bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

An Extended Coatomer Binding Motif in the SARS-CoV-2 Spike Protein

Debajit Dey[1,†], Suruchi Singh[1,†], Saif Khan[2,‡] , Matthew Martin[1,§], Nicholas J. Schnicker[2], Lokesh Gakhar[2,3,¶], Brian G. Pierce[4,5], and S. Saif Hasan[1,6,7,*]

[1]Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, 108 N. Greene Street, Baltimore, MD 21201

[2]Protein and Crystallography Facility, Carver College of Medicine, University of Iowa, 51 Newton Road, Iowa City, IA 5224

[3]Department of Biochemistry, Carver College of Medicine, University of Iowa, 51 Newton Road, Iowa City IA 52245

[4]W.M. Keck Laboratory for Structural Biology, University of Maryland Institute for Bioscience and Biotechnology Research, 9600 Gudelsky Drive, Rockville, MD 20850

[5]Department of Cell Biology and Molecular Genetics, University of Maryland, 4062 Campus Drive, College Park, MD 20742

[6]University of Maryland Marlene and Stewart Greenebaum Cancer Center, University of Maryland Medical Center, 22. S. Greene Street, Baltimore, MD 21201

[7]Center for Biomolecular Therapeutics, University of Maryland School of Medicine, 9600 Gudelsky Drive, Rockville, MD 20850

[†]Shared first authorship

[‡]Current address: Bridge Institute, USC Michelson Center for Convergent Bioscience, University of Southern California, 1002 Childs Way, Los Angeles, CA 90089

[§]Current address: University of Pittsburgh Kenneth P. Dietrich School of Arts and Sciences, 139 University Pl, Pittsburgh, PA 15260

[¶]Current address: PAQ Therapeutics, 625 Massachusetts Ave, Cambridge, MA 02139

*Correspondence: Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, 108 N. Greene Street, Baltimore, MD 21201; sshasan@som.umaryland.edu

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

β-Coronaviruses such as SARS-CoV-2 hijack coatomer protein-I (COPI) for spike protein retrograde trafficking to the progeny assembly site in endoplasmic reticulum-Golgi intermediate compartment (ERGIC). However, limited residue-level details are available into how the spike interacts with COPI. Here we identify a novel extended COPI binding motif in the spike that encompasses the canonical K-x-H dibasic sequence. This motif demonstrates selectivity for αCOPI subunit. Guided by an *in silico* analysis of dibasic motifs in the human proteome, we employ mutagenesis and binding assays to show that the spike motif terminal residues are critical modulators of complex dissociation, which is essential for spike release in ERGIC. αCOPI residues critical for spike motif binding are elucidated by mutagenesis and crystallography and found to be conserved in the zoonotic reservoirs, bats, pangolins, camels, and in humans. Collectively, our investigation on the spike motif identifies key COPI binding determinants with implications for retrograde trafficking.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

β-Coronaviruses have been responsible for major human respiratory diseases in the last two decades. In 2002, the severe acute respiratory syndrome coronavirus (SARS-CoV) was implicated in an epidemic first reported in China before spreading to 27 countries, which resulted in 774 deaths (Stadler et al., 2003). A decade later, Middle East respiratory syndrome (MERS) was reported in Saudi Arabia in 2012 with over 30% fatality in patients (Ahmed, 2017; Zaki et al., 2012). Most recently, the novel SARS-CoV-2 has been implicated in the COVID-19 global pandemic that has claimed over four million lives. Current efforts to contain the pandemic are focused primarily on vaccinations using the viral spike protein that is responsible for SARS-CoV-2 entry into host cells (Polack et al., 2020; USFDA, 2020). Fundamental insights into spike biogenesis will advance the understanding of how β-coronaviruses exploit host resources during viral infection and may potentially lead to the development of novel therapeutics.

The trimeric β-coronavirus spike is organized into an ectodomain, a transmembrane domain, and a cytosolic domain (Beniac et al., 2006; Song et al., 2004). In infected cells, the newly synthesized and post-translationally modified spike is transported from Golgi to the ERGIC compartment, which is the site of β-coronavirus progeny assembly (Klumperman et al., 1994; Lontok et al., 2004; McBride et al., 2007). This retrograde trafficking of the spike from Golgi to ERGIC involves a cytosolic dibasic motif, K-x-H-x-x (Lys-x-His, where x is any amino acid) (Lontok et al., 2004; McBride et al., 2007). Such C-terminal dibasic motifs and variants such as K-x-K-x-x and K-K-x-x are widely reported in the cytosolic tail of host membrane proteins that undergo retrograde trafficking (Cosson and Letourneur, 1994; Gaynor et al., 1994; Jackson et al., 1993; Townsley and Pelham, 1994). As such, the β-coronavirus spike demonstrates molecular mimicry of dibasic trafficking motifs (Lontok et al., 2004; McBride et al., 2007). This recycling of spike protein has been suggested to enhance interactions with the viral membrane (M) protein localized in ERGIC during progeny assembly and is crucial for spike maturation (Jennings et al., 2021; Lontok et al., 2004; McBride et al., 2007). These observations establish a key role of the dibasic motif in SARS-CoV and SARS-CoV-2 infection and propagation cycles. Interestingly, the spike dibasic motif and adjacent residues are completely conserved in sarbecoviruses, i.e., SARS-CoV and SARS-CoV-2, although sequence divergence in residues neighboring the dibasic motif is noted in MERS-CoV (Lontok et al., 2004; McBride et al., 2007).

On the host side, retrograde trafficking is mediated by the interactions of dibasic motifs with the coatomer protein-I (COPI) complex (Cosson and Letourneur, 1994; Letourneur et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1994). Seven subunits, namely, α, β, β', γ, δ, ε and ζ, assemble into a COPI complex on retrograde trafficking vesicles that carry cargo (Dodonova et al., 2015; Duden et al., 1991; Hara-Kuge et al., 1994; Harrison-Lavoie et al., 1993; Malhotra et al., 1989; Serafini et al., 1991; Stenbeck et al., 1993; Waters et al., 1991). Prior genetic, biochemical, biophysical, and structural investigations have shown that the binding site for host cargo dibasic motifs maps to the N-terminal β-propeller WD40 domains of α and β'COPI subunits, which are structural homologs (Cosson and Letourneur, 1994; Eugster et al., 2004; Fiedler et al., 1996; Jackson et al., 2012; Ma and Goldberg, 2013; Schroder-Kohne et al., 1998). Mutagenesis analyses of α and β' subunit N-terminal WD40 domains have identified residues critical for binding of host protein dibasic motifs (Jackson et al., 2012). This and another study (Ma and Goldberg, 2013) provided important structural details of how dibasic host and viral peptides bind to αCOPI-WD40 and β'COPI-WD40 domains.

Cellular and biochemical investigations in recent years and during the ongoing COVID-19 pandemic have suggested a role of COPI interactions in sarbecovirus spike trafficking, maturation, glycan processing, and syncytia formation during infection (Cattin-Ortolá et al., 2021; Jennings et al., 2021; McBride et al., 2007). These studies have established a platform to investigate the underlying chemistry of spike-COPI interaction. For instance, it is not known whether the K-x-H motif is sufficient to determine the strength of this interaction, whether adjacent residues in the spike cytosolic tail play a role in this binding, and what are the sequence determinants of spike-COPI disassembly crucial for spike release in ERGIC. On the host side, it is presently not known which COPI residues are critical for spike interactions. As such, key facets of this initial binding event in spike trafficking remain largely unknown for SARS-CoV-2 as well as SARS-CoV. In the present investigation, we address these questions using a combination of bio-layer interferometry (BLI), molecular modeling, mutagenesis, X-ray crystallography, and an *in silico* analysis of the human membrane proteome. Employing a sarbecovirus spike hepta-peptide corresponding to the K-x-H-x-x motif, we identify critical residues in αCOPI-WD40 for hepta-peptide binding and demonstrate structural alterations in an αCOPI-WD40 mutant. Amino acid propensity is described in human dibasic motifs and adjacent downstream residues, and mutagenesis experiments driven by this analysis provide insights into how sarbecovirus spike modulates strength of binding to COPI. Collectively, our study advances the structural and biophysical understanding of how the dibasic motif hijacks COPI for spike retention in endo-membranes and trafficking to the plasma membrane during sarbecovirus infections.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

pH Modulated direct binding of sarbecovirus spike hepta-peptide to αCOPI-WD40 domain

In this investigation, we heterologously expressed and purified the N-terminal WD-40 domain of αCOPI-WD40 (residues 1-327) from *Schizosaccharomyces pombe* (**SI Figure S1**). Although SARS-CoV-2 infects mammals, we chose this *S. pombe* construct for two reasons. First, the putative interaction interface for dibasic peptides is conserved between this construct and the human αCOPI-WD40 homologue, COPA (Jackson et al., 2012; Ma and Goldberg, 2013). Second, this construct has been previously crystallized and structurally characterized (Ma and Goldberg, 2013). This is consistent with our aim of understanding the structural basis of spike-COPI interactions. A crystal structure of the purified αCOPI-WD40 domain was determined to 1.75Å resolution (**Figure 1a, Table 1**). The αCOPI-WD40 domain is organized into a β-propeller and is consistent with previously described structures of αCOPI-WD40 (Cα root-mean-square-deviation is <0.5Å) (Ma and Goldberg, 2013). However, a peripheral loop and a short α-helix (Gly^{168}-Ala^{188}, shown with an arrow in **Figure 1a**) demonstrate substantial differences from previously described αCOPI-WD40 structures likely due to altered crystal packing. An N-terminal acetylation of the αCOPI-WD40 polypeptide was identified in the structure. Importantly, the αCOPI-WD40 domain interface for putative interactions with dibasic motifs is similar between the structure determined here and previously published structures (Ma and Goldberg, 2013).

Recent investigations of SARS-CoV-2 spike and previously of SARS-CoV spike with COPI have employed cellular lysates (Cattin-Ortolá et al., 2021; Jennings et al., 2021; McBride et al., 2007). Hence, we asked if there is direct interaction between the purified components. To address this question, we first established a BLI assay to test this interaction. A hepta-peptide of the sarbecovirus spike (^{1267}Gly-Val-Lys-Leu-His-Tyr-Thr^{1273}, **Figure 1b**) was synthesized with an N-terminal biotin tag attached via a linker. The hepta-peptide C-terminus has a free carboxylate to mimic the C-terminus of a polypeptide. This hepta-peptide (or its sequence variants) was immobilized on a streptavidin biosensor for BLI analysis. The purified αCOPI-WD40 domain was provided as the analyte in the BLI assay (**Figure 1c-i**). It was observed that the spike hepta-peptide binds directly to the purified αCOPI-WD40 domain with an equilibrium dissociation constant (K_D)=4.17±0.04 μM and a kinetic K_D=2.75±0.09 μM at pH 7.5 (**Figure 1c-e**). A scrambled sequence of this hepta-peptide showed no detectable interaction with αCOPI-WD40 (**Figure 1f**), suggesting that the binding of the wild-type hepta-peptide was sequence specific. This binding analysis demonstrates that the C-terminal peptide of the sarbecovirus spike contains sufficient sequence and structural information to interact directly with αCOPI-WD40. This is consistent with prior COPI binding analyses with peptides corresponding to host dibasic motifs (Jackson et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2012; Ma and Goldberg, 2013). In contrast, this sarbecovirus hepta-peptide demonstrates extremely weak binding to β'COPI-WD40 (**SI Figure S2a-d**). This selectivity for αCOPI-WD40 is consistent with that reported for a similar spike hepta-peptide (^{1377}Phe-Glu-Lys-Val-His-Val-Gln^{1383}) from porcine epidemic diarrhea virus (PEDV), an α-coronavirus (Ma and Goldberg, 2013).

One of the first analyses of COPI involvement in SARS-CoV spike trafficking employed cellular lysate pull-downs to show enhanced spike-COPI interactions under acidification to pH 6.5 (McBride et al., 2007). As the sarbecovirus hepta-peptide contains His^{1271} in the K-x-H motif, we asked if this acidification would affect hepta-peptide interaction with αCOPI-WD40. A nearly 3-fold enhancement in binding between the wild-type hepta-peptide and αCOPI-WD40 (equilibrium K_D=1.40±0.15 μM) was observed upon acidification to pH 6.5, likely due to partial protonation of the His residue in the hepta-peptide K-x-H motif (**Figure 1d, g**). Relative to pH 7.5, this lower pH accelerated the association rate of the hepta-peptide with αCOPI-WD40 by a factor of 1.7 while concomitantly suppressing complex dissociation by another 1.7-fold (**Figure 1h, i**). As such, acidification was inferred to be a key factor in stabilizing the hepta-peptide complex with αCOPI-WD40. All subsequent BLI assays were performed at pH 6.5 The biochemical basis of cytosolic acidification in sarbecovirus infections is discussed below.

