Running title: Structural model of a2N V-ATPase and its binding interface for CTH2

Structural model of a2-subunit N-terminus and its binding interface for Arf-GEF CTH2: Implication for regulation of V-ATPase, CTH2 function and rational drug design

Vladimir Marshansky[1,*], Hiroyuki Hosokawa[2], Maria Merkulova[2], Anastasia Bakulina[3], Phat Vinh Dip[4], Youg Raj Thaker[4], Ajay Bjargava[5], James R. Tonra[6], Dennis A. Ausiello[7], Gerhard Grüber[4]

[1] Neuro-Horizon Pharma, Sharon, MA, 02067, USA

[2] Center for Systems Biology, Program in Membrane Biology, MGH, Boston, MA, 02114, USA

[3] Novosibirsk State University, 630090, Novosibirsk, Russia

[4] Nanyang Technological University, School of Biological Sciences, Singapore, 637551, Republic of Singapore

[5] The Francis Crick Institute, London, NW1 1ST, UK

[6] BeyondSpring Pharmaceuticals, New York, 10005, NY, USA

[7] Center for Assessment for Technology and Continuous Health, MGH, Boston, MA, 02114, USA

* Corresponding author: Vladimir Marshansky MD, PhD

E-mail address: vlad@neurohorizonpharma.com; Tel: (617) 797-4364

ABSTRACT

We have previously identified the interaction between mammalian V-ATPase a2-subunit isoform and cytohesin-2 (CTH2) and studied molecular details of binding between these proteins. In particular, we found that six peptides derived from the N-terminal cytosolic domain of a2 subunit ($a2N_{1-402}$) are involved in interaction with CTH2 (Merkulova, Bakulina, Thaker, Grüber, & Marshansky, 2010). However, the actual 3D binding interface was not determined in that study due to the lack of high-resolution structural information about a-subunits of V-ATPase. Here, using a combination of homology modeling and NMR analysis, we generated the structural model of complete $a2N_{1-402}$ and uncovered the CTH2-binding interface. First, using the crystal-structure of the bacterial *M. rubber* I_{cyt}-subunit of A-ATPase as a template (Srinivasan, Vyas, Baker, & Quiocho, 2011), we built a homology model of mammalian $a2N_{1-352}$ fragment. Next, we combined it with the determined NMR structures of peptides $a2N_{368-395}$ and $a2N_{386-402}$ of the C-terminal section of $a2N_{1-402}$. The complete molecular model of $a2N_{1-402}$ revealed that six CTH2 interacting peptides are clustered in the distal and proximal lobe sub-domains of $a2N_{1-402}$. Our data indicate that the proximal lobe sub-domain is the major interacting site with the Sec7 domain of first CTH2 protein, while the distal lobe sub-domain of $a2N_{1-402}$ interacts with the PH-domain of second CTH2. Indeed, using Sec7/Arf-GEF activity assay we experimentally confirmed our model. The interface formed by peptides $a2N_{1-17}$ and $a2N_{35-49}$ is involved in specific interaction with Sec7 domain and regulation of GEF activity. These data are critical for understanding of the cross-talk between V-ATPase and CTH2 as well as for the rational drug design to regulate their function.

INTRODUCTION

The V-ATPases are ubiquitous proton pumps, that use the energy of ATP to translocate protons from cytosol to intracellular compartments or extracellular space. V-ATPases maintain pH homeostasis at the cellular and the whole organism level, and also play a critical role in cellular function via direct interaction with a variety of proteins, whose functions are unrelated to pH homeostasis. The direct and indirect roles of V-ATPase were previously reviewed by us (Marshansky & Futai, 2008; Marshansky, Rubinstein, & Grüber, 2014) and others (Forgac, 2007; Hinton, Bond, & Forgac, 2009).

V-ATPases are very complex multi-subunit enzymes that function as proton-pumping rotary nano-motors (Marshansky & Futai, 2008; Marshansky, Rubinstein, & Grüber, 2014). Functional V-ATPases are composed of two parts: a cytoplasmic V_1- and a transmembrane V_O-sector, which may dissociate from each other in response to some stimuli. This dissociation results in reduced ATPase activity and a shutting down of proton translocation of V-ATPase, and is one of the main mechanisms of down-regulation of V-ATPase function (Marshansky & Futai, 2008; Marshansky, Rubinstein, & Grüber, 2014). Each of the two sectors are composed of multiple different subunits. In mammalian cells eight different proteins are combined in the following stoichiometry $A_3B_3C_1D_1E_3F_1G_3H_1$ to form a V_1-sector; while at least six different proteins $a_1c_5c"_1d_1e_1Ac45_1$ form the transmembrane V_O-sector. Proton pumping across the membrane occurs via coupling of ATP-hydrolysis with rotary-mechanism of protons translocation in the interface between the rotating c-ring and the stationary a-subunit of V-ATPase. The rotation of the c-ring is driven by ATP hydrolysis catalyzed by the stationary A_3B_3 headpiece of the V_1-sector. In order to accomplish an efficient transfer of

ATP hydrolysis energy to c-ring rotation, the rotation of A_3B_3 headpiece is prevented by three peripheral stalks (Zhang et al., 2008). Each of these stalks is composed by a G/E heterodimer, which anchors the A_3B_3 headpiece to the membrane through either: i) direct binding to the N-terminal tail of a-subunit, or ii) indirectly through subunit C. In particular, while two of three G/E-stalks are directly bound to a-subunit, the third G/E-stalk is bound to subunit C, which itself interacts simultaneously with a-subunit and the second G/E-peripheral stalk (Oot & Wilkens, 2012; Oot et al., 2017; Zhang et al., 2008). It is noteworthy that during the disassembly of V_1-sector from V_O-sector, the protein-protein interactions between a-subunit and G/E-stalks as well as C-subunit and G/E-stalk are destabilized through a yet unknown molecular mechanism (Oot et al., 2017).

In mammals there are four a-subunit isoforms, (a1, a2, a3 and a4) that contain two major domains: a cytosolic N-terminal domain (aN ~ 400 aa) and a membrane-integrated C-terminal domain (aC ~ 400 aa), containing eight transmembrane spanning helices (Marshansky, 2007; Toei, Toei, & Forgac, 2011). While a-subunit isoforms are highly homologous to one another, nevertheless they perform non-redundant functions. Initially we found, that the cytosolic N-terminal domain of a2-subunit ($a2N_{1\text{-}402}$) directly interacts with cytohesin-2 (CTH2), which acts as Arf-GEF (GDP/GTP exchange factor) and activator of Arf-family small GTPases. Since CTH2 and Arf small GTPases are the key regulators of receptors signaling, vesicular trafficking and actin cytoskeleton rearrangement (Hurtado-Lorenzo et al., 2006; Merkulova et al., 2010), these findings provided new insights into the functional link between V-ATPase dependent organellar acidification and vesicular trafficking. Importantly, our recent study revealed that the other three a-subunit isoforms (a1, a3 and a4) also bind to CTH2, suggesting its

ubiquitous nature and cell biological significance of signaling between V-ATPase and CTH2 (Merkulova et al., 2011).

Cytohesin-2 (CTH2, also known as ARNO) together with cytohesins-1, -3 and -4, belongs to cytohesin subfamily of Arf-GEFs, activators of Arf small GTPases (Casanova, 2007). Cytohesins are highly homologous proteins composed of four distinct structural domains: i) an N-terminal coiled-coil domain; ii) a central Sec7-domain; iii) a pleckstrin homology (PH) domain; and iv) a C-terminal polybasic domain. In our previous study we addressed the molecular details of binding between CTH2 and $a2N_{1-402}$ (Merkulova et al., 2010) and uncovered very complex interactions between these proteins. We found multiple binding sites within the Sec7 domain of CTH2 as the strongest interaction sites with a-subunit of V-ATPase (Merkulova et al., 2010). Moreover, homology modeling of CTH2 performed in our study also suggested the multisite binding and complex character of interactions between these two proteins (Merkulova et al., 2010). Indeed, pull-down experiments using synthetic peptides demonstrated that six $a2N_{1-402}$ specific peptides $a2N_{1-17}$, $a2N_{35-49}$, $a2N_{198-214}$, $a2N_{215-230}$, $a2N_{313-331}$ and $a2N_{386-402}$ are involved in the interaction with full-length CTH2 (Merkulova et al., 2010). Intriguingly, these peptide motifs were not clustered but instead randomly distributed throughout the sequence of $a2N_{1-402}$ of V-ATPase (Merkulova et al., 2010). Additional pull-down and surface plasmon resonance experiments with purified domains of CTH2 revealed that peptide $a2N_{1-17}$ interacted specifically and strongly with Sec7 domain, while peptide $a2N_{198-214}$ interacted with PH domain (Merkulova et al., 2010). This data suggested that the epitope formed by $a2N_{1-17}$ is most likely involved in interaction with Sec7 domain, while the epitope formed by $a2N_{198-214}$ is probably part of a different binding site, involving the PH-

domain of CTH2. However, this hypothesis could not be verified due to the lack of structural features of the $a2N_{1-402}$ subunit of V-ATPase.

