UCSF
UC San Francisco Electronic Theses and Dissertations

Title
Molecular phenomenology of membrane-bound protein machines

Permalink
https://escholarship.org/uc/item/30b367dp

Author
Asarnow, Daniel

Publication Date
2021

Supplemental Material
https://escholarship.org/uc/item/30b367dp#supplemental

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Molecular phenomenology of membrane-bound protein machines

by
Daniel Asarnow

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biophysics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Yifan Cheng
87A0F4E0B50C4F6...

Yifan Cheng
Chair

DocuSigned by:
James Fraser

James Fraser

DocuSigned by:
ADAM FROST
1E2FF75D7B5A48F...

ADAM FROST

Committee Members

Copyright 2021

by

Daniel Asarnow

Acknowledgements

I first thank my Ph.D. advisor, Professor Yifan Cheng, for providing me with a scientific home these past six years, for endless encouragement, and for unfailing support in my academic endeavors. Without his wisdom, compassion, and boundless enthusiasm for discovery, this dissertation would not have been possible. I also thank Professor Philip A. Beachy of Stanford University, a close collaborator and advisor for most of my Ph.D., who introduced me to Hedgehog pathway structural biology and has taught me so much of what it means to be a scientist. I am likewise very deeply grateful to my direct collaborators and co-first authors, Drs. Qianqian Wang and Bei Wang (no relation), thank you so much for welcoming me into your scientific lives. It has truly been an honor to work with you.

I must next acknowledge my other friends and mentors among the faculty at UCSF: Professors Charles S. Craik, Aashish Manglik, Robert Stroud, David Julius, Adam Frost, and James S. Fraser. Each of you has contributed invaluable lessons that I will bring forward consciously into my future efforts in independent research and mentorship.

I also thank my many friends and colleagues at who contributed so much to my professional training and personal journey at UCSF. In particular, Zanlin Yu, David P. Bulkley, Alexander Myasnikov, and Sam Li for their indispensable technical assistance on the microscope.

To my postdoc mentors Henriette E. Autzen, Amber M. Smith, Melody G. Campbell, and Meghna Gupta, you have taught me so, so much and I will be indebted to you forever. To my labmates and dear friends, Evan M. Green and Eugene J. Palovcak, Kaihua Zhang, Jianhua Zhao, Bryan Faust, Adamo Mancino, and Marcell Zimanyi; to my partner in crime Raghav Kalia; to my classmates Paul V. Thomas, Cole V.M. Helsell, Nat Hendel, Nadja Kern, Sergei Pourmal; to my friend and confidante Amanda Loshbaugh: I am so excited to continue knowing you all as we

embark on the next chapters of our lives. I would be remiss not to also credit Benjamin Barad and Ilan Chemmama for helping me maintain a sense of humility through many lessons on the Bakar squash courts, as well as Amanda's dog Olie, who was truly the goodest boy.

Before I came to UCSF, I was fortunate to work with the other aspiring young scientists in Professor Raymond Esquerra's lab at San Francisco State University. Pooncharas Tipgunlakant, David Poole, Bushra Mariam, Dana Kennedy, Benjamin Rodriguez, Damon Robles, and Ignacio Lopez befriended me, helped me, and taught me how to make real scientific contributions by any means necessary. I am also deeply indebted to my earliest mentors in academic research, Drs. Eefei Chen, Istvan Szundi, and Robert Goldbeck, who welcomed me into Professor David Kliger and Professor Olaf Einarsdottir's groups at UC Santa Cruz.

Going back farther, I must also recognize the great impact of my time at the Exploratorium, San Francisco's museum of science, art, and human perception. As a high school "explainer" there I learned how to *do* hands-on science education from the best operators in the business, including Darlene Librero, Ron Hipschman, Paul Doherty, and Tom Thompkins, as well as fellow explainers Dexter Red, Kai Flemming, and Zoë Clarke. At the graduate level I still teach using our methods.

To my partner, Keely, I love you and I'm so fortunate to have been in it together with you through the last four years. And to my oldest friends, Chris and Sara and Justin…see you soon.

Above all, I thank my family for their unwavering love and support throughout my 34-year journey from 2383B Virginia St. to the conclusion of my doctorate. It is my parents, David Asarnow and Mary Taylor-Asarnow, who have taught me and continue to teach me that life's highest values are love and righteousness, and my sisters, Sarah and Lizzy, who have challenged me to put those values into practice.

This dissertation is dedicated to the memory of my grandmother, Carmel Hara.

Contributions

Chapter 2

D. Asarnow, B. Wang et al., "Structural insight into SARS-CoV-2 neutralizing antibodies and modulation of syncytia," *Cell*, vol. 184, no. 12, pp. 3192-3204.e16, 2021.

Chapter 3

Q. Wang, D. Asarnow, et al., "Dispatched uses Na^+ flux to power lipid-modified Hedgehog release," *Nature*, in press, 2021.