The terminal residues in the spike hepta-peptide are key modulators of αCOPI-WD40 binding and dissociation

Co-crystallization was attempted to map the interaction interface between the spike hepta-peptide and the purified αCOPI-WD40 domain. Although crystals were obtained in several conditions, no density for the hepta-peptide was observed as the putative binding site on αCOPI-WD40 was blocked by symmetry related molecules in the crystal lattice (data not shown). This tendency of self-association between WD40 β-propeller domains in a crystal lattice has been previously described (Jackson et al., 2012; Ma and Goldberg, 2013). Hence, homology modeling was employed to analyze the structural basis of interaction between SARS-CoV-2 spike hepta-peptide and αCOPI-WD40 domain. This modeling was based on a prior co-crystal structure of αCOPI-WD40 domain with a dibasic peptide (Ma and Goldberg, 2013). Apart from the Lys^{1269} and His^{1271} residues in the K-x-H motif, the terminal Tyr^{1272}-Thr^{1273} residues in the hepta-peptide are within interaction distance of αCOPI-WD40 surface residues (**Figure 2a**). This is intriguingly suggestive of a role of these two spike residues in binding αCOPI. Interestingly, the two N-terminal residues in the hepta-peptide, i.e., Gly^{1267}-Val^{1268}, make no contact with the αCOPI-WD40 surface. To evaluate the role of the spike residues in binding αCOPI-WD40, *in silico* alanine scanning mutagenesis of the modeled spike hepta-peptide was performed (**Table 2**). The spike Lys^{1269} and

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

His^{1271} residues that constitute the K-x-H dibasic motif are predicted to be most crucial for binding αCOPI-WD40 domain. The *in silico* mutations of these residues to Ala yield highly unfavorable free energy changes suggestive of substantially weakened binding to αCOPI-WD40 (**Table 2**). The Ala mutation of Tyr^{1272} in the spike peptide implies a significant role of this residue in stabilization of the spike-αCOPI-WD40 complex. This is likely due to the side chain interaction between the oxygen atom in Tyr^{1272} side-chain hydroxyl group with the αCOPI-WD40 His^{31} side-chain NE2 atom, along with main chain interactions of Tyr^{1272}. The terminal residue in the spike, i.e., Thr^{1273}, is predicted to contribute modestly to the stabilization of the complex with αCOPI-WD40 (**Table 2**).

Next, we tested this *in silico* model of interactions between the spike hepta-peptide and αCOPI-WD40 using BLI assays (**Figure 2b-k**). The mutagenesis of Lys^{1269} or His^{1271} in the spike K-x-H motif to Ala residues abolished binding to αCOPI-WD40 (**Figure 2b, c**). As expected, the dual Ala mutation of the K-x-H motif does not demonstrate any substantial binding to αCOPI-WD40 (**Figure 2d**). As such, both basic residues in SARS-CoV-2 spike K-x-H motif are individually and concomitantly required for αCOPI-WD40 binding. Replacement of either residue is sufficient to disrupt αCOPI-WD40 binding to the spike hepta-peptide. These data are consistent with the *in silico* predictions described above as well as with cellular assays on SARS-CoV and SARS-CoV-2 spike trafficking (Cattin-Ortolá et al., 2021; Jennings et al., 2021; McBride et al., 2007). We next tested the contribution of spike hepta-peptide Tyr^{1272} residue to αCOPI-WD40 binding. A BLI assay of a mutant Tyr^{1272}→Ala spike hepta-peptide (^{1267}Gly-Val-Lys-Leu-His-<u>Ala</u>-Thr^{1273}) yielded an equilibrium K_D=3.77±0.34 μM, which is 2.7-fold weaker than the wild-type spike peptide (**Figure 2e, f)**. Although this mutation only reduced the rate of complex formation by 1.3-fold relative to the wild-type hepta-peptide (**Figure 2h**), it accelerated complex dissociation by 1.9-fold (**Figure 2i**). This suggests weakened interactions of the spike hepta-peptide with αCOPI-WD40 when the aromatic side chain interactions of Tyr^{1272} are abrogated. Collectively, this BLI analysis indicates that Tyr^{1272} is important for complex stability. These experimental results are consistent with the above described *in silico* model (**Table 2**). Next, we evaluated the C-terminal position Thr^{1273} in the spike. A BLI assay of a mutant Thr^{1273}→Ala hepta-peptide (^{1267}Gly-Val-Lys-Leu-His-Tyr-<u>Ala</u>^{1273}) yielded an equilibrium K_D=1.13±0.05 μM, which is similar to the wild type hepta-peptide (K_D=1.40±0.15 μM) (**Figure 2f, j**). This Thr^{1273}→Ala mutation caused a slowing down of αCOPI-WD40 association-dissociation kinetics by 2.3 and 1.9-fold respectively (**Figure 2h, i**). This suggested that a β-branched residue at the C-terminus may be an important determinant in complex formation kinetics. To probe further, we generated a Thr^{1273}→Val mutant hepta-peptide, which maintains a β-branched residue at the C-terminus. Val has methyl groups at the two side

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

chain γ positions, which replace a methyl and a hydroxyl group at equivalent γ positions in Thr. A BLI assay of this mutant hepta-peptide showed a 2.3-fold weakened interaction with αCOPI-WD40 relative to the wild type, with an equilibrium K_D=3.20±0.04 μM (**Figure 2f, k**). Interestingly, this mutant demonstrated 1.5-fold slower association kinetics and 1.5-fold more rapid dissociation than for the wild type hepta-peptide (**Figure 2h, i**). Compared to the Thr^{1273}→Ala hepta-peptide, this Thr^{1273}→Val mutant weakened binding by 2.8-fold while accelerating αCOPI-WD40 complex association and dissociation kinetics by 1.5 and 2.8-fold respectively (**Figure 2f, h, and i**). These data suggest that a β-branched residue at the spike C-terminus is involved in dissociation of the complex with αCOPI-WD40. Interestingly, a prior analysis implicated β-branched residues at the penultimate position in the PEDV spike sequence in modulating interactions with COPI-WD40 domains (Ma and Goldberg, 2013).

Electrostatics of spike hepta-peptide C-terminus drive dissociation from αCOPI-WD40

We performed an *in silico* analysis of the human proteome to gain insights into whether the spike extended dibasic motif demonstrates consistency with host dibasic motifs and their environment. We identified 119 sequences predicted to be membrane proteins that terminate with K-x-H-x-x and K-x-K-x-x dibasic motifs (**SI Table T1**). These sequences were aligned and analyzed for the frequency of 20 amino acids at each of the positions in the dibasic motif and the two terminal residues following the motif (**Figure 3a, SI Table T2**). This analysis revealed novel details about the dibasic motif. First, it was inferred that the predominant dibasic motif is K-x-K-x-x rather than K-x-H-x-x by nearly an order of magnitude. Second, only a low frequency (0.07) of the sequences have an aromatic residue at the penultimate position, which corresponds to Tyr^{1272} in the SARS-CoV-2 spike. β-Branched residues Leu, Ile, Val, Ser, and Thr are found at a high frequency of 0.38 at this penultimate position. Third, acidic residues at the C-terminus are observed in nearly a quarter (frequency=0.24) of the sequences. Overall, with a frequency of 0.42, the C-terminal position has a strong tendency to be occupied by charged residues such as Arg, Asp, Glu, His, and Lys. Hydroxyl side chain containing Thr, which corresponds to Thr^{1273} in SARS-CoV-2 spike, is a low frequency residue (0.05). In our *in silico* model of the hepta-peptide complexed with αCOPI-WD40, the side chain of this Thr^{1273} residue is within interaction distance of a cluster of basic residues in αCOPI-WD40 (Arg^{13}, Lys^{15}, and Arg^{300}, **Figure 3b**). Hence, we hypothesized that the presence of a charged residue at this spike position would modulate interactions with αCOPI-WD40. This was supported by our *in silico* analysis, which predicted stabilization of the complex when an acidic Glu residue replaced Thr^{1273} in the spike hepta-peptide (**Table 2**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We next tested the role of this spike C-terminal residue in modulating αCOPI-WD40 binding using a BLI assay (**Figure 3c-i**). We employed three distinct mutations of the spike hepta-peptide at this position, i.e., acidic (Glu), basic (Arg), and neutral (Gln) (**Figure 3c-e**). The presence of an acidic Glu residue at the C-terminus was found to substantially strengthen binding of the hepta-peptide to αCOPI-WD40 with an equilibrium K_D=0.31±0.01 μM, which is 4.5-fold tighter than the binding of the wild type hepta-peptide sequence (**Figure 3f**). This is consistent with our *in silico* model and is strongly suggestive of an electrostatic interaction between the Glu^{1273} side chain and αCOPI Arg^{13}, Lys^{15}, and Arg^{299} side chains to stabilize the complex. Furthermore, the rate of dissociation of αCOPI-WD40 domain from the Glu^{1273} hepta-peptide is 6-fold slower than that of the wild-type spike hepta-peptide (**Figure 3h**). In fact, during the time course of our experiment, we did not observe complete dissociation of this complex with the Glu^{1273} containing hepta-peptide. To eliminate the possibility of non-specific interactions, we employed a hepta-peptide with a scrambled sequence (**SI Figure S3**). Next, we tested whether modifying side-chain charge at the C-terminal position of the spike hepta-peptide affects complex formation with αCOPI-WD40. Relative to Glu^{1273}, the binding between αCOPI-WD40 domain was weakened when neutral Gln^{1273} was substituted into the hepta-peptide (equilibrium K_D=0.78±0.04 μM, **Figure 3d, f**). However, the binding of the hepta-peptide with Gln^{1273} was still 1.8-fold tighter than that of the wild-type hepta-peptide (**Figure 3f**). In contrast, basic Arg^{1273} in the spike hepta-peptide (equilibrium K_D=1.73±0.24 μM) yielded an interaction strength similar to the wild type sequence (equilibrium K_D=1.40±0.15 μM, **Figure 3e, f**). The amide carbonyl group in the Gln^{1273} side-chain likely interacts with the basic residue cluster on αCOPI-WD40 through hydrogen bonding. This stabilizing interaction is disrupted when Gln is replaced by Arg^{1273} in the spike hepta-peptide. Intriguingly, the rate of association with αCOPI-WD40 was slowed down relative to the wild type hepta-peptide by a factor of 2.6 for Glu^{1273} and Gln^{1273}, whereas it was similar to that of Arg^{1273} containing hepta-peptide (**Figure 3i**). Overall, these data establish a critical role of the C-terminal position in the SARS-CoV-2 spike in modulating binding to αCOPI-WD40.

A polar αCOPI-WD40 interface for spike hepta-peptide binding

We subsequently focused our attention on the spike binding residues in αCOPI-WD40. The *in silico* modeling of SARS-CoV-2 spike hepta-peptide shows that interaction with αCOPI-WD40 domain involves predominantly polar residues (**Table 3**). Amongst these residues, Arg^{57}, Asp^{115}, and Tyr^{139} provide the highest level of stabilization to spike hepta-peptide binding. The Arg^{57} side-chain interacts with the main chain carbonyl of spike His^{1271}, Tyr^{1272}, and Thr^{1273} (**Figure 4a**). The Asp^{115} side-chain forms a bond with the terminal NZ atom in the spike Lys^{1269} side-chain

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(**Figure 4b**). This side-chain of spike Lys^{1269} is further stabilized by an interaction with the hydroxyl oxygen in Tyr^{139} side-chain (**Figure 4c**). Hence, the side chains of αCOPI-WD40 Arg^{57}, Asp^{115}, and Tyr^{139} residues provide an extensive and polar interaction network for binding of the spike hepta-peptide. Therefore, mutagenesis of these three αCOPI-WD40 residues to Ala is predicted to disrupt interactions with the spike hepta-peptide as suggested by our *in silico* analysis (**Table 4**).