Previously, a high-resolution crystal structure of cytosolic N-terminus of I-subunit (I_{cyt}) of *Meiothermus ruber* A-ATP synthase was solved (Srinivasan et al., 2011). This protein is homologous to yeast Vph1 and mammalian a-subunit isoforms of V-ATPase. Overall, I_{cyt} is composed of a curved long central α-helical bundle capped on both ends by two lobes with similar α/β architecture that shows remarkable structural similarity with yeast C-subunit of V-ATPase (Srinivasan et al., 2011). The crystal structure of I_{cyt} also fits well into the corresponding place in an electron microscopy density map of intact A-ATP synthase from *Thermus thermophilus* bacteria (Lau & Rubinstein, 2012; Srinivasan et al., 2011). Importantly, extensive analysis of the multiple alignments of all currently known a-subunits of V-ATPase demonstrated the evolutionarily conserved character of these proteins and their structures. This in turn opens the possibility to generate homology models of N-terminal cytosolic domains of a-subunit isoforms from other species including mammals.

Here, we used the crystal structure of I_{cyt} from *Meiothermus ruber* to build a homology model of mouse a2-subunit ($a2N_{1-352}$) of V-ATPase. Moreover, we also solved the NMR structures of two synthetic overlapping peptides derived from $a2N_{339-402}$ in order to determine the structural features of this unknown part of the protein and to build a structural model of the entire $a2N_{1-402}$ cytosolic tail. Using this model, we determined the distribution of CTH2 interacting epitopes clustered in two spatially separate binding regions on 3D structural model of $a2N_{1-402}$. We also outlined and analyzed a G/E-stalks binding interfaces on $a2N_{1-402}$ based on previously published data from the Forgac

laboratory (Qi & Forgac, 2008). Our data suggest the complex interaction between a2-subunit V-ATPase and CTH2 that may modulate binding between a2-subunit and G/E-subunits of the peripheral stalks. We propose this interplay between these interactions as a molecular mechanism of cross-talk between trans-membrane V-ATPase complex, cytosolic CTH2 and Arf small GTPases.

RESULTS

The structure of *Meiothermus ruber* I_{cyt} subunit as a template for mouse a2N$_{1-352}$ homology model. The first high-resolution structure of A-ATP synthase *M. ruber* I_{cyt} has been solved previously (Srinivasan et al., 2011), which opened the possibility to generate homology models of cytosolic N-terminal domains of a-subunits from other species. We performed multiple amino acid sequence alignment of a-subunit homologues from various species in order to assess a general overview of a-subunit evolution. This analysis revealed that high level conservation of amino acid sequences of cytosolic N-terminal tail of a-subunits from different species is not required for preservation of the overall secondary and tertiary structures. Thus, we concluded that I_{cyt} could be used as a template to build a reliable model of 3D structure of the mouse a2N protein.

The homology model of a2N$_{1-352}$ structure. The homology modeling of a2N$_{1-352}$ was performed using the crystal structure of I_{cyt} from *M. ruber* (PDB ID: 2RRK) as a template as described in Methods. The crystal structure of *M. rubber* I_{cyt} represents the amino acids 1-301 (I_{cyt_1-301}), whereby the C-terminal 50 amino acids of I_{cyt} including the

residues 302-351 ($I_{cyt\_302-351}$) remained unresolved (Srinivasan et al., 2011). Since I_{cyt_1-301} is an independently folded structural unit, it most likely represents the separate domain of *M. ruber* I subunit. In the mouse a2N sequence this separate domain corresponds to $a2N_{1-352}$ region (Figure 1). The homology-modeled structure of mouse $a2N_{1-352}$ adopts the same overall shape of a barbell as its I_{cyt} template and contains the same three major structural elements (Figure 2a-c, Supplementary Movie S1). The domain composition of $a2N_{1-402}$ is shown in Figure 2a and the color and amino acids composition of the structural elements also correspond to the ribbon structures shown in Figure 2b-c, respectively. The proximal lobe (*PL*) with the mixed α/β architecture is formed by two distinct amino acid regions 1-42 aa (*PL*, red) and 322-352 aa (*PL*, red). The distal lobe (*DL*) with the similar mixed architecture is formed by a single continuous amino acid stretch 173-244 aa (*DL*, blue). The central bar domain (*BD*) is also formed by two distinct amino acid regions 43-172 aa (*BD*, green) and 245-321 aa (*BD*, green) which contains two anti-parallel rows of long α-helices, that link two lobes together (Figures 1 and 2b,c). The proximal lobe consists of 4 anti-parallel β-strands (Figure 1, strands 1, 2, 6 and 7) and 2 α-helices (Figure 1, helices I and XI) that fold into two-layer α/β sandwich. The strands-2 and 7 are not visible in the model since these regions assigned by PyMOL as loops, Swiss-PDB Viewer (Guex, Peitsch, & Schwede, 2009) assigns β-strands for these regions, while PsiPred predicts β-strand for strand-2 (Figure 1, in blue). Thus we suggest that this is not a significant difference between the model and the template. The distal lobe is composed of 3 antiparallel β-strands (Figure 1, strands 3, 4 and 5) and 2 α-helices (Figure 1, helices VI and VII) that fold into similar two-layer α/β sandwich. The 'extra' 51 residues of a2N are predominantly located in very N-terminus (additional 9 aa) and

contribute to the formation of the longer loops between α-helices II and III, α-helix V and β-strand 3, and between–β-strands 4 and 5 (Figures 1 and 2). Thus the overall structure of $a2N_{1\text{-}352}$ model is very similar to the crystal structure of I_{cyt}.

The NMR structure of $a2N_{353\text{-}402}$ region. The cytosolic N-terminal tail of mouse a2-subunit ($a2N_{1\text{-}402}$) has been predicted to span 1-402aa (Figure 1) (Merkulova et al., 2010). Thus the homology model of $a2N_{1\text{-}352}$ is not a complete part of the cytosolic N-terminal tail, since the last 50 aa ($a2N_{353\text{-}402}$) could not been modeled using $I_{cyt\_1\text{-}301}$ as a template. Since, this part of $a2N_{1\text{-}402}$ is connecting the rest of the tail to its predicted first transmembrane domain, we called this part of the protein the stem domain (*SD*) (Figure 2a and 3a, 353-402 aa in white). In addition, the region 386-402 of mouse V-ATPase subunit a2-isoform has been shown to be essential for binding to CTH2 Arf GEF small GTPase (Merkulova et al., 2010). Here were synthesized the overlapping peptides $a2N_{339\text{-}395}$ and $a2N_{386\text{-}402}$, which were derived from the $a2N_{339\text{-}402}$ region (Figure 3a, in cyan). The solution structure of entire $a2N_{386\text{-}402}$ (in cyan) and part of $a2N_{339\text{-}395}$ (in cyan), shown as $a2N_{368\text{-}395}$ (in gray), were solved by NMR spectroscopy. Amino acids of the peptides were sequentially assigned using both NOESY and TOCSY data. 2D TOCSY and 2D NOESY raw data was processed using in-built Topspin software (Bruker). Figure 4a-b show the assigned NH region of the 2D NOESY spectrum of $a2N_{339\text{-}395}$ and $a2N_{386\text{-}402}$ (Figure 3a, in cyan). Primary sequence amino acid marking was followed by the assignment of cross peaks by overlaying 2D TOCSY and 2D NOESY spectra (Figure 4a-b). The secondary structure elements of $a2N_{339\text{-}395}$ were analyzed based on $^{1}H\alpha$ chemical shifts with respect to the random coil values. In case of $a2N_{339\text{-}395}$ the data indicate an α-

