic analysis. Blue: aromatic residues >50% conserved in the respective subfamily; red: aromatic residues >50% conserved in the respective subfamily and signature residue; black: not conserved residue present in the used structure; orange: not aromatic residue in the used structure, but present as an aromatic in >50% in the respective subfamily. Inset: Aromatic core in rTRPV1. The specific positions of the aromatics are indicated. Used structures: rTRPV1, PDB:7LP9; mTRPC5, PDB:6AEI; pmTRPM8, PDB:606A; hTRPA1, PDB:3J9P; CrTRP1, PDB:6PW4; mTPC1, PDB:6C96.

Figure 4 - Figure supplement 1. Characterization of Aromatic Core. (a) Upper: Histogram of the number of aromatic residues contained in the bigger cluster of each structure, fitted to a Gaussian function (center at x=6.78 and width=2.87). Bottom: Average size of the core per subfamily. Graph shows means+SEM. The size sample for each subfamily depends on the structure files availability (TRPs n=128; TRPA n=8; TRPC n=10; TRPM n=28; TRPV n=80) (b) Upper: table of size of the larger cluster on non-TRP channels. Bottom: Hatch pattern showing the threshold of p<0.05 for the fitted Gaussian curve.

Figure 4 - Figure supplement 2. Distograms of mean distances and normalized variance of mean distances between pairs of residues within the TM1-TM4 region. (a) Mean distance and variance distograms obtained for pan-TRP. (b) frequency histogram for pairwise distances in all analyzed structures highlighting the AC residues (colored bars). (c-e) Distograms corresponding to TRPV, TRPC, and TRPM subgroups. Blue lines depict the position of the conserved aromatics (>50%) for each subfamily. (f-h) Individual frequency histograms for the pairwise distances depicted in the corresponding insets and the distgrams c-e.
Figure 5. The AC connects to aromatics in P2 from neighboring subunits. (a) A conserved intermolecular connection between residues (licorice) in helices at opposite faces to the AC (gray surface) and P2 (yellow surface). Inset: Upper view of residues establishing the inter-subunit interaction (rTRPV1, PDB:7LP9). (b) Sequence logos showing the position of residues involved in the putative intermolecular interaction in blue, and the fourth signature residue in orange.

Figure 5 - Figure supplement 1. Inter-subunit interaction is conserved in subfamilies. (a) Residues connecting TM4 and TM5 from neighboring subunits align on the same position in TM4 and at one or two positions from the 4th signature residue in TM5. (b) Lateral view of the interaction between LBD and PD from different subunits for two channels from different subfamilies.

Figure 6. Conserved modifications observed in TRP proteins. The different channels studied in this work are presented next to a schematic view of the topology obtained in our phylogenetic analysis. Unique TRP features are highlighted. Previous observations confirmed here are indicated in pink shades. Novel observations from the present work are indicated in gray shades. Lines represent presence while crosses represent absence or loss.
Reference List


Sierra-Valdez F, Azumaya CM, Romero LO, Nakagawa T, Cordero-Morales JF. 2018. Structure–function analyses of the ion channel TRPC3 reveal that its cytoplasmic domain allosterically modulates channel gating. *Journal of Biological Chemistry* **293**:16102–16114. doi:10.1074/jbc.RA118.005066


related channels

conserved & contacting aromatics within TM1-4

TRP channel
TM fingerprint

presence of TDh & connection with 4-5L

modified connectivity between TDh & 4-5L

AC connects TM1-4 using less aromatics

trivia

TRPA

presence of elongated AC & putative connection between P1 and P3

aromatics connect only three TM helices

TRPV

extended TM3-TM4 loop

TRPC

extended linker between preTM1 and TM1

only a portion of AC & bridge disconnected

crTRP1

TRPML2

bridge
no-aromatics

TPC1 A-A

bridge
no-aromatics

TPC1 A-B

bridge
unstructured AC

PKD2

bridge & AC