Next, the role of αCOPI-WD40 Arg^{57}, Asp^{115}, and Tyr^{139} residues in binding the spike hepta-peptide was tested. We generated three single-site mutants of αCOPI-WD40 wherein Arg^{57}, Asp^{115}, and Tyr^{139} residues were individually mutated to Ala. These mutants were expressed heterologously and purified from Expi293 cells. Analysis by SEC suggested an overall similarity in hydrodynamic radius with the wild-type αCOPI-WD40 domain (**SI Figure S1**). These three mutants were analyzed for binding to the wild-type spike hepta-peptide by BLI assays. All three mutants demonstrated extremely weak binding to the wild type sequence of the spike hepta-peptide (**Figure 4d-f**). This demonstrated that αCOPI-WD40 residues Arg^{57}, Asp^{115}, and Tyr^{139} are individually critical for binding the spike hepta-peptide. Disruption of even one of these interactions is likely sufficient to destabilize the spike-COPI complex.

Structural basis of conformational changes in an αCOPI-WD40 mutant

The results of αCOPI-WD40 Arg^{57}, Asp^{115}, and Tyr^{139} mutagenesis led us to ask if the loss of binding to the spike hepta-peptide was due to disruption of a single critical interaction or due to larger alterations in the protein structure. To address this question, we crystallized αCOPI-WD40 Arg^{57}→Ala and Tyr^{139}→Ala mutants. The αCOPI-WD40 Asp^{115}→Ala mutant did not yield crystals in the conditions we tested. The crystal structures of αCOPI-WD40 Arg^{57}→Ala and Tyr^{139}→Ala mutants were determined by X-ray diffraction to a resolution of 1.24Å and 1.49Å, respectively (**Table 1**).

The αCOPI-WD40 Arg^{57}→Ala mutant structure demonstrated novel structural alterations that had previously not been reported in the crystal structures of wild-type αCOPI-WD40 or the related β'COPI-WD40 (**Figure 5a**). The mutation of Arg^{57} to an Ala residue generated a cavity in the spike hepta-peptide binding site. This change led to a 62° rotation of a nearby Tyr^{97} residue side chain into the newly generated cavity in αCOPI-WD40 (**Figure 5b**). In parallel, the residue Asp^{73} underwent a substantial conformational change. This residue interacts with the side chain of Arg^{57} in the wild-type αCOPI-WD40 structure. However, the loss of stabilizing interactions from the Arg^{57} side chain and the reorientation of Tyr^{97} caused the Asp^{73} side chain to rotate away by 73° from its initial position (**Figure 5b**). These conformational changes are accompanied by a

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1.1Å and 0.8Å movement of Tyr^{139} and His^{31} side chains respectively closer towards the spike hepta-peptide as inferred from our *in silico* model. In contrast, the side chain of Lys^{15} moves 1.7Å away from the inferred hepta-peptide position. As such, the binding site and its vicinity demonstrate a substantially modified interaction network in Arg^{57}→Ala mutant.

Next, we asked if Arg^{57}→Ala mutation and the associated rotameric changes caused any main chain reorganization in αCOPI-WD40. To obtain a global overview of changes in the main chain geometry, the differences in Ramachandran angles were calculated between corresponding residues in the wild type and Arg^{57}→Ala crystal structures (**Figure 5c**). The top peak in this difference Ramachandran plot, i.e., peak 1, in this analysis corresponds to a substantial main chain twist at Gly^{72}, which is in the neighborhood of the Arg^{57}→Ala mutation. This conformational change is associated with the Asp^{73} side chain rotation a repositioning of the main chain atoms from Gly^{72} to Val^{77}, which are pushed away from the domain core consistent with the reorientation of the Asp^{73} side chain (**Figure 5d**). Peaks 2 and 4 in this difference Ramachandran plot correspond to changes in surface loops that are 19Å and 25Å from the mutation site and are likely due to crystal contacts. Peaks 3 and 6 correspond to main chain rearrangement in the mutation site, i.e., Arg^{57}→Ala. This is likely a combination of the mutation and modifications to side chain rearrangements in the neighborhood of Ala^{57} described above. Peak 5 is associated with a surface loop 31Å from the mutation site. This loop demonstrates weak electron density and is only partly ordered. Hence, three of the top six peaks, i.e., 1, 3, and 6, in this analysis are associated with considerable rearrangement of the αCOPI-WD40 surface upon the mutation of basic Arg^{57} to neutral Ala including the site for spike hepta-peptide binding.

In contrast to the Arg^{57}→Ala substitution, the crystal structure of αCOPI-WD40 Tyr^{139}→Ala mutant demonstrated no significant changes as compared to the wild-type structure (**Figure 5c, 5e**). No major rearrangements of side chains or main chains were observed. This lack of conformational rearrangement contrasts with the structural changes in the αCOPI-WD40 Arg^{57} →Ala mutant structure. It is likely that the electroneutral change from Tyr^{139} to Ala does not perturb the local electrostatic surface sufficiently to alter protein conformation. Hence, disruption of spike hepta-peptide binding in this Tyr^{139}→Ala mutant is due to the loss of a single side chain hydroxyl group. Collectively, the crystal structures of αCOPI-WD40 Arg^{57}→Ala and Tyr^{139}→Ala mutants reveal distinct and contrasting structural principles by which spike hepta-peptide binding is disrupted.

An analysis of crystal packing in the αCOPI-WD40 structures reported here showed that the peptide binding site residues are in contact with symmetry related chains. We subsequently asked if distinct crystal packing may have contributed to these different structural consequences

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of Arg^{57}→Ala and Tyr^{139}→Ala mutations. However, similar crystal packing interactions are provided by residues His^{267} and Lys^{309} from a symmetry-related αCOPI-WD40 chain to the spike hepta-peptide binding site. As such, conformational differences in these two αCOPI-WD40 mutants are due to altered interaction chemistry in the hepta-peptide binding site.

Conservation of αCOPI-WD40 residues critical for spike hepta-peptide binding

The mutagenesis, BLI, and crystallographic analyses described here are focused on αCOPI. Hence, we asked if αCOPI Arg^{57}, Asp^{115}, and Tyr^{139} residues are conserved in bats, pangolins, camels, and humans, which have been implicated as zoonotic reservoirs and hosts for β-coronaviruses (Bolles et al., 2011; Han et al., 2016; Latinne et al., 2020; Roess et al., 2016; Zhang et al., 2020). Overall, αCOPI is 98.5-99.1% identical in these multicellular higher organisms (**SI Table T3**). In contrast, αCOPI conservation is relatively moderate between these organisms and yeast at 46.8-47.1% sequence identity. However, all three αCOPI residues critical for spike hepta-peptide binding, i.e., Arg^{57}, Asp^{115}, and Tyr^{139}, are found to be 100% identical in yeast, bat, pangolin, camel, and human αCOPI. This suggests a conserved role of these three residues in binding dibasic motifs in host proteins, which is exploited by the sarbecovirus spike to hijack the host COPI machinery. αCOPI residues such as Lys^{15} and His^{31} that are suggested to be involved in spike hepta-peptide binding by our *in silico* analysis demonstrate complete conservation. Interestingly, this conservation extends to *S. cerevisiae* β'COPI wherein αCOPI residues Arg^{57} and Asp^{115} are replaced by Arg^{59} and Asp^{117} in β'COPI, respectively. However, αCOPI Tyr^{139} is semi-conserved and is replaced by Phe^{142} in β'COPI.

Discussion

The trafficking of sarbecovirus spike protein during progeny virus assembly consists of three distinct steps, i.e., spike-COPI binding in donor membranes such as cis-Golgi, inter-organelle trafficking, and dissociation of spike-COPI at the destination, which is ERGIC (Klumperman et al., 1994). This trafficking pathway can be disrupted by either weakening of spike-COPI binding leading to premature complex dissociation or enhanced stability of this complex, which interferes with spike release. This is supported by recent cellular imaging and biochemical analysis of the SARS-CoV-2 spike protein (Jennings et al., 2021). Hence, elucidating the determinants of spike-COPI interactions is fundamental to understanding sarbecovirus assembly.

Employing a spike hepta-peptide and a purified aCOPI-WD40 domain, the present investigation expounds on the biophysical and structural bases of spike-COPI interactions. We demonstrate that direct binding of purified αCOPI-WD40 domain to the SARS-CoV-2 spike hepta-

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

peptide is modulated by an extended coatomer binding motif that stretches beyond the spike K-x-H residues. Our data show that residues such as acidic Glu in the C-terminal position in the spike likely interact with complementary charged basic residues in αCOPI-WD40. This interaction strengthens spike binding to the host αCOPI. This analysis is consistent with a recent preprint that shows a key role of this SARS-CoV-2 spike C-terminal position in pull-down assays of the spike cytosolic domain with COPI subunits (Cattin-Ortolá et al., 2021). A second cellular investigation has recently shown that the inferred stabilization of the spike-COPI complex by a Lys^{1269}-x-His^{1271}→Lys^{1269}-x-Lys^{1271} spike mutation has dramatic effects on SARS-CoV-2 spike processing and trafficking (Jennings et al., 2021). This functional analysis suggests a key role of spike-COPI complex dissociation in modulating spike trafficking and function. Hence, it is likely that residues that strengthen spike-COPI complex stability beyond that from wild type interactions are avoided in the spike C-terminus. This includes acidic Glu and unbranched Ala residue that stabilize the αCOPI-WD40 domain as demonstrated in the present investigation. Interestingly, our analysis of the human membrane proteome suggests that the occurrence of a charged residue such as Glu and β-branched residues is a high probability event at the C-terminus of dibasic motifs. This raises an intriguing question of whether such charged residues present a structural and biophysical disadvantage to spike-COPI interactions, and hence, are selected against in sarbecoviruses. Such a highly stabilized complex may not undergo dissociation in ERGIC to release the spike for processing and downstream virion assembly. Although our investigation does not probe this interesting question in sarbecovirus evolution, we note that complex formation of Glu^{1273} or Gln^{1283} containing spike hepta-peptide is substantially slower than in the wild type. Even Arg^{1273}, which lacks charge complementarity with αCOPI-WD40 residues shows slower association kinetics. Glu, Gln, and Arg have long side chains unlike Thr^{1273}, which is suggestive of a role of side chain size in modulating interactions with COPI. As suggested previously, β-branching presents steric restrictions on protein-protein binding (Ma and Goldberg, 2013). Hence, Thr^{1273} likely presents a means to enhance spike-COPI dissociation. Here, it is significant to note that although αCOPI-WD40 and β'COPI-WD40 domains have high structural similarity (Ma and Goldberg, 2013), our data demonstrate selectivity of the sarbecovirus hepta-peptide in favor of αCOPI-WD40. This is consistent with the binding selectivity of the PEDV hepta-peptide for αCOPI-WD40 (Ma and Goldberg, 2013) and likely points to a functional role of αCOPI in coronavirus biogenesis.

In the context of electrostatics of spike-COPI interactions, acidification of the cytosol is reported preceding apoptosis, which is associated with β-coronavirus infections (Gottlieb et al., 1996; Li et al., 2020a; Yan et al., 2004). More recently, investigations have suggested the

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

involvement of elevated lactate levels in COVID-19, which are associated with lowered pH (Li et al., 2020b). Our investigation demonstrates a role of acidic pH in strengthening spike hepta-peptide binding to αCOPI-WD40, which is consistent with prior cellular data (McBride et al., 2007). The protonation of His^{1271} residue in the spike dibasic motif upon acidification likely contributes to this strengthened binding. It is intriguing that the spike cytosolic domain is highly enriched in charged residues that would respond to pH changes. Concurrently, the dibasic motif binding surface of αCOPI and β'COPI demonstrates clustering of charged residues that would be responsive to pH modifications.

Our investigation identifies three αCOPI residues Arg^{57}, Asp^{115}, and Tyr^{139}, as essential for spike hepta-peptide binding. These αCOPI residues are completely conserved across organisms associated with β-coronavirus infections such as bats, pangolins, camels, and humans. Importantly, the spike dibasic motif and adjacent residues demonstrate complete conservation in β-coronavirus isolates from bats, pangolins, civets, camels, and humans (Ge et al., 2013; Guan et al., 2003; Lau et al., 2005; Li et al., 2005; Raj et al., 2014; Tang et al., 2006; Xiao et al., 2020). This is suggestive of a likely conserved COPI interaction mechanism for spike binding, and more broadly of dibasic motifs, in these organisms. Interestingly, the critical αCOPI residues identified in the present analysis are broadly consistent with a prior genetic and biophysical study that implicated αCOPI Arg^{57} and Lys^{15}, and β'COPI Arg^{59} and Asp^{117} (equivalent to αCOPI Asp^{115}) as critical for dibasic motif binding, retrograde trafficking, and growth of yeast cells (Jackson et al., 2012).