helical formation between the amino acids N_{373} to Y_{386} (Figure 4c). Identified cross peaks in the HN-HN region are shown in Figure 4e, indicating α-helical features of $a2N_{339-395}$. HN–HN, Hα–HN(i, i+3), Hα–HN(i, i+4), and Hα–Hβ (i, i+3) connectivities were plotted from the assigned NOESY spectrum (Figure 4e). Hα–HN(i, i+3), indicating α-helical formation between the residues G_{378} to Y_{387}, whereby Hα–HN(i, i+2) between T_{372} and F_{375} indicates a potential α-helical turn. Data from assigned 2D NOESY spectra, torsion angle calculated from HA values by TALOS software and primary amino acid sequence were used as input for the automated structure calculation by Cyana 3.0 package (Herrmann, Güntert, & Wüthrich, 2002). Since the 2D NOESY assignment of $a2N_{339-395}$ reveals an unstructured N-terminal part, the calculation was performed only for the C-terminal region with the residues 368 to 395, called $a2N_{368-395}$ (Figure 3a, in grey). Out of 100 generated structures the 20 lowest energy structures were taken for further analysis. In total an ensemble of 20 calculated structures resulted in an overall mean root square deviation (RMSD) of 0.297 Å for the residues 378-387 (Figure 3b). All the structures of $a2N_{368-395}$ have energies lower than -100 kcal mol^{-1}, no NOE violations greater than 0.3 Å and no dihedral violations greater than 5º. The statistics for 20 structures are shown in Figure 3b. The structure of $a2N_{368-395}$ shows a total length of 43.05 Å and forms a flexible N-terminal region from residues 368-371 with a helix extending from amino acids 372-375, followed by a short loop from 376-378 and a second helix between the residues 379-387 (Figure 5a,b), while the remaining C-terminal region is flexible. Molecular surface electrostatic potential of $a2N_{368-395}$ reveals an amphiphilic surface (Figure 5c,d). At one side of the peptide, residues E_{377}, D_{384} and E_{393} are forming a negative charged surface (Figure 5c). The opposite side reflects an amphiphilic surface potential, formed by the

charged residues at the N- and C-termini as well as the hydrophobic residues F_{375}, F_{379}, I_{382}, V_{383}, A_{385}, Y_{386} V_{388}, and Y_{391} (Figure 5d).

In parallel, the secondary structure elements of $a2N_{386-402}$ showed α-helical formation between the amino acids P_{396} to I_{402} (Figure 6a,b). HN–HN, Hα–HN(i, i+3), Hα–HN(i, i+4), and Hα–Hβ (i, i+3) connectivities were plotted from the assigned NOESY spectrum (Figure 4f) and reflecting α-helical formation in the C-terminus. In total an ensemble of 20 calculated structures resulted in an overall mean root square deviation (RMSD) of 0.68 Å (Figure 3b). All these structures have energies lower than -100 kcal mol^{-1}, no NOE violations greater than 0.3 Å and no dihedral violations greater than 5º (Figure 3b). The $a2N_{386-402}$ peptide contains an unstructured N-terminal region formed by the amino acid residues 386 to 395 and a C-terminal α-helical region from residues 396 to 402 (Figure 6b). Since $a2N_{386-402}$ in the intact protein is predicted to be locate on the border of the cytoplasmic and membrane-embedded parts of a2-subunit (Merkulova et al., 2010), we propose that this C-terminal helix of $a2N_{386-402}$ corresponds in fact to the beginning of the first transmembrane helix of the a2-subunit of V-ATPase.

The complete model of $a2N_{1-402}$. In order to build a 3D model of the entire $a2N_{1-402}$ region, the homology model of $a2N_{1-352}$ was combined with NMR structures of the two peptides $a2N_{368-395}$ (Figure 5) and $a2N_{386-402}$ (Figure 6) as described in Methods (Figure 7, Supplementary Movie S2). In comparison with the $a2N_{1-352}$ structural model, three additional α-helices can be assigned in the $a2N_{1-402}$ model (Figure 1, helices XII, XIII, and XIV), which are connected by unstructured linkers (Figure 7a,b,d). According

to $a2N_{1-402}$ model only α-helix XIII lays close to α-helix I of $a2N_{1-352}$ region. Overall, $a2N_{353-402}$ region represents an appendage joined to the $a2N_{1-352}$ structural core through a flexible linker formed by amino acids 353-371 between β-strand 7 and α-helix XII (Figure 1). This appendage will most likely contributes to the formation of the short stem domain (SD), which connects the cytosolic and membrane-embedded parts in intact full-length a-subunit of V-ATPase as observed on electron microscopy images (Lau & Rubinstein, 2010, 2012).

The calculated $a2N_{1-402}$ molecular model can be fitted well into the density of the barbell-shaped collar of the 23 Å resolution EM map of the related *P. furiosus* A-ATP synthase (EMD-1542)(Vonck, Pisa, Morgner, Brutschy, & Müller, 2009) with a correlation coefficient being 0.8702, as calculated using the Chimera program [30] and as shown Figure 8a-b. The fit takes into consideration that the regions $a2N_{140-173}$ and $a2N_{206-222}$, which are segments of the eukaryotic V-ATPase subunit and absent in *M. ruber* A-ATP synthase I_{cyt} template, are presented as unstructured regions in the final $a2N_{1-402}$ model (Figure 7a,c,d). In addition, the structures of the subunits E and H of the *T. thermophilus* and *Pyrococcus horikoshii* OT3 A-ATP synthase are well accommodated by the density of the straight and bent peripheral stalk with a correlation of 0.7827 and 0.7704, respectively, as described most recently [32]. The structure of the H^+-translocating *c*-ring (Murata, Yamato, Kakinuma, Leslie, & Walker, 2005; Vonck et al., 2009) was fitted with a correlation coefficient being 0.8601.

The structural model and features of CTH2 binding interface on a2-subunit of V-ATPase. While six CTH2 interacting peptides are evenly distributed in amino acid sequence of $a2N_{1-402}$ (Merkulova et al., 2010), our complete $a2N_{1-402}$ model revealed that epitopes formed by these peptides are forming just two spatially distinct binding sites for two molecules of CTH2 (Figure 7a-d). The CTH2 interacting peptides $a2N_{1-17}$, $a2N_{35-49}$, $a2N_{313-331}$ and $a2N_{386-402}$ are clustered together in the proximal lobe (*PL*) into the first binding site (Figure 7a,d left and Figure 7b). Three of these peptides $a2N_{1-17}$, $a2N_{35-49}$ and $a2N_{313-331}$ form a continuous surface from the β-strands 1, 2 and 6, and the adjacent loops (Figure 1 and Figure 7a,b,d), while $a2N_{386-402}$ peptide is separate and forms an additional binding surface (Figures 1, 5a and 5b Figure 7a,d left and Figure 7b). The second CTH2 binding site is composed by $a2N_{198-214}$ and $a2N_{215-230}$ peptides which form a continuous surface from the β-strands 4 and 5, and the loop between them in the distal lobe (*DL*) of $a2N_{1-402}$ (Figure 1 and Figures 7a,d right and c). Thus, both binding sites have a very similar predominantly β-sheet architecture, except for the presence of an additional unique α-helical region XIV of $a2N_{386-402}$ peptide in the proximal lobe binding interface. Importantly, according to $a2N_{1-402}$ model, both sites are not buried inside the protein molecule but are exposed and readily accessible to binding to CTH2.

Mapping of the G/E peripheral stalk interaction sites on structural model of $a2N_{1-402}$. Previously, Forgac and co-workers performed extensive site-specific cross-linking experiments with the N-terminal domain of yeast Vph1p, a homolog of mammalian V-ATPase a-subunit (Qi & Forgac, 2008). Following five amino acid residues in Vph1 G_{181}, S_{266}, A_{347}, A_{351} and Q_{369} that are located in proximity or

interacting with G/E peripheral stalks of V-ATPase were identified (Qi & Forgac, 2008). We applied this data to $a2N_{1-402}$ structural model, in order to reconstitute the G/E-binding interface on the mouse a2-subunit of V-ATPase. According to multiple sequence alignment, G_{181}, S_{266}, A_{347}, A_{351}, and Q_{369} amino acids of Vph1p correspond to G_{170}, T_{255}, P_{336}, R_{340} and N_{358} in mouse a2N sequence (Figure 1). In the structural model of $a2N_{1-402}$ the G_{170} residue is located in the unstructured loop, connecting α-helix V of the central region and first β-strand 3 of the distal lobe (Figure 1 and 7d, right). The residue of T_{255} is in close proximity and a forming part of the short predicted α-helix connecting unstructured region behind the last α-helix VII of the distal lobe and α-helix VIII of the central bar region (Figure 1 and 7d, right). Thus, both amino acids G_{170} and T_{255} come close together in the structural model and most likely are forming one G/E-binding interface in the distal lobe of $a2N_{1-402}$ (Figure 7d, right). The next two of five residues P_{336} and R_{340} are located in the proximal lobe of $a2N_{1-402}$ forming a single continuous interface in the same α-helix XI (Figure 1 and 7d, left). In close proximity to this surface is residue N_{358}, which is located in the middle of the unstructured linker that connects core and appendage domains of $a2N_{1-402}$. Thus, all three amino acids P_{336}, R_{340} and N_{358} come close together in the structural model and most likely are forming another G/E-binding interface on the proximal lobe of $a2N_{1-402}$ (Figure 7d, left).