Molecular phenomenology of membrane-bound protein machines

Daniel Asarnow

Abstract

The basic functions of living things are predominantly carried out by uniquely adapted protein nanomachines. Molecular phenomenology is an approach to biology driven by the most basic questions about these species: what are their atomic structures, and how do they move? The past two centuries have seen great strides in biology, yet we are only beginning to understand its fundamental units of function. In the past decade, revolutionary advances in electron microscopy have enabled direct imaging of biological molecules in solution, opening what may be the final frontier of molecular phenomenology. My doctoral work has focused especially on applying single-particle electron microscopy to the dynamics of those protein machines that lie at the plasma membrane that divides cells from their environment.

In the first chapter of this dissertation I introduce a previously unpublished method for modeling protein motions through multi-body refinement and principal component analysis with geodesic kernels. While multi-body analysis is not new, other methods have not considered the intrinsic nature of rigid-body motions as a non-linear manifold. As such, they have been forced to rely on *ad-hoc* parameter scaling, and are valid only over very small motions where manifold curvature can be neglected. The kernel approach, in contrast, eliminates the parameter scaling problem and is applicable to arbitrary rigid-body motions. The basic method may be of use outside cryo-EM of biological molecules, for example in 3D animation of articulated figures or controlling swarms of drones or satellites.

In the second chapter, I and others apply molecular phenomenology to uncover the mechanisms of action underlying antibody modulation of the fusogenic activity of the SARS-CoV-

2 Spike protein. This work was the first study demonstrating that Spike-reactive antibodies can either inhibit or enhance Spike-mediated membrane fusion leading to viral entry or pathological cell-cell fusion and formation of syncytia (giant, multinucleated cells). The results epitomize the deep connection between basic biophysical understanding and biological function, by tying molecular structure and conformation to the virulence of an emergent pathogen responsible for millions of deaths and hundreds of millions of injuries.

In the third chapter, I and others report high resolution structures of the murine Dispatched and its complex with Sonic hedgehog. I then use 3D variability analysis, a state-of-the-art phenomenological method, to uncover coupling between transmembrane ion flux and release of Sonic hedgehog from organizer cells during animal development. While multi-body refinement and 3D variability analysis both visualize interdomain motions, the latter also captures intradomain conformational changes by direct 3D reconstruction of orthogonal axes of particle covariance. This approach enables the production of true molecular movies that go beyond animations based on multi-body refinement. Dispatched bioactivity is characterized by subtle conformational rearrangements in its transmembrane domain that direct the energy of the plasma membrane Na^+ gradient towards extraction of lipid-modified Sonic hedgehog from the plasma membrane and activation of Sonic hedgehog for efficient recruitment of its critical adapter protein SCUBE2. High-resolution molecular movies of active Dispatched, validation through traditional 3D classification, and detailed mutational analysis using a direct read-out of protein function yield one of the most complete studies in molecular phenomenology to-date, foreshadowing a new era of biophysics in which modeling dynamic structural ensembles may become *de rigueur*.

Table of Contents

Geodesic kernels for multi-body analysis 1

ABSTRACT 2

PRINCIPLES OF MULTI-BODY REFINEMENT 2

MULTI-BODY PHENOMENOLOGY OF FLEXIBLE STRUCTURES 5

METHODOLOGY 9

FIGURES 14

SUPPLEMENTAL FILES 19

ACKNOWLEDGEMENTS 20

REFERENCES 21

Structural insight into SARS-CoV-2 neutralizing antibodies and modulation of syncytia 24

ABSTRACT 26

INTRODUCTION 26

RESULTS 28

DISCUSSION 36

FIGURES 39

TABLES 57

SUPPLEMENTAL FILES 62

METHODS 63

ACKNOWLEDGEMENTS 76

AUTHOR CONTRIBUTIONS 76

REFERENCES 78

Dispatched uses Na^+ flux to power lipid-modified Hedgehog release 86

ABSTRACT 87

INTRODUCTION 87

RESULTS 88

DISCUSSION 94

SUPPLEMENTAL DISCUSSION 96

FIGURES 99

TABLES 120

SUPPLEMENTAL FILES 121

METHODS 122

DATA AVAILABILITY 130

ACKNOWLEDGEMENTS 130

AUTHOR CONTRIBUTIONS 131

REFERENCES 132

List of Figures

FIGURE 1.1 PRINCIPLES OF MULTI-BODY REFINEMENT 15

FIGURE 1.2 RESIDUAL ALIGNMENTS LIE ON THE NON-LINEAR MANIFOLD SE(3) 16

FIGURE 1.3 KERNEL PCA OF TRPM4 MHR1-2 MULTI-BODY REFINEMENT DATA 17

FIGURE 2.1 ISOLATION OF SARS-COV-2 RECEPTOR-BLOCKING ANTIBODIES FROM A NAÏVE HUMAN LIBRARY 39

FIGURE 2.2 SARS-COV-2 NEUTRALIZATION BY RECEPTOR BLOCKING ANTIBODIES 40

FIGURE 2.3 ANTI-SARS-COV-2 SPIKE RBD IGG ANTIBODIES AFFECT TRYPSIN INDUCED CELL SYNCYTIA FORMATION 42

FIGURE 2.4 STRUCTURES OF SPIKE-FAB COMPLEXES 43

FIGURE 2.5 BINDING MODE AND EPITOPE OF 5A6 45

FIGURE 2.6 SPIKE FUNCTIONAL MODULATION BY RECEPTOR-BLOCKING ANTIBODIES 46

FIGURE 2.7 CHARACTERIZATIONS OF SARS-COV-2 RECEPTOR-BLOCKING ANTIBODIES 47

FIGURE 2.8 SARS-COV-2 LIVE VIRUS NEUTRALIZATION AND ANTIBODY BINDING TO SPIKE TRIMER MEASURED BY SURFACE PLASMON RESONANCE 48