Building on this prior investigation, our crystallographic analysis of αCOPI Arg^{57}→Ala and Tyr ^{139}→Ala mutants presents two complementary structural results to substantially advance the understanding of how these residues are critical for COPI architecture. The αCOPI Arg^{57}→Ala mutant demonstrates a rearrangement of the spike hepta-peptide binding site and of neighboring residues whereas the Tyr^{139}→Ala mutant structure is largely similar to the wild type αCOPI structure. Yet, both mutants demonstrate the same functional outcome, i.e., loss of spike hepta-peptide binding. Given this structural sensitivity of αCOPI, and presumably β'COPI, to changes in electrostatics, this raises an interesting question about the structural basis of how mutations in these subunits alter normal retrograde trafficking. It is relevant to note that the *COPA* gene, which encodes the human homolog of αCOPI, has been implicated in a set of clinical disorders collectively known as the COPA syndrome (Vece et al., 2016; Watkin et al., 2015). Here, mis-sense mutations including ones that modify side chain charge in the WD40 domains compromise COPA protein function in retrograde trafficking (Watkin et al., 2015). Based on data presented

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

here, it would be of interest to investigate the structural basis of this dysfunction to gain deeper insights into COPI biology.

In conclusion, our present analysis and supporting prior investigations demonstrate that the extended dibasic motif in the sarbecovirus spike functions as an effective tool to hijack the COPI complex involved in retrograde trafficking. In broader terms, our structural analysis provides a basis to further investigate the structural and functional consequences of αCOPI and β'COPI mutations in disrupting retrograde trafficking.

Methods

Protein and peptide production

The *S. pombe* αCOPI-WD40 domain was synthesized by TOPGENE and cloned in pcDNA3.1(+) with a C-terminal strep-tag for affinity purification. Five mutations (Leu^{181}→Lys, Leu^{185}→Lys, Ile^{192}→Lys, Leu^{196}→Lys and Phe^{197}→Lys) were incorporated in the gene to improve solubility as suggested previously (Ma and Goldberg, 2013). Expression was performed in Expi293 mammalian cells using the Thermo Fisher ExpiFectamine expression kit. Protein purification was performed by affinity chromatography of the clarified cellular lysate followed by SEC in a Superdex 75 chromatography column. Arg^{57}→Ala, Asp^{115}→Ala, and Tyr^{139}→Ala mutants of αCOPI-WD40 domain were expressed and purified as described for the wild type protein. The purified αCOPI-WD40 domain in 150 mM NaCl, 5 mM dithiotreitol (DTT), 10% glycerol, and either 20 mM Tris-HCl (pH 7.5) or 50 mM MES-NaOH (pH 6.5) was flash-frozen in liquid nitrogen until further experimentation.

β'COPI-WD40 (residues 1-304) from *Saccharomyces cerevisiae* was cloned in pGEX-6P-1 vector and expressed overnight in *E. coli* pLysS cells at 18°C. This GST fusion protein was purified by affinity chromatography and SEC in 150 mM NaCl, 5 mM dithiotreitol (DTT), 10% glycerol, and 50 mM MES-NaOH (pH 6.5) and was flash-frozen in liquid nitrogen until further experimentation.

Peptide synthesis was performed by Biomatik (USA) with an N-terminal biotin tag and a $(PEG)_4$ linker between the tag and the peptide. No modification was performed at the C-terminus of the peptides thereby leaving a free terminal carboxylate group.

BLI assay

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Biotinylated spike hepta-peptides were tethered to streptavidin (SA) biosensors (FortéBio) in a 96-well plate format. Purified αCOPI-WD40 domain was provided as the analyte. Kinetics measurements for determination of binding affinity were performed on an Octet RED96 system (FortéBio). Data acquisition was carried out using the Data Acquisition 11.1 suite. Briefly, SA biosensors were hydrated in 200 µL of kinetics buffer (20mM Tris-HCl (pH 7.5) or 50 mM MES-NaOH (pH 6.5), 150 mM NaCl, 5 mM DTT, 10% glycerol, 0.2 mg/ml bovine serum albumin (BSA), and 0.002% Tween 20) for 10 minutes prior to binding. The spike hepta-peptide (5 µg/ml) was loaded on the biosensors for 15 seconds. A baseline was established by rinsing the biosensor tips in the kinetic buffer for 30 seconds. This was followed by association with αCOP in varying concentrations over 60 seconds and dissociation in the baseline well for 90 seconds. A temperature of 25°C and a shake speed of 1000 rpm was maintained during acquisition. All experiments were carried out in triplicates. A new sensor was used for each replicate. Data processing and analysis were performed in the FortéBio Data Analysis 11.1 software suite. Raw data was subtracted from the 0 µM αCOP signal as a reference. The baseline step immediately before the association step was used for the alignment of the y-axis. An inter-step correction between the association and dissociation steps was performed. Reference subtracted curves were processed with the Savitzky-Golay filtering method and subjected to global fitting using a 1:1 binding model. All fits to BLI data had R^2 value (goodness of fit) >0.9.

Crystallization and structure determination

Purified αCOPI-WD40 domain was concentrated to 2 mg/ml in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM DTT and 10% glycerol buffer. Crystal trays were set up with the hanging drop vapor diffusion method with 0.5 µL of αCOPI-WD40 mixed with an equal volume of reservoir buffer. Crystals were within 48 hours at 22°C in 20% PEG3350 and 0.25 M sodium citrate tribasic dihydrate. Crystals were cryo-protected in mother liquor supplemented with 20% ethylene glycol and flash-frozen in liquid nitrogen. Purified Arg^{57}→Ala and Tyr^{139}→Ala mutants of αCOP were concentrated to ~2.2 mg/ml and crystallized as described for the wild type protein. Crystals for Arg^{57}→Ala were obtained in 22% PEG3350 and 0.2 M trisodium citrate and for Tyr^{139}→Ala in 18% PEG3350 and 0.2M potassium-sodium tartrate. The crystals for Arg^{57}→Ala and Tyr^{139}→Ala αCOPI-WD40 mutants were cryoprotected in 20% glycerol and 20% ethylene glycol, respectively. X-ray diffraction data for wild-type αCOPI-WD40 was collected at the beamline GM/CA 23-ID-D of the Advanced Photon Source at the Argonne National Laboratory and at the National Synchrotron Light Source II (NSLS II) beamline 17-ID-1 AMX at the Brookhaven National Laboratory for the mutants. The X-ray diffraction data for the wild type protein crystals was

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

indexed, integrated and scaled using HKL3000 (Minor et al., 2006) whereas those for the mutants were processed using XDS (Kabsch, 2010) as part of the data acquisition and processing pipeline at the beamline. The data processing statistics are given in **Table 1**. The scaled data were merged using AIMLESS in CCP4 suite (Evans and Murshudov, 2013). Molecular replacement was performed in Phenix using a previously determined αCOPI-WD40 domain structure (PDB ID 4J87) as the search model (McCoy et al., 2007; Rossmann and Blow, 1962). Iterative model building and refinement were performed in Phenix.refine (Afonine et al., 2012) and Coot (Emsley and Cowtan, 2004). Figures were generated in PyMol. Part of the software used here was curated by SBGrid (Morin et al., 2013).

Analysis of Ramachandran angles

The crystal structures of wild type, Arg^{57}→Ala, and Tyr^{139}→Ala αCOPI-WD40 were analyzed in Molprobity (Williams et al., 2018). For each structure pair, i.e., wild type with Arg^{57}→Ala or wild type with Tyr^{139}→Ala, per residue difference in Ramachandran angles was determined using equation [1]-

$$(\delta=\sqrt{((\psi_{WT}-\psi_m)^2+(\phi_{WT}-\phi_m)^2)}) \quad [1]$$

Here, Ramachandran angles for wild type and mutant structures are represented as (ψ_{WT}, ϕ_{WT}) and (ψ_m, ϕ_m) respectively. Each structure-pair was superimposed in PyMol and inspected to ensure consistency with the results of the Ramachandran angle analysis. We identified three instances of surface exposed residues (Asp^{96}, Asn^{257} in the wild type coordinates, and Ser^{11} in Arg^{57}→Ala coordinates) where the structures were highly similar between corresponding main chain atoms in the wild type and mutant but the sign of a dihedral angle close to 180° had been flipped. The Cα rmsd of short penta-residue stretches of the polypeptide chain centered at each of these residues was 0.14Å, 0.06Å, and 0.18Å, respectively. The signs of the Ramachandran angles for these residues were corrected manually.

***In silico* analysis of sarbecovirus spike hepta-peptide with αCOPI-WD40**

Structural modeling of the SARS-CoV-2 spike C-terminus peptide (sequence: GVKLHYT) in complex with the αCOPI-WD40 domain was performed using homology modeling in Modeller (Webb and Sali, 2014) and the structure of the αCOP-WD40 complexed with Emp47p peptide (PDB ID 4J8B) as a template. Prior to computational mutagenesis, models were processed with FastRelax (Conway et al., 2014) in Rosetta (v. 3.5), with backbone and side chain atoms

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

constrained to the input coordinates. The command line parameter settings for FastRelax execution ("relax" executable) used were:

-relax:constrain_relax_to_start_coords

-relax:coord_constrain_sidechains

-relax:ramp_constraints false

-ex1

-ex2

-use_input_sc

-correct

-no_his_his_pairE

-no_optH false

-flip_HNQ

-nstruct 1

Computational mutagenesis simulations to predict effects on binding affinities (ΔΔGs) for point substitutions were performed using a previously described protocol implemented in Rosetta (v. 2.3) (Kortemme and Baker, 2002). Default parameters were used, with the exception of extra rotamers allowed during packing of modeled side chains, specified by command line parameters:

-extrachi_cutoff 1 -ex1 -ex2 -ex3

Sequence analysis of dibasic motifs in the human membrane proteome

UNIPROT identifiers of secreted and membrane-bound human proteins, as well as secreted/membrane-bound protein isoforms, were downloaded from the Human Protein Atlas (http://www.proteinatlas.org) (Thul et al., 2017). The corresponding protein sequences were obtained from UNIPROT, leading to approximately 6800 sequences, which were parsed using an in-house Perl script to identify C-terminal motif residues. Of these, 119 sequences that

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

demonstrated a C-terminal dibasic motif were analyzed further for amino acid propensities in the dibasic motif and neighboring residues.

Accession numbers of protein sequences used in this investigation

αCOPI (yeast, NP_595279.1; bat, XP_031291824.1; pangolin, XP_031291824.1; camel, XP_031291824.1; human, P53621), β'COPI (yeast, Q96WV5), spike protein (bat coronavirus 279/2005, Q0Q475; bat coronavirus Rp3/2004, Q3I5J5; bat coronavirus HKU3, Q3LZX1; bat SARS-like coronavirus WIV1, KC881007.1; pangolin coronavirus isolate, MT799526.1; civet SARS-CoV isolate SZ3, AY304486.1; MERS coronavirus isolate camel/Qatar_2_2014, KJ650098.1; MERS-CoV, K9N5Q8; SARS-CoV, P59594; SARS-CoV-2, P0DTC2).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

569 **Table 1: Crystallographic data and refinement statistics**

Structure	**αCOPI-WD40 WT**	**αCOPI-WD40 Arg^{57}→Ala**	**αCOPI-WD40 Tyr^{139}→Ala**
PDB ID	7S22	7S16	7S23
X-ray Source	23-ID-D	AMX 17-ID-1	AMX 17-ID-1
Wavelength (Å)	1.03	0.92	0.92
Temperature (K)	100	100	100
Space group	P 1 2_1 1	P 1 2_1 1	P1 2_1 1
Unit cell (Å, º)	a= 36.99, b=171.82, c=71.42; α=γ=90, β=99.57	a=35.87, b=56.71, c=70.55; α=γ=90, β=99.37	a=37.31, b=171.55, c=71.28; α=γ=90, β=99.71
Resolution (Å)	44.44-1.75 (1.78-1.75)	69.61-1.24 (1.26–1.24)	171.55 -1.49 (1.52-1.49)
$^{a}R_{merge}$ (%)	9.2 (54.7)	5.4 (57.9)	8.6(71.9)
<I/σ(I)>	8.3 (2.2)	11.5 (1.6)	5.0 (1.1)
CC1/2 (%)	98.1 (65.3)	99.8 (66.3)	99.6 (67.0)
No. of reflections	196291 (10446)	319698 (6270)	502939 (25265)
No. of unique reflections	85435 (4461)	74665(2292)	140251 (6840)
Completeness (%)	97.0 (95.4)	94.2 (58.9)	97.9 (96.8)
Redundancy	2.3 (2.3)	4.3 (2.7)	3.6 (3.7)
Refinement Statistics			
Resolution (Å)	44.48-1.75 (1.79-1.75)	43.97-1.24 (1.26-1.24)	65.02-1.49 (1.51-1.49)
No. of reflections (F>0) used in refinement	81123 (2645)	74567 (1653)	140153 (4335)
$^{b}R_{work}$ (%)	17.3	13.5	14.0
$^{c}R_{free}$ (%)	21.4	15.9	18.4
RMS bond length (Å)	0.007	0.005	0.005
RMS bond angle (°)	0.938	0.910	0.696