Identification of V-ATPase interface involved in binding with CTH2 and modulation of its GEF activity. According to our studies, two V-ATPase derived peptides $a2N_{1-17}$ and $a2N_{35-49}$ are forming a specific interaction interface on proximal lobe (PL_{SD}) sub-domain of a2N that is involved in interaction with Sec7 domain of CTH2

(Figure 8a, shown by gray rectangle). To confirm our model, we therefore performed an additional GEF activity experiments using recombinant Sec7 CTH2 domain and a2N V-ATPase derived synthetic peptides (Bio-a2N$_{1-17}$ -TAT; Bio-a2N$_{18-34}$ -TAT; Bio-a2N$_{35-49}$ -TAT; Bio-a2N$_{75-91}$ -TAT and Bio-a2N$_{313-331}$-TAT)(Figure 8g). Experimental testing of our model demonstrated a potent inhibition of Sec7/Arf1-GEF activity by both Bio-a2N$_{1-17}$-TAT (IC_{50} = 1.5 μM) (Figure 8b) and Bio-a2N$_{35-49}$-TAT (IC_{50} = 0.9 μM)(Figure 8c), peptides forming an interface highlighted in gray rectangle on Figure 8a. However, peptides of a2N V-ATPase that are not involved in formation of a binding-competent interface: Bio-a2N$_{18-34}$-TAT (Figure 8d), Bio-a2N$_{75-91}$-TAT(Figure 8e), Bio-a2N$_{313-331}$-TAT (Figure 8f), are not able to modulate the function of CTH2 Sec7 domain in a Sec7/Arf1-GEF activity assay. These data indicate that the interface formed by a2N$_{1-17}$ and a2N$_{35-49}$ of V-ATPase (Figure 8a) is specifically involved in the interaction with Sec7 domain and regulation of enzymatic GEF activity of CTH2.

DISCUSSION

In this study we generated the structural model of a2N$_{1-402}$ and uncovered two binding interfaces for Arf GEF CTH2. The complete molecular model of a2N$_{1-402}$ revealed that all CTH2 interacting peptides are located in distinct regions of the proximal lobe (*PL*) and distal lobe (*DL*) of a2N$_{1-402}$ and there are no interacting epitopes in the central bar domain (*BD*) of the protein (Figure 7). The average distance between two CTH2 binding sites on a2N$_{1-402}$ was estimated as ~ 73 Å. The first CTH2 binding interface located in *PL* is formed in part by the externally exposed a2N$_{1-17}$, and a2N$_{35-49}$ peptides (Figure 7a,b, Figure 9). According to our previous studies these two peptides are

specifically interacting with Sec7 domain of cytohesin-2 (Merkulova et al., 2010). Therefore, here we conclude that this first *PL* located binding interface is the major CTH2 binding site, which is involved in interaction with catalytic Sec7 domain of first CTH2. Two other peptides $a2N_{313-331}$ and $a2N_{386-402}$ are also located in this region. However, these peptides do not interact with Sec7 domain, while they are interacting with full-length CTH2 in pull-down experiments (Merkulova et al., 2010). We conclude, that epitopes formed by these peptides are most likely involved in the interaction with regulatory elements of CTH2 such as Sec7/PH-linker, PH- and/or PB-domains, which are located in the close proximity to the Sec7 domain (Merkulova et al., 2010). Finally, based on our previous data showing low affinity interaction of peptide $a2N_{198-214}$ with purified PH-domain, we hypothesized that the second *DL* located epitope should be involved in interaction with only regulatory PH-domain of second CTH2 (Figure 7a,c, Figure 9).

What would be the molecular role and cell biological functional significance of such a complex and differential interaction between a2-subunit of V-ATPase and two molecules of CTH2? Previously, we reported that interaction between a2-subunit and CTH2 is depending upon V-ATPase driven acidification of the endosomal lumen which suggests, that V-ATPase itself could function as pH-sensing receptor (Hurtado-Lorenzo et al., 2006). Thus, we indeed identified V-ATPase as a novel signaling receptor that on one hand, is sensing the acidification status of endosomal lumen, and on the other hand, could transmit this information across the membrane and modulate activity of the Arf-GEF CTH2 and cognate Arf small GTPases. The Arf family of small GTPases (Arf1 - Arf6) belongs to the Ras-superfamily small GTPases that function as “molecular

switches" and regulate multiple cellular functions. CTH2 as an Arf-GEF is a regulatory protein, whose main function is the activation of Arf small GTPases and thus regulating multiple down-stream effectors. Previously, we suggest that cytohesin/Arf's might function as "molecular on/off switches" for self-regulation of the V-ATPase function (Marshansky & Futai, 2008). In particular we proposed that cross-talk and signaling between a2-subunit and CTH2 might be involved in self-regulated acidification-dependent disassembly of V-ATPase and shutting down of its function as a proton-pumping nano-motor (Marshansky & Futai, 2008). Here using structural insights from $a2N_{1\text{-}402}$ model we test this hypothesis.

Insights from structural model of $a2N_{1\text{-}402}$: Regulatory role of CTH2 in modulation of V-ATPase function. Importantly, our structural model of $a2N_{1\text{-}402}$ and identification of its interacting interfaces with both CTH2 and G/E-stalks provide basis for analysis of the potential molecular mechanism of self-regulation of V-ATPase function. Cryo-electron microscopy study of the intact holo-complex of V-ATPase demonstrates that a-subunit and C-subunit of V-ATPase are oriented parallel to the membrane surface forming so called "colar" separating V_1- and V_O-sectors of V-ATPase (Lau & Rubinstein, 2010, 2012). Thus, one of the major structural roles of the a-subunit as part of V-ATPase complex is to hold together V_1- and V_O-sectors during its rotational catalysis, in particular via its interaction with G/E peripheral stalks (Forgac, 2007). This model is supported by various cross-linking, two-hybrid, co-immunoprecipitation and electron microscopy experiments showing that cytosolic N-terminal tail of a-subunit directly interacts with A-, G-, E-, H- and C-subunits of V_1- sector of V-ATPase (Forgac,

2007). If the a-subunit is involved in such multiple interaction with other subunits of V_1-sector of V-ATPase, the question arises of how and when subunit *a* interacts with CTH2? One possibility is that CTH2 interacts with V_O-sector only after disassembly of the V_1/V_O-complex complex when a-subunit becomes more exposed for its binding. In this scenario CTH2 may share the same binding interfaces with other subunits of V_1-sector, such as G/E, competing for their binding with a-subunit, and thus, preventing the assembly of yet unassembled V-ATPase or the re-assembly of previously disassembled V-ATPase. In another scenario the interaction between CTH2 and a-subunit may take place in the intact V_1V_O-ATPase. In this case the G/E peripheral stalk subunits of V_1-sector and CTH2 would have different and not overlapping binding interfaces on $a2N_{1-402}$, which however, might trigger structural rearrangements upon CTH2 binding resulting in the disassembly of V_1/V_O-complex.

The structural model of $a2N_{1-402}$ and identification of its interacting interfaces with both CTH2 and G/E peripheral stalks helps to evaluate these possibilities. Our data demonstrated that there are two non-identical and non-overlapping sites on $a2N_{1-402}$ for both G/E stalks and CTH2. In case of G/E stalks, each of these two sites directly bind two out of three G/E peripheral stalks in intact V-ATPase complex (Zhang et al., 2008). The first G/E-binding interface is located in proximal lobe and is formed by α-helix XI and unstructured linker that connects core and appendage domains of $a2N_{1-402}$ (Figure 7d, left, *PL*). The second G/E-binding interface is located in distal lobe and is formed by the unstructured loop and the short predicted α-helix, preceding α-helix VII (Figure 7d, right, *DL*). Similarly, CTH2 has also got two distinct and non-overlapping binding sites.

While the CTH2 binding interface is in close proximity to the G/E interaction sites, they are not, however, identical since the G/E- and CTH2 binding motifs form separate non-overlapping interaction surfaces (Figure 7d). This non-overlapping character of CTH2 and G/E-binding surfaces identified in $a2N_{1-402}$ model suggests that CTH2 might be recruited and interacts with intact V_1/V_O-complex. This possibility was recently confirmed in our experiments with intact *S. cerevisiae* V-ATPase and a truncated form of CTH2 (Hosokawa et al., 2013). Thus our data indicate that CTH2 does not compete with G/E subunits for binding to a2-subunit of V-ATPase. However, its binding may compromise the stability of interaction between the a2-subunit and the G/E peripheral stalks. Thus, according to this model we conclude, that the acidification-dependent recruitment of CTH2 to V-ATPase previously uncovered in our studies (Hurtado-Lorenzo et al., 2006) is most likely involved in regulating disassembly of intact V_1/V_O-complex rather than being involved in prevention of assembly of V-ATPase (Figure 9a). We have also recently proposed an alternative mechanism of V-ATPase regulation by CTH2 (Dip, Saw, Roessle, Marshansky, & Grüber, 2012). According to this model, binding of the CTH2 to the proximal lobe of the N-terminal domain of Vph1 (yeast homolog of a-subunit) of V-ATPase could affect ion-translocation in the interface of the *c*-ring and the C-terminal membrane-embedded domain of a-subunit (Figure 9b).

In summary, combining of homology modeling and NMR structural analysis give rise to the structural model of $a2N_{1-402}$, which allow to uncover the molecular features of CTH2-binding interfaces on V-ATPase. Our model reveals the existence of two CTH2 binding sites located on the distal and proximal lobes of $a2N_{1-402}$. We also determined

that while Sec7-domain is a major interaction site on the proximal lobe of $a2N_{1-402}$, its distal lobe most likely interacts with only the PH-domain of CTH2. The further analysis revealed, that these two CTH2 binding sites are in a close proximity to two G/E-binding sites on subunit a. Thus, we hypothesize, that binding of CTH2 to the V-ATPase complex may destabilize binding of both G/E peripheral stalks to the a-subunit, triggering disassembly of V-ATPase into a V_1 and V_O-complex (Figure 9a) and/or shutting down proton pumping activity of the V-ATPase nano-motor (Figure 9b). However, this hypothesis is only the first step to uncover the molecular details of the interaction, interplay and signaling between V-ATPase, CTH2 and Arf family small GTPases. While recently performed multi-sequence alignment analysis of interaction sites between V-ATPase and CTH2 have uncovered the evolutionary conservation of their binding-mechanism (unpublished data), these models are awaiting experimental testing and confirmation using alternative molecular and cell biological approaches. Our final goal in uncovering these interactions is a structure-based design of small molecules that would specifically interfere with interactions between V-ATPase and CTH2. These small molecules can be used as drugs to modulate: i) either function of V-ATPase; or ii) signaling and function of CTH2 and Arf family small GTPases in variety of diseases, since both V-ATPase and CTH2/Arf6 small GTPases are emerging as an important drug targets (Marshansky, Rubinstein, & Grüber, 2014).

In particular, one of the subunits of eukaryotic V-ATPase was identified as (pro)renin receptor (PRR) playing a central role in the activation of the local renin-angiotensin system (RAS)(Danser, 2009; Ichihara & Kinouchi, 2011; Marshansky et al., 2014). Previously, it was uncovered that over-activation of PRR is strongly associated

with cardiovascular hypertensive and kidney diseases (Danser & Deinum, 2005; Ichihara, Kaneshiro, Takemitsu, Sakoda, & Itoh, 2007). Thus, future studies in this area could lead to the development of novel therapeutic approaches for the treatment of hypertension and its complications (Li, Zhang, & Zhuo, 2017; Sun, Danser, & Lu, 2017). Moreover, the V-ATPase is also emerging as an important drug target, able to control signaling and trafficking of another receptors such as EGFR/ErbB, Fz/LRP6 and Notch, and thus useful for treatment of lung cancers among others (Lu et al., 2013; Marshansky et al., 2014; McGuire, Cotter, Stransky, & Forgac, 2016; Stransky, Cotter, & Forgac, 2016).

On the other hand, cytohesin2 (CTH2) and Arf6 small GTPase have emerged as critical regulators of Slit2–Robo4–paxillin–GIT1 network which controls vascular stability and leak (Jones et al., 2009; London & Li, 2011). Inhibition of CTH2/Arf6 function can significantly reduce vascular permeability, which identified them as a new therapeutic target for ameliorating diseases involving the vascular system. In particular, cardiovascular hypertension and acute lung injury (ALI)/adult respiratory distress syndrome (ARDS), diseases in which inflammation, cytokine storm and vascular leak are present (Jones et al., 2009; London & Li, 2011), may be effectively treated by drugs targeting these proteins.

METHODS

Bioinformatics analysis and a2N$_{1-402}$ homology modeling.

The prediction of the secondary structure of a2N$_{1-402}$ was performed using PsiPred on-line server (http://bioinf.cs.ucl.ac.uk/PSIPRED/)(Buchan et al., 2010). The model of 1-352 part of a2 was built by homology modeling with MODELLER software (Eswar et al., 2006). The crystal structure of recently solved cytoplasmic N-terminal domain of

subunit I, homolog of subunit a, of V-ATPase (PDB ID 2RRK) was used as a template for homology modeling (Srinivasan et al., 2011). The alignment for homology modeling is based on profile-profile alignment of a2 homologs alignment and subunit I homologs alignment. Also subunit I secondary structure and a2 PSIPRED predicted secondary structure were taken into account (Buchan et al., 2010). For joining the homology $a2N_{1-352}$ model and the NMR peptide structures, Autodock Vina docking program was used to build several complexes of $a2N_{1-352}$ model and $a2N_{378-387}$ fragment from NMR structure of $a2N_{368-395}$ peptide (Trott & Olson, 2010).

Then the NMR structure of the entire $a2N_{368-395}$ peptide was fitted to these complexes and the complex was selected where the peptide did not intersect with the protein. The final model was builT by Modeller based on two templates: 1) the aforementioned complex of $a2N_{1-352}$ and $a2N_{368-395}$ peptide, and 2) NMR structure of $a2N_{386-402}$ peptide. Lastly, in the final model dihedral angles of Glu393 were manually changed for better fit into the cryo-EM map segment corresponding to I subunit of the *T. thermophilus* V-ATPase (Lau & Rubinstein, 2012) and fitting itself is described below separately.

Also I-TASSER software was used for building $a2N_{1-402}$ model (Roy, Kucukural, & Zhang, 2010). We made some modification to I-TASSER, which allow to set several preferred templates, and used 2RRK and NMR peptides structures as templates. $a2N_{1-352}$ parts of I-TASSER models were very similar to corresponded parts of the Modeller homology model. But all I-TASSER models were not in good agreement with Cryo-EM map and our knowledge about a2N structure, with 353-402 part interacting with BD domain (data not shown). The possible reason is following: I-TASSER is optimized for

globular proteins and can not take into account proximity of membrane which influences $a2N_{1\text{-}402}$ structure.

NMR data collection, processing and peptide structure determination.