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Overall B value (Å^2)	21.1	21.8	21.7
Ramachandran Plot Statistics[d]			
Residues	924	321	923
Favored (%)	96.1	95.6	95.8
Allowed (%)	3.8	4.4	4.2
Disallowed (%)	0.1	0.0	0.0

570 [a]R_{merge} = [∑h∑i|Ih – Ihi|/∑h∑iIhi] where Ih is the mean of Ihi observations of reflection h.
571 Numbers in parenthesis represent highest resolution shell. [b]R_{work} and [c]R_{free} = ∑||Fobs| - |Fcalc|| /
572 ∑|Fobs| x 100 for 95% of recorded data (R_{work}) or 5% data (R_{free}). [d]From MolProbity(Williams et
573 al., 2018).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2: *In silico* mutagenesis of extended dibasic motif in SARS-CoV-2 spike

Spike hepta-peptide mutation	Change in interaction energy (ΔΔG, Rosetta units)[1]	Predicted binding change with αCOPI-WD40
Lys^{1269}→Ala	1.9	Weakened
Leu^{1270}→Ala	0.1	Moderate effect
His^{1271}→Ala	1.4	Weakened
Tyr^{1272}→Ala	0.8	Weakened
Thr^{1273}→Ala	0.1	Moderate to no effect
Thr^{1273}→Glu	-1.4	Strengthened

[1]Rosetta energy unit ΔΔG values correspond approximately to energies in kcal/mol, as seen in previous studies (Kortemme and Baker, 2002; Pierce et al., 2014), therefore predicted moderate disruptive substitutions have scores > 0.6 and highly disruptive substitutions have scores > 1.2, while predicted substantial improved binding corresponds to scores < -1.2.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

580 **Table 3: Interaction residues in *in silico* model of spike hepta-peptide with αCOPI-WD40**

Residue in SARS-CoV-2 spike hepta-peptide	αCOPI-WD40 contact residue within 3.5Å
Lys^{1269}	Asp^{96}, Asp^{115}, Tyr^{139}
Leu^{1270}	Arg^{99}
His^{1271}	Arg^{57}, Arg^{99}, Met^{141}, Leu^{157}, Asn^{216}
Tyr^{1272}	His^{31}, Arg^{57}
Thr^{1273}	Arg^{13}, Lys^{15}, His^{31}, Arg^{57}, Arg^{300}, Trp^{302}

581

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 4: Effects of *in silico* αCOPI-WD40 mutagenesis on spike hepta-peptide binding

αCOPI-WD40 mutation	**Change in interaction energy (ΔΔG, Rosetta Units)**	**Predicted binding change with αCOPI-WD40**
Lys^{15}→Ala	0.8	Weakened
His^{31}→Ala	0.8	Weakened
Arg^{57}→Ala	1.2	Weakened
Asp^{115}→Ala	0.9	Weakened
Tyr^{139}→Ala	1.1	Weakened

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends

Figure 1: Direct binding interaction of sarbecovirus spike hepta-peptide with αCOPI-WD40. **(a)** Structural conservation of αCOPI-WD40 domain determined in the present investigation (yellow) and a previous structure (magenta). Arrow highlights main chain differences between these two αCOPI-WD40 structures in Gly^{168}-Ala^{188}. **(b)** Conservation of dibasic motif in spike C-terminii of SARS-CoV, SARS-CoV-2, and MERS-CoV-2. The basic Lys and His residues are highlighted in blue. **(c-i)** BLI assay of N-biotinylated spike hepta-peptide with αCOPI-WD40 domain. One representative experiment of three is shown in panels **(c, f, g)**. Color code for concentrations is given at the bottom of the figure. **(c)** The spike wild type peptide sequence demonstrates dose-dependent binding to αCOPI-WD40 domain. **(d, e)** Mean values of equilibrium and kinetic K_D are shown. **(f)** Scrambling of the hepta-peptide sequence abolishes binding suggesting sequence-specific interaction. **(g)** Acidification enhances binding between the wild type spike hepta-peptide and αCOPI-WD40 domain. **(h, i)** Rate constants for complex association and dissociation are shown. The error bars represent the standard error and "n.d." implies not determined for weak interactions.

Figure 2: Structure-guided mutagenesis of spike hepta-peptide and binding analysis with αCOPI-WD40 domain. **(a)** *In silico* model of the spike hepta-peptide complexed with αCOPI-WD40 domain (yellow surface). The hepta-peptide is shown as a ribbon in rainbow colors from N (blue) to C (red) terminus. The Cα-atoms in the hepta-peptide are shown as spheres. The side chains of residues that interact with αCOPI-WD40 are shown as a stick. **(b-k)** BLI analysis of αCOPI-WD40 binding to spike hepta-peptide mutants. The color code of BLI traces is given at the bottom of the figure. One representative experiment of three is shown. The mutation in the spike hepta-peptide sequence is highlighted in bold and is underlined. Mutagenesis of, **(b)** Lys^{1269}, **(c)** His^{1271}, or **(d)** both abolishes binding to αCOPI-WD40. **(e)** In contrast, Tyr^{1272}→Ala mutation only weakens binding to αCOPI-WD40. The middle panel shows weak binding of αCOPI-WD40 domain with a hepta-peptide wherein Lys^{1269} has been mutated to Ala. Panels **(f-i)** show a comparative analysis of K_D and rate constants for these mutants. **(j)** Mutagenesis of Thr^{1273} to Ala in the spike hepta-peptide leads to moderately enhanced binding to αCOPI-WD40 whereas mutagenesis to Val^{1273} weakens binding **(k)**. In panels **(f-i)**, the error bars represent the standard error and "n.d." implies not determined for weak interactions. For reference, the values of kinetic and equilibrium constants for the wild type hepta-peptide sequence ("WT") are shown from **Figure 1**.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3: *In silico* and biophysical analysis of spike C-terminal position in αCOPI-WD40 binding. (a) Sequence logo generated from the alignment of K-x-H(K)-x-x sequence in 119 proteins predicted to be in the human membrane proteome. This shows the abundance of Lys in the first and third positions, low frequency of aromatic residues in the penultimate position, and the abundance of Asp and Glu in the C-terminal position. **(b)** An *in silico* model of the spike hepta-peptide on αCOPI-WD40 (yellow surface) shows an abundance of basic residues in the vicinity of the terminal Thr^{1273} spike residue. Panels **(c-i)** show results of a BLI analysis of binding between spike hepta-peptide mutants and αCOPI-WD40. Stabilization of the spike hepta-peptide complexed with αCOPI-WD40 is observed when the terminal position contains either, **(c)** acidic Glu^{1273}, or **(d)** neutral Gln^{1273} residue. **(e)** In contrast, basic Arg^{1273} in the spike hepta-peptide does not favor enhanced binding. Panels **(f-i)** show K_D and rate constants for these assays. These data implicate a role of this terminal hepta-peptide position in modulating tight binding to αCOPI-WD40. In **(c-e)**, one representative experiment of three is shown. The error bars in **(f-i)** represent the standard error and "n.d." implies not determined for weak interactions. For reference, the values of equilibrium and rate constants for the wild type hepta-peptide sequence ("WT") are shown from **Figure 1**.

Figure 4: Structure-guided mutagenesis of αCOPI-WD40 domain and binding analysis with spike hepta-peptide. Panels **(a-c)** highlight the hepta-peptide interactions (within 4Å) of **(a)** Arg^{57}, **(b)** Asp^{115}, and **(c)** Tyr^{139} residues in αCOPI-WD40 in an *in silico* model. These three interacting αCOPI-WD40 residues are shown as yellow-red-blue sticks whereas the other residues are shown as a yellow surface for simplicity. The corresponding interacting residues in the spike hepta-peptide are labelled and shown as green-red-blue sticks and spheres for Cα atoms. The BLI analysis of Arg^{57}→Ala, Asp^{115}→Ala, and Tyr^{139}→Ala mutants with the wild type spike hepta-peptide is shown in panels **(d)**, **(e)**, and **(f)** respectively. All three mutants demonstrate no substantial binding of the spike hepta-peptide. One representative experiment of three is shown in panels **(d-f)**.

Figure 5: Crystallographic analysis αCOPI-WD40 Arg^{57}→Ala and Tyr^{139}→Ala mutants. (a) Electron density around the side chains of residues Ala^{57}, Asp^{73}, and Tyr^{97} (blue mesh, 2Fo-Fc map contoured at 1.0 σ) in the structure of the Arg^{57}→Ala mutant. For simplicity, other residues are shown as a yellow surface. **(b)** Conformational changes in αCOPI-WD40 caused by Arg^{57}→Ala mutation. The mutant and wild type αCOPI-WD40 structures are shown in pink and

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

yellow as the primary colors, respectively. The Arg^{57}→Ala mutation generates a cavity in αCOPI-WD40. The nearby Tyr^{97} residue side chain rotates into this cavity. This is accompanied by the outward rotation of the Asp^{73} side chain, which is bonded to Arg^{57} in the wild type structure. **(c)** An analysis of differences in main chain conformation between the wild type and mutant αCOPI-WD40 crystal structures. The difference in Ramachandran angles was calculated for each residue, ($\delta=\sqrt{((\psi_{WT}-\psi_m)^2+(\phi_{WT}-\phi_m)^2)}$), where ($\psi_{WT}$, ϕ_{WT}) and (ψ_m, ϕ_m) are Ramachandran angles for wild type and each mutant crystal structure. This analysis shows larger conformational changes in the Arg^{57}→Ala mutant (blue) than in Tyr^{139}→Ala mutant (orange). The top six peaks are highlighted. Peak 1 corresponds to a main chain rearrangement coincident with an outward movement of Asp^{73} as shown in panel **(d)**. Cα atoms are shown as spheres in panel **(d)**. Upto 1.2Å and 0.8Å shifts in the Cα atoms are observed for Gly^{72} and Lys^{76} respectively, between the wild type and mutant structures. The intervening residues demonstrate substantial conformational rearrangement of the main chain. **(e)** In contrast, the αCOPI-WD40 Tyr^{139}→Ala mutant structure (primary color cyan) shows limited changes from the wild type structure (primary color yellow).

Acknowledgements: We thank Prof. David Owen (University of Cambridge, Cambridge, England) for critical comments, Dr. Jonathan Goldberg (Memorial Sloan Kettering Cancer Center, New York City, USA), Dr. Elena Goldberg (Memorial Sloan Kettering Cancer Center, New York City, USA), and Dr. Lauren Jackson (Vanderbilt University, Nashville USA) for advice on protein purification, Dr. Travis Gallagher (National Institute of Standards and Technology, Rockville USA) for advice on crystallographic data collection, Dr. Andrey Galkin (University of Maryland Baltimore USA) for advice on BLI assays, and Dr. Jean Jakoncic (NSLSII beamline 17-ID-1 AMX), Dr. Stephan Corcoran (APS beamline 23-ID-D), and Dr. Darren Sherrell (APS beamline 19-ID) for advice during crystallographic data collection and processing, and Ms. Corrinne Wilson for assistance in manuscript preparation. SSH acknowledges support from the University of Maryland School of Medicine, University of Maryland MPower, and MPower COVID-19 Response Fund Award. This article was supported by funds through the Maryland Department of Health's Cigarette Restitution Fund Program, University of Maryland Marlene and Stewart Greenebaum Comprehensive Cancer Center (National Cancer Institute - Cancer Center Support Grant (CCSG) - P30CA134274), The Holden Comprehensive Cancer Center at The University of Iowa and its National Cancer Institute Award P30CA086862. This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility, operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. GM/CA@APS has been funded by the National Cancer Institute (ACB-12002) and

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the National Institute of General Medical Sciences (AGM-12006, P30GM138396). Access to Sector 84 laboratories at the Advanced Protein Characterization Facility (APCF) was possible through funding provided by NIH grant GM115586 and DOE contract DE-AC02-06CH11357. This research used resources AMX of the National Synchrotron Light Source II, a U.S. DOE Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under contract no. DE-SC0012704.