NMR spectra of $a2N_{339\text{-}395}$ and $a2N_{386\text{-}402}$, respectively, were collected at temperature ranging from 288 K to 298 K on Avance cryo-probed 600 or 700 MHz instruments (Bruker, Billerica, MA). All NMR data were processed using Bruker Avance spectrometer in-built software Topspin program. To elucidate the structure of $a2N_{339\text{-}395}$ and $a2N_{386\text{-}402,}$ data were collected on Bruker Avance at 600 MHz or 700 MHz. Pulse calibrations and other parameter including temperature, buffer and peptide concentration were optimized before making final measurements. 10 % D_2O was used to lock the NMR signal in each experiment. The two dimensional (2D) ^1H NMR spectra including total correlation spectroscopy (TOCSY) and nuclear overhauser enhancement spectroscopy (NOESY) were obtained at the temperature of 298 K. 2D experiments with TOCSY were performed with mixing time of 60 to 80 ms, whereby NOESY used 200 to 300 ms mixing time. Baseline corrections were applied wherever necessary. The proton chemical shift was referenced to the methyl signal of DSS (2, 2-dimethyl-2-silapentane-5-sulphonate [Cambridge Isotope Laboratories]), an external standard reference to 0 ppm. Peak-assignment and data analysis of the Fourier transformed spectra were performed with SPARKY 3.1 program (Kneller & Goddard, 1997). Assignments were carried out according to the classical procedure including spin system identification and sequential assignment (Wüthrich, 1986). Inter proton distance were obtained from the NOESY spectra. NOESY peaks were categorized as strong, medium and weak based on the signal

intensity and were translated into distance constraints as 3.0 Å, 4.0 Å and 5.0 Å, respectively. Dihedral angle restraints were calculated from the spectra assigned chemical shift using torsion angle likelihood obtained from shift and sequence similarity (TALOS) (Cornilescu, Delaglio, & Bax, 1999). The secondary structures were predicted using the chemical shift index (CSI) and NOE pattern. Structure calculations were performed by using Cyana program which uses simulated annealing with molecular dynamics in torsion angle space (Güntert, Mumenthaler, & Wüthrich, 1997). Final Cyana ensemble structures were visualized by Pymol (DeLano, 2002).

Reagents, Peptides Synthesis, Labeling, and Purification

If not otherwise specified, all reagents were purchased from Sigma. Peptide synthesis resins and Fmoc-protected amino acids were purchased from EMD Chemicals. All peptides were synthesized, purified by HPLC, and analyzed by mass spectrometry as previously described (Hosokawa et al., 2013). The following a2N-derived peptides were synthesized:

$a2N_{339-395}$

(RRALEEGSRESGATIPSFMNIIPTKETPPTRIRTNKFTEGFQNIVDAYGVGSYREV)

$a2N_{386-402}$

(YGVGSYREVNPALFTII)

Bio- $a2N_{1-17}$ -TAT

(Bio-MGSLFRSESMCLAQLFL-YGRKKRRQRRR)

Bio-$a2N_{18-34}$-TAT

(Bio-QSGTAYECLSALGEKGLC-YGRKKRRQRRR)

Bio-a2N$_{35-49}$-TAT

(Bio-VQFRDLNQNVSSFQRC-YGRKKRRQRRR)

Bio-a2N$_{75-91}$-TAT

(Bio-ADIPLPEGEASPPAPPL-YGRKKRRQRRR)

Bio-a2N$_{313-331}$-TAT

(Bio-NMCSFDVTNKCLIAEVWCP--YGRKKRRQRRR)

Recombinant proteins and GEF Activity Assay

In these experiments, we used radiolabel-based assay to determine the GDP/GTP exchange activity of Sec7 domain and cytohesin-2 (CTH2). This assay allows the steady-state enzymatic GEF activity analysis of Sec7 domain and cytohesin-2 (CTH2) with delta17-Arf1 in the presence of PIP_2-containing liposomes. Preparation of recombinant proteins (CTH2, Sec7, delta17-Arf1) and experimental procedures of GEF activity assay were previously described (Hosokawa et al., 2013).

FIGURE LEGENDS

Figure 1. Amino acid sequence alignment of the cytosolic N-terminal regions from bacterial *M. ruber* subunit I, fungal *S. cerevisiae* Vph1p and mammalian *M. musculus* a2-subunit isoform. The alignment is performed as described in Methods. Identical residues are highlighted in red. The secondary structures of α-helices (α) and β-strands (β) on top in black are from bacterial I_{cyt} crystal structure. The secondary structures of mouse a2N$_{1-402}$, predicted by PsiPred, or assigned by PyMol for Modeller and I-Tasser models are on the bottom in blue. These secondary structures look similar to

each other and to the secondary structure of I_{cyt}, however while two long continuous α-helices are present in a2N model, the corresponding two regions in I_{cyt} each contains three α-helices (III, IV, V and VIII, IX, X) disrupted by kinks. The α-helices in the very bottom row are from NMR solution structures of peptides $a2N_{368-395}$ and $a2N_{386-402}$. Five residues of Vph1 including G_{181}, S_{266}, A_{347}, A_{351}, and Q_{369} (Qi & Forgac, 2008) and the corresponding mouse residues G_{170}, T_{255}, P_{336}, R_{340} and N_{358} are highlighted in green. Cytohesin-2 interacting peptides $a2N_{1-17}$, $a2N_{35-49}$, $a2N_{198-214}$, $a2N_{215-230}$, $a2N_{313-331}$ and $a2N_{386-402}$ are underlined on $a2N_{1-402}$ sequence, while discontinuous proximal I_{cyt_1-33} and $I_{cyt\_274-301}$ and continuous $I_{cyt\_135-196}$ distal lobe regions are underlined on I_{cyt} sequence. They also are shown as separate stack of amino acid sequences. Numbers below a2-subunit lines are from a2N amino acid sequence. The very bottom stack of amino acid sequences ($a2N_{353-402}$ in a2N) is an appendage region of these proteins with previously unknown structure.

Figure 2. Structural homology model of $a2N_{1-352}$. a) Schematic representation of structural domains of $a2N_{1-402}$. These structures are indicated as follows: i) proximal lobe (*PL*) formed by 1-42 aa and by 322-352 aa in red; ii) bar domain (*BD*) formed by 43-172 aa and 245-321 aa in green; iii) distal lobe (*DL*) formed by 173-244 aa in blue; and v) stem domain (*SD*) formed by 353-402 aa in white. **b,c)** Ribbon representations of a homology model of $a2N_{1-352}$. Structural domains are colored as above. It should be noted that stem domain is not present in this model. **b)** Bottom view from the membrane towards V_1-sector. **c)** Top view from the V_1-sector towards the membrane. This view is rotated by 180° vertically relative to the view in b.

Figure 3. Synthesis and NMR analysis of peptides overlapping and corresponding to stem domain of $a2N_{1-402}$. a) The C-terminal amino acid sequence of $a2N_{1-402}$. The end of proximal lobe (PL) domain (332-353 aa) is shown in peach and stem domain (SD) is in white. Overlapping synthetic peptides $a2N_{339-395}$ and $a2N_{386-402}$ are shown in cyan. The C-terminal part of $a2N_{339-395}$ peptide (called $a2N_{368-395}$), for which structure was solved by NMR spectroscopy is shown in grey. **b)** Table of structural statistics for the amino acids of $a2N_{368-395}$ and $a2N_{386-402}$ peptides determined by NMR analysis in solution.

Figure 4. NMR data and analysis of $a2N_{339-395}$ and $a2N_{386-402}$ peptides. Assignment of cross-peaks in the NOESY spectrum of **a)** $a2N_{339-395}$ and **b)** $a2N_{386-402}$ Peak picking was done in Sparky 3.1 software and peaks were identified based on TOCSY spectrum. Since the 2D NOESY assignment of $a2N_{339-395}$ reveals an unstructured N-terminal part, the calculation was performed only for the C-terminal region with the residues 368 to 395, called $a2N_{368-395}$. The amino acid sequence of **c)** $a2N_{368-395}$ and **d)** $a2N_{386-402}$ peptides and their secondary structure elements based on 1Hα chemical shifts with respect to the random coil values. NOESY connectivity plot of peptides **e)** $a2N_{368-395}$ and **f)** $a2N_{386-402}$ are indicative of the residues connected in space revealing the presence of a helical structure.

Figure 5. NMR structure of $a2N_{368-395}$ peptide. a) NMR structure of the peptide $a2N_{368-395}$ showing superimposition of 20 structures calculated by Cyana 3.1 package and revealing a partially α-helical formation. **b)** Cartoon representation of a single NMR structure of $a2N_{368-395}$. **c, d)** The molecular surface electrostatic potential of peptide **c)**

$a2N_{368-395}$ and **d)** its 180 view respectively generated by Pymol (DeLano, 2002). The positive potentials are drawn in blue, negative in red and hydrophobic in light grey.

Figure 6. NMR structure of $a2N_{386-402}$ peptide. a) NMR structure of the peptide $a2N_{386-402}$ showing superimposition of 20 structures calculated by Cyana 3.1 package and revealing a partially α-helical formation. **b)** Cartoon representation of a single NMR structure of $a2N_{386-402}$. **c, d)** The molecular surface electrostatic potential of peptide **c)** $a2N_{386-402}$ and **d)** its 180 view respectively generated by Pymol (DeLano, 2002). The positive potentials are drawn in blue, negative in red and hydrophobic in light grey.

Figure 7. Structural model of complete $a2N_{1-402}$ and identification of CTH2 and G/E-peripheral stalks binding sites. a-d) Structural molecular model of $a2N_{1-402}$ and identification of two distinct CTH2 binding sites. **a)** Ribbon representation of a top view of a molecular model of $a2N_{1-402}$ showing six cytohesin-2 binding peptides as follows: i) $a2N_{1-17}$ (in red); ii) $a2N_{35-49}$ (in purple); iii) $a2N_{198-214}$ (in blue); iv) $a2N_{215-230}$ (in cyan); v) $a2N_{313-331}$ (in yellow) and vi) $a2N_{386-402}$ (in orange). Note, that α-helix XIII of appendage region $a2N_{353-402}$ (lower left part with peptide $a2N_{386-402}$) lays close to α-helix I of core region $a2N_{1-352.}$ **b)** View of proximal lobe (*PL*) and details of the first cytohesin-2 binding site formed by $a2N_{1-17}$, $a2N_{35-49}$, $a2N_{313-331}$ and $a2N_{386-402}$ peptides. Note that three peptides $a2N_{1-17}$, $a2N_{35-49}$ and $a2N_{313-331}$ form a continuous surface while $a2N_{386-402}$ peptide is separate. This view is rotated by 90° vertically and then counter-clockwise by 90° horizontally relatively to the view in a. **c)** View of distal lobe (*DL*) and details of the second CTH2 binding site formed by $a2N_{198-214}$ and $a2N_{215-230}$ peptides. This view is rotated clockwise 90° horizontally relatively to the view in a. **d)** Ribbon

representation of a top view of a molecular model of $a2N_{1-402}$. A ll six CYH2 binding peptides are colored in red, while five amino acids G_{170}, T_{255}, P_{336}, R_{340} and N_{358} identified in G/E-stalks binding vicinity are colored in blue (Qi & Forgac, 2008). Amino acids P_{336}, R_{340} and N_{358} forming first G/E-stalk binding site in *PL* are on the left, while amino acids G_{170} and T_{255} forming second G/E-stalk binding site in *DL* are on the right.

Figure 8. Identification of V-ATPase interface involved in binding and signaling with CTH2. a) Two V-ATPase derived peptides $a2N_{1-17}$ and $a2N_{35-49}$ are forming a specific interaction interface on proximal lobe sub-domain of a2N (shown by gray rectangle) that is involved in interaction with Sec7 domain of CTH2. **b,c**) Experimental testing of our model demonstrated a potent inhibition of Sec7/Arf1-GEF activity by both **b**) Bio-$a2N_{1-17}$ -TAT (IC_{50} = 1.5 μM) and **c**) Bio-$a2N_{35-49}$ -TAT (IC_{50} = 0.9 μM) interface forming peptides. **d,e,f**) The V-ATPase-derived peptides: **d**) Bio-$a2N_{18-34}$ -TAT; **e**) Bio-$a2N_{75-91}$-TAT and **f**) Bio-$a2N_{313-331}$-TAT that are not involved in the formation of binding-competent interface, are unable inhibit enzymatic GEF activity of Sec7 domain. **g**) Sequences of V-ATPase-derived synthetic peptides tested in enzymatic GEF-activity assay of CTH2 Sec7 domain.

Figure 9. Model of interactions between CTH2 and V-ATPase: Implication for cross-talk and regulation of CTH2 and V-ATPase functions. a-b) The 3D reconstruction EM map of the related A-ATP synthase of the archaeon *P. furiosus* enzyme. The presented molecular model of $a2N_{1-402}$ (*yellow*), the *c*-ring (*wheat*) and G/E-peripheral stalks (G, *cyan* and E, *green*/red) were used for the fitting. Cytohesin-2

(CTH2) (*green*) has been accommodated based on the interactions described in the Figures 7a-c. **a)** This model proposes that binding of CTH2 to the V-ATPase may destabilize the binding of both G/E peripheral stalks to a-subunit, which could promote the disassembly of the V_1- and V_O-sectors and inactivation of H^+-pumping of the V-ATPase. **b)** This model proposes that ion-translocation in the interface of the *c*-ring and the C-terminal membrane-embedded domain of a-subunit (*yellow cylinder*), could be affected by the binding of the CTH2 to the proximal lobe of N-terminal domain of 2-subunit of V-ATPase.

KEYWORDS

V-ATPase a2-subunit isoform N-terminus; Arf GEF cytohesin-2; homology modeling; NMR analysis; Sec7 domain enzymatic GEF activity; cross-talk between V-ATPase and CTH2; rational drug design

ACKNOWLEDGMENTS

This study was supported by A*STAR BMRC (09/1/22/19/609) (to GG). PD is grateful to receive the Singapore International Graduate Award (SINGA).

CURRENT AFFILIATIONS

Dr. Phat Vinh Dip: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

Dr. Youg Raj Thaker: School of Biological Science, Protein Structure and Disease Mechanisms, University of Essex, CO4 3SQ, UK

COMPETING FINANCIAL INTERESTS

The authors declare that they have no competing financial interests.

SUPPLEMENTARY MATERIAL

Movie S1. Structural homology model of a2N$_{1-352}$.

Movie S2. Structural model of complete a2N$_{1-402}$.

REFERENCES

Bill, A., Schmitz, A., Konig, K., Heukamp, L. C., Hannam, J. S., & Famulok, M. (2012). Anti-proliferative effect of cytohesin inhibition in gefitinib-resistant lung cancer cells. *PLoS One, 7*(7), e41179. (doi:10.1371/journal.pone.0041179)

Buchan, D. W., Ward, S. M., Lobley, A. E., Nugent, T. C., Bryson, K., & Jones, D. T. (2010). Protein annotation and modelling servers at University College London. *Nucleic Acids Res, 38*(Web Server issue), W563-568. (doi:gkq427 [pii] 10.1093/nar/gkq427)

Casanova, J. E. (2007). Regulation of Arf activation: the Sec7 family of guanine nucleotide exchange factors. *Traffic, 8*(11),1476-1485. (doi:TRA634 [pii]10.1111/j.1600-0854.2007.00634.x)

Cornilescu, G., Delaglio, F., & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR, 13*(3), 289-302.

Danser, A. H. (2009). (Pro)renin receptor and vacuolar H+-ATPase. *Hypertension, 54*(2), 219-221. (doi:10.1161/HYPERTENSIONAHA.109.135236)

Danser, A. H., & Deinum, J. (2005). Renin, prorenin and the putative (pro)renin receptor. *Hypertension, 46*(5), 1069-1076. (doi:10.1161/01.HYP.0000186329.92187.2e)

DeLano, W. (2002). The PyMOLMolecular Graphics System. *DeLano Scientific.*

Dip, P. V., Saw, W. G., Roessle, M., Marshansky, V., & Grüber, G. (2012). Solution structure of subunit a, a (104-363), of the Saccharomyces cerevisiae V-ATPase and the importance of its C-terminus in structure formation. *J Bioenerg Biomembr, 44*(3), 341-350. (doi:10.1007/s10863-012-9442-3)

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper U. & Sali, A. (2006). Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics, Chapter 5*, Unit 5 6. (doi:10.1002/0471250953.bi0506s15)

Forgac, M. (2007). Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology. *Nat Rev Mol Cell Biol, 8*(11), 917-929. (doi:nrm2272 [pii]10.1038/nrm2272)

Guex, N., Peitsch, M. C., & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis, 30 Suppl 1*, S162-173. (doi:10.1002/elps.200900140)

Guntert, P., Mumenthaler, C., & Wuthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol, 273*(1), 283-298. (doi:S0022-2836(97)91284-5 [pii]10.1006/jmbi.1997.1284)

Herrmann, T., Güntert, P., & Wüthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol, 319*(1), 209-227. (doi:S0022283602002413)

Hinton, A., Bond, S., & Forgac, M. (2009). V-ATPase functions in normal and disease processes. *Pflugers Arch, 457*(3), 589-598. (doi:10.1007/s00424-007-0382-4)

Hosokawa, H., Dip, P. V., Merkulova, M., Bakulina, A., Zhuang, Z., Khatri, A., Jian X., Keating S.M., Bueler S.A., Rubinstein J.L., Randazzo P.A., Ausiello A.A. & Grüber G.,

& Marshansky, V. (2013). The N termini of a-subunit isoforms are involved in signaling between vacuolar H+-ATPase (V-ATPase) and cytohesin-2. *J Biol Chem, 288*(8), 5896-5913. (doi:10.1074/jbc.M112.409169)