Author contributions:

DD: Data acquisition, analysis, interpretation, and manuscript preparation.

SS: Data acquisition, analysis, interpretation, and manuscript preparation.

SK: Data acquisition, analysis, and interpretation.

MM: Data acquisition.

NJS: Data acquisition, analysis, interpretation, and manuscript preparation.

LG: Data acquisition, analysis, interpretation, and manuscript preparation.

BGP: Data acquisition, analysis, interpretation, and manuscript preparation.

SSH: Conception, data acquisition, analysis, interpretation, and manuscript preparation.

Competing interests: The authors declare no competing interests.

Data availability: Coordinates for the crystal structures have been deposited in the Protein Data Bank with IDs: 7S22 (wild type), 7S16 (Arg^{57}→Ala), and 7S23 (Tyr^{139}→Ala). The plasmids for αCOPI-WD40, αCOPI-WD40 mutants (Arg^{57}→Ala and Tyr^{139}→Ala), and GST-β'COPI-WD40 expression will be deposited in www.addgene.com.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References:

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr *68*, 352-367.

Ahmed, A.E. (2017). The predictors of 3- and 30-day mortality in 660 MERS-CoV patients. BMC Infect Dis *17*, 615.

Beniac, D.R., Andonov, A., Grudeski, E., and Booth, T.F. (2006). Architecture of the SARS coronavirus prefusion spike. Nat Struct Mol Biol *13*, 751-752.

Bolles, M., Donaldson, E., and Baric, R. (2011). SARS-CoV and emergent coronaviruses: viral determinants of interspecies transmission. Curr Opin Virol *1*, 624-634.

Cattin-Ortolá, J., Welch, L., Maslen, S.L., Skehel, J.M., Papa, G., James, L.C., and Munro, S. (2021). Sequences in the cytoplasmic tail of SARS-CoV-2 Spike facilitate expression at the cell surface and syncytia formation. bioRxiv, 2020.2010.2012.335562.

Conway, P., Tyka, M.D., DiMaio, F., Konerding, D.E., and Baker, D. (2014). Relaxation of backbone bond geometry improves protein energy landscape modeling. Protein Sci *23*, 47-55.

Cosson, P., and Letourneur, F. (1994). Coatomer interaction with di-lysine endoplasmic reticulum retention motifs. Science *263*, 1629-1631.

Dodonova, S.O., Diestelkoetter-Bachert, P., von Appen, A., Hagen, W.J., Beck, R., Beck, M., Wieland, F., and Briggs, J.A. (2015). VESICULAR TRANSPORT. A structure of the COPI coat and the role of coat proteins in membrane vesicle assembly. Science *349*, 195-198.

Duden, R., Griffiths, G., Frank, R., Argos, P., and Kreis, T.E. (1991). Beta-COP, a 110 kd protein associated with non-clathrin-coated vesicles and the Golgi complex, shows homology to beta-adaptin. Cell *64*, 649-665.

Emsley, P., and Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Eugster, A., Frigerio, G., Dale, M., and Duden, R. (2004). The alpha- and beta'-COP WD40 domains mediate cargo-selective interactions with distinct di-lysine motifs. Mol Biol Cell *15*, 1011-1023.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr *69*, 1204-1214.

Fiedler, K., Veit, M., Stamnes, M.A., and Rothman, J.E. (1996). Bimodal interaction of coatomer with the p24 family of putative cargo receptors. Science *273*, 1396-1399.

Gaynor, E.C., te Heesen, S., Graham, T.R., Aebi, M., and Emr, S.D. (1994). Signal-mediated retrieval of a membrane protein from the Golgi to the ER in yeast. J Cell Biol *127*, 653-665.

Ge, X.Y., Li, J.L., Yang, X.L., Chmura, A.A., Zhu, G., Epstein, J.H., Mazet, J.K., Hu, B., Zhang, W., Peng, C.*, et al.* (2013). Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature *503*, 535-538.

Gottlieb, R.A., Nordberg, J., Skowronski, E., and Babior, B.M. (1996). Apoptosis induced in Jurkat cells by several agents is preceded by intracellular acidification. Proc Natl Acad Sci U S A *93*, 654-658.

Guan, Y., Zheng, B.J., He, Y.Q., Liu, X.L., Zhuang, Z.X., Cheung, C.L., Luo, S.W., Li, P.H., Zhang, L.J., Guan, Y.J.*, et al.* (2003). Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science *302*, 276-278.

Han, H.J., Yu, H., and Yu, X.J. (2016). Evidence for zoonotic origins of Middle East respiratory syndrome coronavirus. J Gen Virol *97*, 274-280.

Hara-Kuge, S., Kuge, O., Orci, L., Amherdt, M., Ravazzola, M., Wieland, F.T., and Rothman, J.E. (1994). En bloc incorporation of coatomer subunits during the assembly of COP-coated vesicles. J Cell Biol *124*, 883-892.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Harrison-Lavoie, K.J., Lewis, V.A., Hynes, G.M., Collison, K.S., Nutland, E., and Willison, K.R. (1993). A 102 kDa subunit of a Golgi-associated particle has homology to beta subunits of trimeric G proteins. EMBO J *12*, 2847-2853.

Jackson, L.P., Lewis, M., Kent, H.M., Edeling, M.A., Evans, P.R., Duden, R., and Owen, D.J. (2012). Molecular basis for recognition of dilysine trafficking motifs by COPI. Dev Cell *23*, 1255-1262.

Jackson, M.R., Nilsson, T., and Peterson, P.A. (1993). Retrieval of transmembrane proteins to the endoplasmic reticulum. J Cell Biol *121*, 317-333.

Jennings, B.C., Kornfeld, S., and Doray, B. (2021). A weak COPI binding motif in the cytoplasmic tail of SARS-CoV-2 spike glycoprotein is necessary for its cleavage, glycosylation, and localization. FEBS Lett *595*, 1758-1767.

Kabsch, W. (2010). Xds. Acta Crystallogr D Biol Crystallogr *66*, 125-132.

Klumperman, J., Locker, J.K., Meijer, A., Horzinek, M.C., Geuze, H.J., and Rottier, P.J. (1994). Coronavirus M proteins accumulate in the Golgi complex beyond the site of virion budding. J Virol *68*, 6523-6534.

Kortemme, T., and Baker, D. (2002). A simple physical model for binding energy hot spots in protein-protein complexes. Proc Natl Acad Sci U S A *99*, 14116-14121.

Latinne, A., Hu, B., Olival, K.J., Zhu, G., Zhang, L., Li, H., Chmura, A.A., Field, H.E., Zambrana-Torrelio, C., Epstein, J.H.*, et al.* (2020). Origin and cross-species transmission of bat coronaviruses in China. Nat Commun *11*, 4235.

Lau, S.K., Woo, P.C., Li, K.S., Huang, Y., Tsoi, H.W., Wong, B.H., Wong, S.S., Leung, S.Y., Chan, K.H., and Yuen, K.Y. (2005). Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci U S A *102*, 14040-14045.

Letourneur, F., Gaynor, E.C., Hennecke, S., Demolliere, C., Duden, R., Emr, S.D., Riezman, H., and Cosson, P. (1994). Coatomer is essential for retrieval of dilysine-tagged proteins to the endoplasmic reticulum. Cell *79*, 1199-1207.

Li, S., Zhang, Y., Guan, Z., Li, H., Ye, M., Chen, X., Shen, J., Zhou, Y., Shi, Z.L., Zhou, P.*, et al.* (2020a). SARS-CoV-2 triggers inflammatory responses and cell death through caspase-8 activation. Signal Transduct Target Ther *5*, 235.

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J.H., Wang, H., Crameri, G., Hu, Z., Zhang, H.*, et al.* (2005). Bats are natural reservoirs of SARS-like coronaviruses. Science *310*, 676-679.

Li, Z., Liu, G., Wang, L., Liang, Y., Zhou, Q., Wu, F., Yao, J., and Chen, B. (2020b). From the insight of glucose metabolism disorder: Oxygen therapy and blood glucose monitoring are crucial for quarantined COVID-19 patients. Ecotoxicol Environ Saf *197*, 110614.

Lontok, E., Corse, E., and Machamer, C.E. (2004). Intracellular targeting signals contribute to localization of coronavirus spike proteins near the virus assembly site. J Virol *78*, 5913-5922.

Ma, W., and Goldberg, J. (2013). Rules for the recognition of dilysine retrieval motifs by coatomer. EMBO J *32*, 926-937.

Malhotra, V., Serafini, T., Orci, L., Shepherd, J.C., and Rothman, J.E. (1989). Purification of a novel class of coated vesicles mediating biosynthetic protein transport through the Golgi stack. Cell *58*, 329-336.

McBride, C.E., Li, J., and Machamer, C.E. (2007). The cytoplasmic tail of the severe acute respiratory syndrome coronavirus spike protein contains a novel endoplasmic reticulum retrieval signal that binds COPI and promotes interaction with membrane protein. J Virol *81*, 2418-2428.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006). HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. Acta Crystallogr D Biol Crystallogr *62*, 859-866.

Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P.C., Timony, M.A., Ottaviano, M., and Sliz, P. (2013). Collaboration gets the most out of software. Elife *2*, e01456.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Pierce, B.G., Hellman, L.M., Hossain, M., Singh, N.K., Vander Kooi, C.W., Weng, Z., and Baker, B.M. (2014). Computational design of the affinity and specificity of a therapeutic T cell receptor. PLoS Comput Biol *10*, e1003478.

Polack, F.P., Thomas, S.J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J.L., Perez Marc, G., Moreira, E.D., Zerbini, C.*, et al.* (2020). Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. N Engl J Med *383*, 2603-2615.

Raj, V.S., Farag, E.A., Reusken, C.B., Lamers, M.M., Pas, S.D., Voermans, J., Smits, S.L., Osterhaus, A.D., Al-Mawlawi, N., Al-Romaihi, H.E.*, et al.* (2014). Isolation of MERS coronavirus from a dromedary camel, Qatar, 2014. Emerg Infect Dis *20*, 1339-1342.

Roess, A., Carruth, L., Lahm, S., and Salman, M. (2016). Camels, MERS-CoV, and other emerging infections in east Africa. Lancet Infect Dis *16*, 14-15.

Rossmann, M.G., and Blow, D.M. (1962). The detection of sub-units within the crystallographic asymmetric unit. Acta Crystallographica *15*, 24-31.

Schroder-Kohne, S., Letourneur, F., and Riezman, H. (1998). Alpha-COP can discriminate between distinct, functional di-lysine signals in vitro and regulates access into retrograde transport. J Cell Sci *111 (Pt 23)*, 3459-3470.

Serafini, T., Stenbeck, G., Brecht, A., Lottspeich, F., Orci, L., Rothman, J.E., and Wieland, F.T. (1991). A coat subunit of Golgi-derived non-clathrin-coated vesicles with homology to the clathrin-coated vesicle coat protein beta-adaptin. Nature *349*, 215-220.

Song, H.C., Seo, M.Y., Stadler, K., Yoo, B.J., Choo, Q.L., Coates, S.R., Uematsu, Y., Harada, T., Greer, C.E., Polo, J.M.*, et al.* (2004). Synthesis and characterization of a native, oligomeric form of recombinant severe acute respiratory syndrome coronavirus spike glycoprotein. J Virol *78*, 10328-10335.

Stadler, K., Masignani, V., Eickmann, M., Becker, S., Abrignani, S., Klenk, H.D., and Rappuoli, R. (2003). SARS--beginning to understand a new virus. Nat Rev Microbiol *1*, 209-218.

Stenbeck, G., Harter, C., Brecht, A., Herrmann, D., Lottspeich, F., Orci, L., and Wieland, F.T. (1993). beta'-COP, a novel subunit of coatomer. EMBO J *12*, 2841-2845.

Tang, X.C., Zhang, J.X., Zhang, S.Y., Wang, P., Fan, X.H., Li, L.F., Li, G., Dong, B.Q., Liu, W., Cheung, C.L.*, et al.* (2006). Prevalence and genetic diversity of coronaviruses in bats from China. J Virol *80*, 7481-7490.

Thul, P.J., Akesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Bjork, L., Breckels, L.M.*, et al.* (2017). A subcellular map of the human proteome. Science *356*.

Townsley, F.M., and Pelham, H.R. (1994). The KKXX signal mediates retrieval of membrane proteins from the Golgi to the ER in yeast. Eur J Cell Biol *64*, 211-216.

USFDA (2020). U.S. Food and Drug Administration. Moderna COVID-19 vaccine [FDA briefing document]. Silver Spring, MD: U.S. Food and Drug Administration, Vaccines and Related Biological Products Advisory Committee; 2020.

Vece, T.J., Watkin, L.B., Nicholas, S., Canter, D., Braun, M.C., Guillerman, R.P., Eldin, K.W., Bertolet, G., McKinley, S., de Guzman, M.*, et al.* (2016). Copa Syndrome: a Novel Autosomal Dominant Immune Dysregulatory Disease. J Clin Immunol *36*, 377-387.

Waters, M.G., Serafini, T., and Rothman, J.E. (1991). 'Coatomer': a cytosolic protein complex containing subunits of non-clathrin-coated Golgi transport vesicles. Nature *349*, 248-251.

Watkin, L.B., Jessen, B., Wiszniewski, W., Vece, T.J., Jan, M., Sha, Y., Thamsen, M., Santos-Cortez, R.L., Lee, K., Gambin, T.*, et al.* (2015). COPA mutations impair ER-Golgi transport and cause hereditary autoimmune-mediated lung disease and arthritis. Nat Genet *47*, 654-660.

Webb, B., and Sali, A. (2014). Protein structure modeling with MODELLER. Methods Mol Biol *1137*, 1-15.

Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B.*, et al.* (2018). MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci *27*, 293-315.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.03.458953; this version posted September 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Xiao, K., Zhai, J., Feng, Y., Zhou, N., Zhang, X., Zou, J.J., Li, N., Guo, Y., Li, X., Shen, X.*, et al.* (2020). Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. Nature *583*, 286-289.

Yan, H., Xiao, G., Zhang, J., Hu, Y., Yuan, F., Cole, D.K., Zheng, C., and Gao, G.F. (2004). SARS coronavirus induces apoptosis in Vero E6 cells. J Med Virol *73*, 323-331.

Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D., and Fouchier, R.A. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med *367*, 1814-1820.

Zhang, T., Wu, Q., and Zhang, Z. (2020). Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. Curr Biol *30*, 1346-1351 e1342.

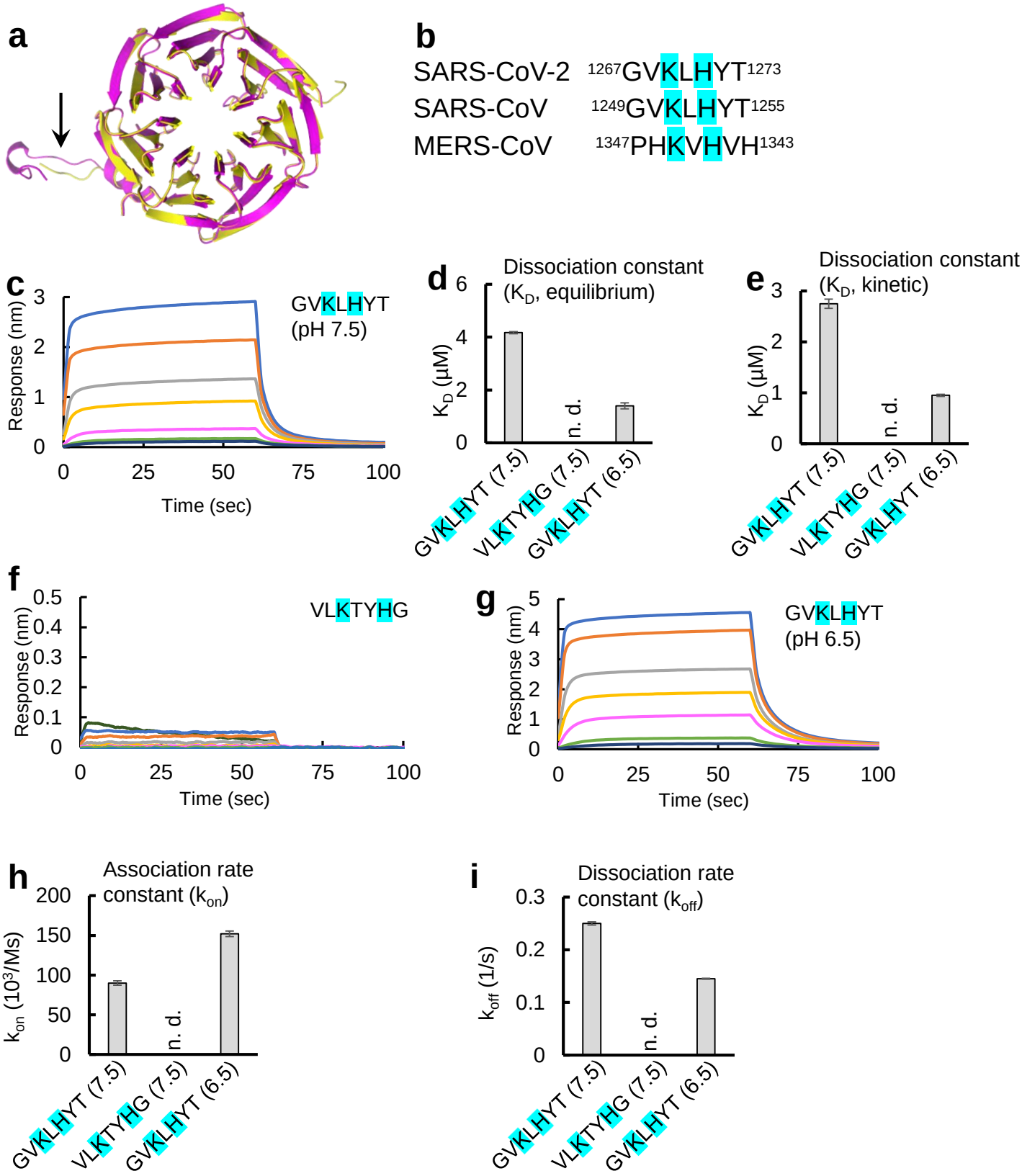

a
b
SARS-CoV-2 1267GVKLHYT1273
SARS-CoV 1249GVKLHYT1255
MERS-CoV 1347PHKVHVH1343
c
Response (nm)
Time (sec)
GVKLHYT
(pH 7.5)
d
Dissociation constant
(K_D, equilibrium)
K_D (μM)
n. d.
GVKLHYT (7.5)
VLKTYHG (7.5)
GVKLHYT (6.5)
e
Dissociation constant
(K_D, kinetic)
K_D (μM)
n. d.
GVKLHYT (7.5)
VLKTYHG (7.5)
GVKLHYT (6.5)
f
Response (nm)
Time (sec)
VLKTYHG
g
Response (nm)
Time (sec)
GVKLHYT
(pH 6.5)
h
Association rate
constant (k_on)
k_on (10³/Ms)
n. d.
GVKLHYT (7.5)
VLKTYHG (7.5)
GVKLHYT (6.5)
i
Dissociation rate
constant (k_off)
k_off (1/s)
n. d.
GVKLHYT (7.5)
VLKTYHG (7.5)
GVKLHYT (6.5)
8μM
5μM
2μM
1μM
0.3μM
0.1μM
0.05μM

Fig 1

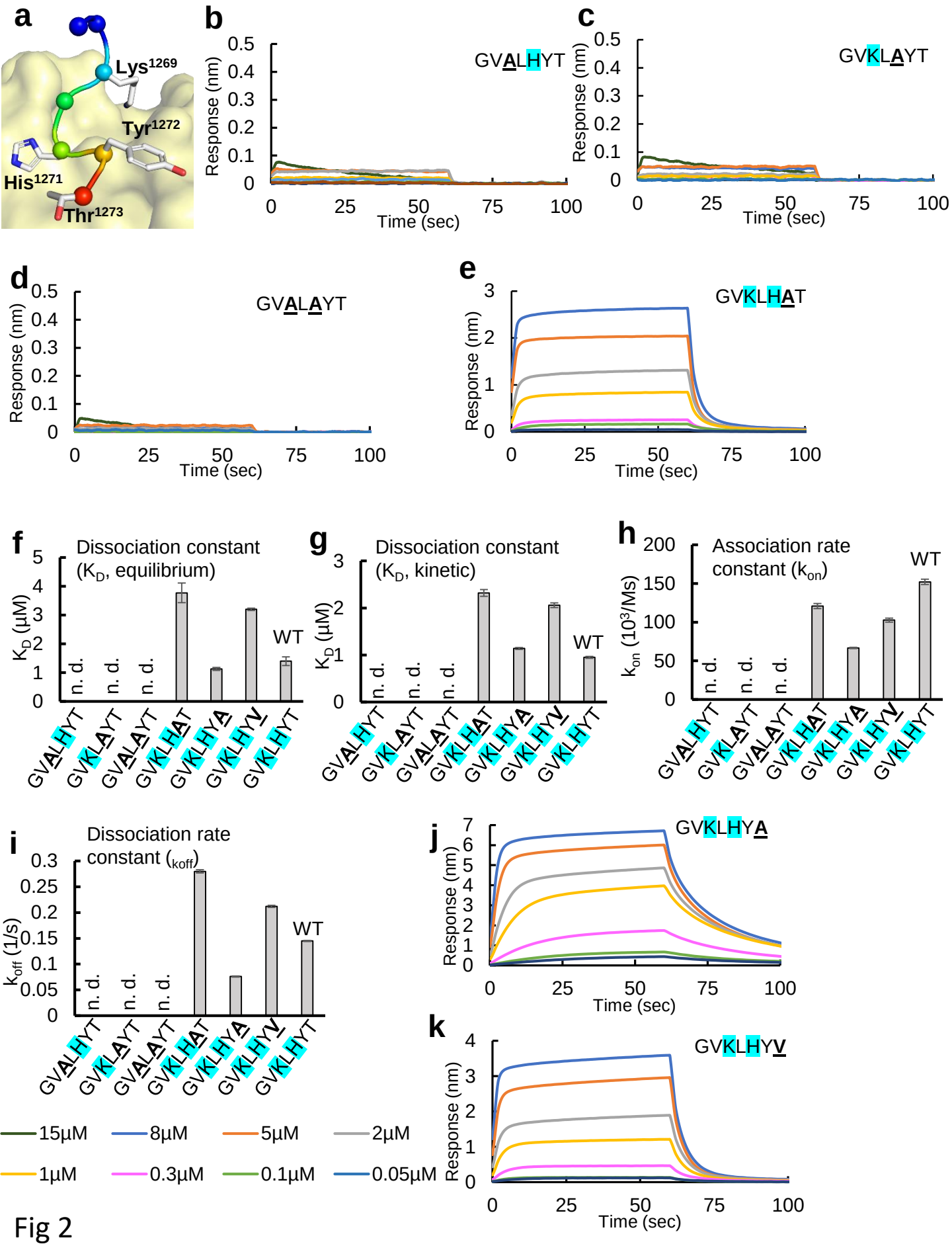

a
Lys1269
Tyr1272
His1271
Thr1273
b
GVALHYT
Response (nm)
Time (sec)
c
GVKLAYT
d
GVALAYT
e
GVKLHAT
f
Dissociation constant
(KD, equilibrium)
KD (μM)
n. d.
WT
GVALHYT
GVKLAYT
GVALAYT
GVKLHAT
GVKLHYA
GVKLHYV
GVKLHYT
g
Dissociation constant
(KD, kinetic)
h
Association rate
constant (kon)
kon (10³/Ms)
i
Dissociation rate
constant (koff)
koff (1/s)
j
GVKLHYA
k
GVKLHYV
15μM
8μM
5μM
2μM
1μM
0.3μM
0.1μM
0.05μM