Hurtado-Lorenzo, A., Skinner, M., El Annan, J., Futai, M., Sun-Wada, G. H., Bourgoin, S., Casanova J., Wildeman A., Bechoua S., Ausiello A.A., Brown D. & Marshansky, V. (2006). V-ATPase interacts with ARNO and Arf6 in early endosomes and regulates the protein degradative pathway. *Nat Cell Biol, 8*(2), 124-136. (doi:ncb1348 [pii] 10.1038/ncb1348)

Ichihara, A., Kaneshiro, Y., Takemitsu, T., Sakoda, M., & Itoh, H. (2007). The (pro)renin receptor and the kidney. *Semin Nephrol, 27*(5), 524-528. (doi:10.1016/j.semnephrol.2007.07.005)

Ichihara, A., & Kinouchi, K. (2011). Current knowledge of (pro)renin receptor as an accessory protein of vacuolar H+-ATPase. *J Renin Angiotensin Aldosterone Syst, 12*(4), 638-640. (doi:10.1177/1470320311429227)

Jones, C. A., Nishiya, N., London, N. R., Zhu, W., Sorensen, L. K., Chan, A. C., Lim C.J., Chen H., Zhang Q., Schultz P.G., Hayallah A.M., Thomas K.R., Famulok M., Zhang K., Ginsberg M.H. & Li, D. Y. (2009). Slit2-Robo4 signalling promotes vascular stability by blocking Arf6 activity. *Nat Cell Biol, 11*(11), 1325-1331. (doi:10.1038/ncb1976)

Kneller, D., & Goddard, T. (1997). SPARKY 3.105 *(Edit. San Francisco, CA, University of California)*.

Lau, W. C., & Rubinstein, J. L. (2010). Structure of intact Thermus thermophilus V-ATPase by cryo-EM reveals organization of the membrane-bound V(O) motor. *Proc Natl Acad Sci U S A, 107*(4), 1367-1372. (doi:0911085107 [pii] 10.1073/pnas.0911085107)

Lau, W. C., & Rubinstein, J. L. (2012). Subnanometre-resolution structure of the intact Thermus thermophilus H+-driven ATP synthase. *Nature, 481*(7380), 214-218. (doi:nature10699 [pii] 10.1038/nature10699)

Li, X. C., Zhang, J., & Zhuo, J. L. (2017). The vasoprotective axes of the renin-angiotensin system: Physiological relevance and therapeutic implications in cardiovascular, hypertensive and kidney diseases. *Pharmacol Res, 125*(Pt A), 21-38. (doi:10.1016/j.phrs.2017.06.005)

London, N. R., & Li, D. Y. (2011). Robo4-dependent Slit signaling stabilizes the vasculature during pathologic angiogenesis and cytokine storm. *Curr Opin Hematol, 18*(3), 186-190. (doi:10.1097/MOH.0b013e328345a4b9)

Lu, Q., Lu, S., Huang, L., Wang, T., Wan, Y., Zhou, C. X., Zhang C., Zhang Z. & Li, X. (2013). The expression of V-ATPase is associated with drug resistance and pathology of non-small-cell lung cancer. *Diagn Pathol, 8*, 145. (doi:10.1186/1746-1596-8-145)

Marshansky, V. (2007). The V-ATPase a2-subunit as a putative endosomal pH-sensor. *Biochem Soc Trans, 35*(Pt 5), 1092-1099. (doi:BST0351092 [pii] 10.1042/BST0351092)

Marshansky, V., & Futai, M. (2008). The V-type H+-ATPase in vesicular trafficking: targeting, regulation and function. *Curr Opin Cell Biol, 20*(4), 415-426. (doi:S0955-0674(08)00056-2 [pii] 10.1016/j.ceb.2008.03.015)

Marshansky, V., Rubinstein, J. L., & Grüber, G. (2014). Eukaryotic V-ATPase: novel structural findings and functional insights. *Biochim Biophys Acta, 1837*(6), 857-879. (doi:10.1016/j.bbabio.2014.01.018)

McGuire, C., Cotter, K., Stransky, L., & Forgac, M. (2016). Regulation of V-ATPase assembly and function of V-ATPases in tumor cell invasiveness. *Biochim Biophys Acta, 1857*(8), 1213-1218. (doi:10.1016/j.bbabio.2016.02.010)

Merkulova, M., Bakulina, A., Thaker, Y. R., Grüber, G., & Marshansky, V. (2010). Specific motifs of the V-ATPase a2-subunit isoform interact with catalytic and regulatory domains of ARNO. *Biochim Biophys Acta, 1797*(8), 1398-1409. (doi:S0005-2728(10)00055-1 [pii] 10.1016/j.bbabio.2010.02.009)

Merkulova, M., Hurtado-Lorenzo, A., Hosokawa, H., Zhuang, Z., Brown, D., Ausiello, D. A., & Marshansky, V. (2011). Aldolase directly interacts with ARNO and modulates cell morphology and acidic vesicle distribution. *Am J Physiol Cell Physiol, 300*(6), C1442-1455. (doi:ajpcell.00076.2010 [pii] 10.1152/ajpcell.00076.2010)

Murata, T., Yamato, I., Kakinuma, Y., Leslie, A. G., & Walker, J. E. (2005). Structure of the rotor of the V-Type Na+-ATPase from Enterococcus hirae. *Science, 308*(5722), 654-659. (doi:1110064 [pii] 10.1126/science.1110064)

Oot, R. A., & Wilkens, S. (2012). Subunit interactions at the V1-Vo interface in the yeast vacuolar ATPase. *J Biol Chem*. (doi:M112.343962 [pii] 10.1074/jbc.M112.343962)

Oot, R. A., Couoh-Cardel, S., Sharma, S., & Wilkens, S. (2017). Breaking up and making up: The secret life of the vacuolar H^+ -ATPase. *Protein Sci. 26,* 896-909. (doi: 10.1002/pro.3147)

Qi, J., & Forgac, M. (2008). Function and subunit interactions of the N-terminal domain of subunit a (Vph1p) of the yeast V-ATPase. *J Biol Chem, 283*(28), 19274-19282. (doi:M802442200 [pii] 10.1074/jbc.M802442200)

Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc, 5*(4), 725-738. (doi:nprot.2010.5 [pii] 10.1038/nprot.2010.5)

Srinivasan, S., Vyas, N. K., Baker, M. L., & Quiocho, F. A. (2011). Crystal structure of the cytoplasmic N-terminal domain of subunit I, a homolog of subunit a, of V-ATPase. *J Mol Biol, 412*(1), 14-21. (doi:S0022-2836(11)00768-6 [pii] 10.1016/j.jmb.2011.07.014)

Stransky, L., Cotter, K., & Forgac, M. (2016). The Function of V-ATPases in Cancer. *Physiol Rev, 96*(3), 1071-1091. (doi:10.1152/physrev.00035.2015)

Sun, Y., Danser, A. H. J., & Lu, X. (2017). (Pro)renin receptor as a therapeutic target for the treatment of cardiovascular diseases? *Pharmacol Res, 125*(Pt A), 48-56. (doi:10.1016/j.phrs.2017.05.016)

Toei, M., Toei, S., & Forgac, M. (2011). Definition of membrane topology and identification of residues important for transport in subunit a of the vacuolar ATPase. *J Biol Chem, 286*(40), 35176-35186. (doi:M111.273409 [pii] 10.1074/jbc.M111.273409)

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem, 31*(2), 455-461. (doi:10.1002/jcc.21334)

Vonck, J., Pisa, K. Y., Morgner, N., Brutschy, B., & Müller, V. (2009). Three-dimensional structure of A1A0 ATP synthase from the hyperthermophilic archaeon Pyrococcus furiosus by electron microscopy. *J Biol Chem, 284*(15), 10110-10119. (doi:M808498200 [pii] 10.1074/jbc.M808498200)

Wüthrich, K. (1986). NMR of proteins and nucleic acids. *(New York: Wiley), XV*, 292p.

Zhang, Z., Zheng, Y., Mazon, H., Milgrom, E., Kitagawa, N., Kish-Trier, E., Heck, A.L., Kane, P.M., & Wilkens, S. (2008). Structure of the yeast vacuolar ATPase. *J Biol Chem, 283*(51), 35983-35995. (doi:M805345200 [pii] 10.1074/jbc.M805345200)