Fig 2

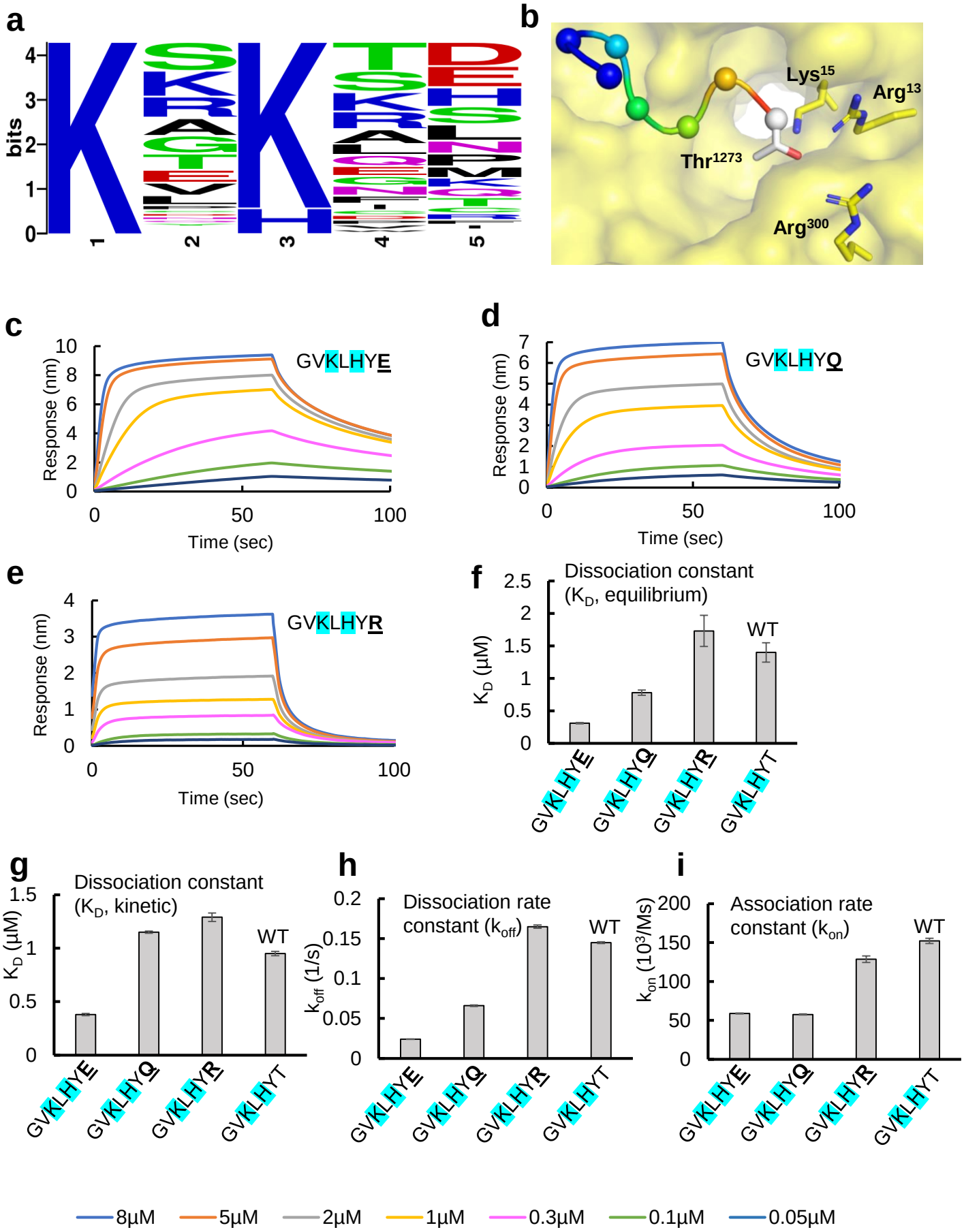

a
bits
1
2
3
4
5
b
Lys15
Arg13
Thr1273
Arg300
c
GVKLHYE
Response (nm)
Time (sec)
d
GVKLHYQ
Response (nm)
Time (sec)
e
GVKLHYR
Response (nm)
Time (sec)
f
Dissociation constant
(KD, equilibrium)
KD (μM)
WT
GVKLHYE
GVKLHYQ
GVKLHYR
GVKLHYT
g
Dissociation constant
(KD, kinetic)
KD (μM)
WT
h
Dissociation rate
constant (koff)
koff (1/s)
WT
i
Association rate
constant (kon)
kon (10³/Ms)
WT
8μM
5μM
2μM
1μM
0.3μM
0.1μM
0.05μM

Fig 3

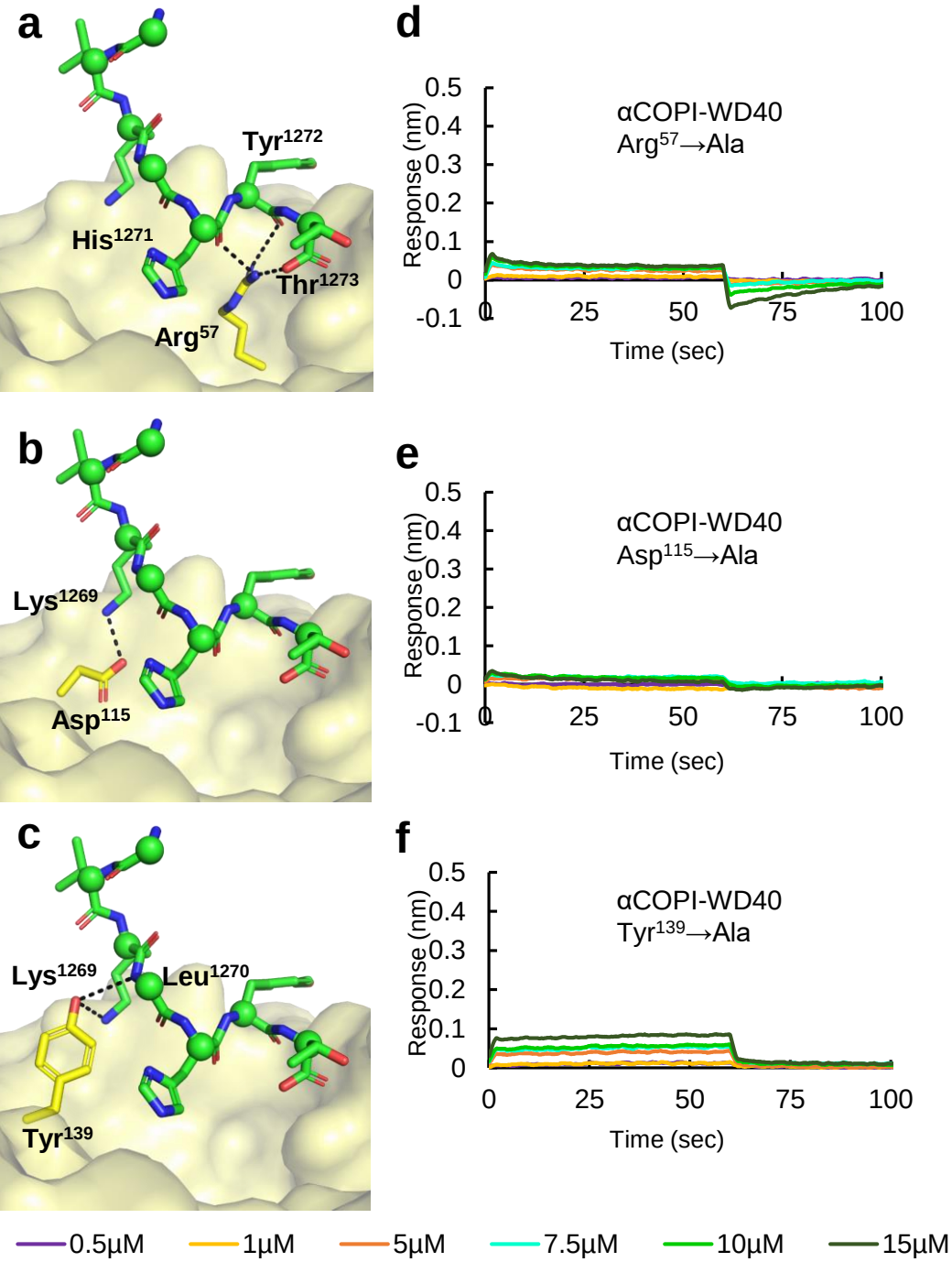
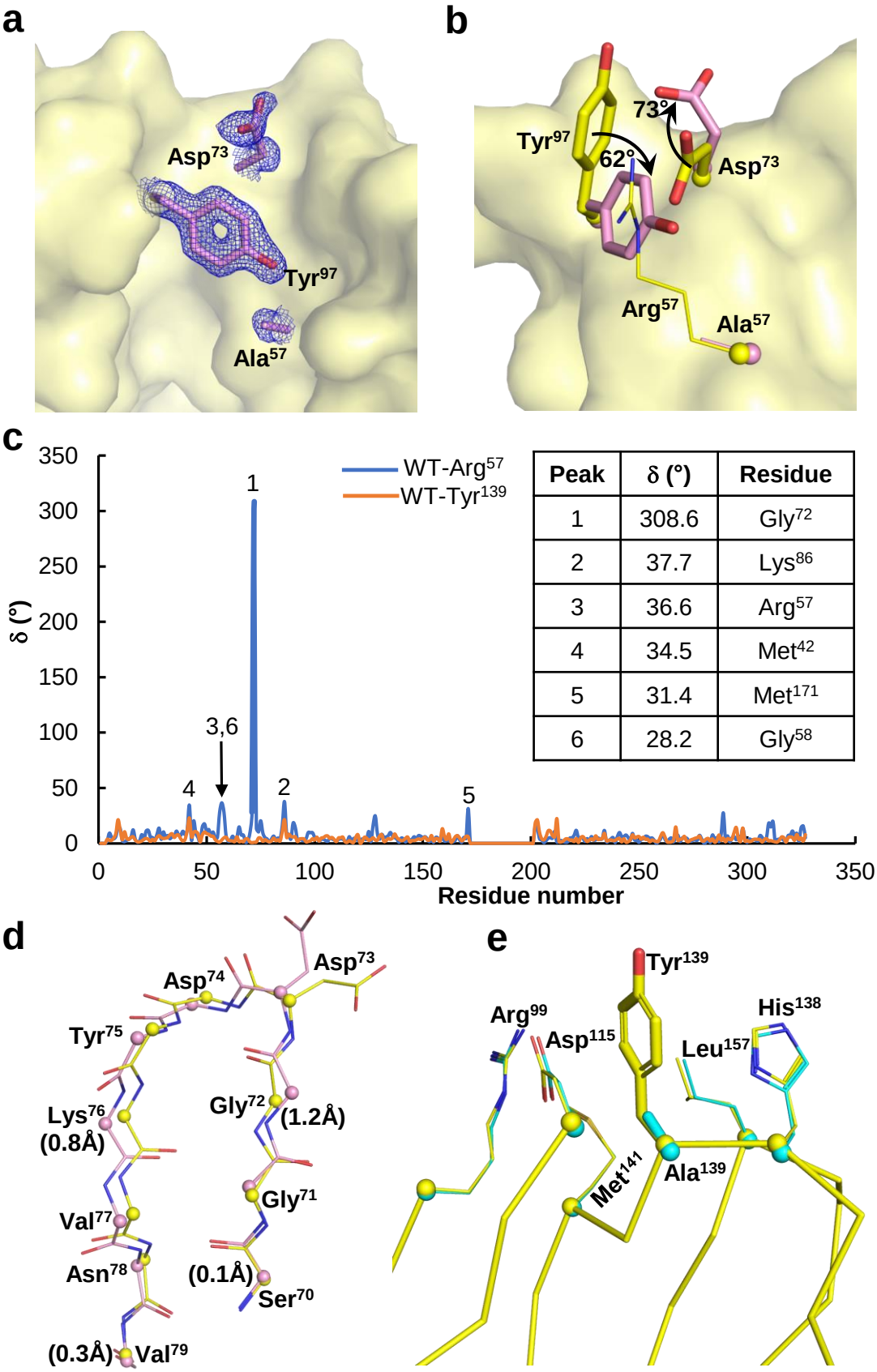

a
Tyr1272
His1271
Thr1273
Arg57
b
Lys1269
Asp115
c
Lys1269
Leu1270
Tyr139
d
αCOPI-WD40
Arg57→Ala
Response (nm)
Time (sec)
e
αCOPI-WD40
Asp115→Ala
Response (nm)
Time (sec)
f
αCOPI-WD40
Tyr139→Ala
Response (nm)
Time (sec)
0.5μM
1μM
5μM
7.5μM
10μM
15μM

Fig 4



Peak	δ (°)	Residue
1	308.6	Gly72
2	37.7	Lys86
3	36.6	Arg57
4	34.5	Met42
5	31.4	Met171
6	28.2	Gly58

Fig 5