FIGURE 2.9 CRYO-EM DENSITIES AND RESOLUTION ESTIMATION 50

FIGURE 2.10 ADDITIONAL STRUCTURAL DETAILS 52

FIGURE 2.11 CRYO-EM PROCESSING 54

FIGURE 2.12 NEUTRALIZATION OF THE SARS-COV-2 PSEUDOVIRUS WITH SPIKE MUTANT D614G 56

FIGURE 3.1 STRUCTURE AND LIPID BINDING SITES OF DISP1 99

FIGURE 3.2 THREE Na^+ DENSITIES IN THE TRANSMEMBRANE DOMAIN. 100

FIGURE 3.3 A TRANSMEMBRANE Na^+ PERMEATION CHANNEL IS DISRUPTED IN THE STRUCTURE OF DISP1-A-NNN 101

FIGURE 3.4 DISP1-A BINDS SHHN BETWEEN ITS DISTAL ECDS 102

FIGURE 3.5 STRUCTURALLY RELATED TRANSPORTERS DISPATCHED AND PATCHED IN HEDGEHOG SIGNALING 103

FIGURE 3.6 DISPATCHED PROTEIN SEQUENCE ALIGNMENT AND STRUCTURAL FEATURES 105

FIGURE 3.7 CRYO-EM DATA AND IMAGE PROCESSING FLOW FOR DISP1-A AND DISP1-A-NNN 107

FIGURE 3.8 CRYO-EM DENSITY AND ATOMIC MODEL QUALITY 109

FIGURE 3.9 COINCIDENT MODES OF DISP1 CONFORMATIONAL FLEXIBILITY REVEALED BY 3D CLASSIFICATION AND 3D VARIABILITY ANALYSIS (3DVA) 110

FIGURE 3.10 REPRESENTATIVE CRYO-EM DENSITIES FROM SELECTED STRUCTURAL FEATURES 111

FIGURE 3.11 Na^+-COORDINATING AMINO ACID RESIDUES IN DISP1, AND CONFORMATIONAL REARRANGEMENTS WITHIN THE TRANSMEMBRANE Na^+ PATHWAY IN DISP1-A-NNN 113

FIGURE 3.12 QUANTITATIVE LUCIFERASE-BASED ASSAY OF SHHNP RELEASE 114

FIGURE 3.13 CRYO-EM DATA AND IMAGE PROCESSING FLOW FOR SHHN:DISP1-A COMPLEX, AND COMPARISON TO *DROSOPHILA* HHN:DISP COMPLEX 115

FIGURE 3.14 CONFORMATIONAL DYNAMICS LINK INTRAMEMBRANE Na^+ SITE OCCUPANCY TO HEDGEHOG RELEASE 117

FIGURE 3.15 Na^+ ION UTILIZATION BY DISP1, PTCH1, AND OTHER MEMBERS OF THE RND TRANSPORTER FAMILY 118

List of Tables

TABLE 2.1 PERCENT FRAMEWORK IDENTITY TO THE GERMLINE SEQUENCES 57

TABLE 2.2 BINDING KINETICS OF ANTIBODIES AND FABS MEASURED AGAINST FC-RBD BY BLI OR SPIKE TRIMER BY SPR 58

TABLE 2.3 CRYO-EM DATA COLLECTION, REFINEMENT AND VALIDATION STATISTICS 59

TABLE 2.5 KEY RESOURCES 60

TABLE 3.1 CRYO-EM DATA COLLECTION, REFINEMENT AND VALIDATION STATISTICS 120

Chapter 1

Geodesic kernels for multi-body analysis

Contributing Authors

Daniel Asarnow[1,2], Yifan Cheng[1,2]

[1]Department of Biochemistry and Biophysics, University of California San Francisco (UCSF) School of Medicine, San Francisco, CA, USA.

[2]Howard Hughes Medical Institute, UCSF, San Francisco, CA, USA.

Abstract

Complex macromolecules and multimeric biological complexes are often characterized by intersubunit flexibility that precludes the precise global alignments needed for high-resolution reconstructions from single-particle cryo-EM images. Multi-body refinement is an approach that seeks to resolve this dilemma by separately alignment each locally ordered component of a complex structure. Treating these 'subparticles' independently can allow more accurate and more precise image registration, and systems exhibiting local (i.e. non-crystallographic) symmetry also expands effective dataset size. TRPM4 and PTCH1 are two membrane proteins which form homomultimers characterized by significant flexibility that severely limits local resolution. TRPM4 exists as a rigidly bound tetramer, but its peripheral MHR domains are highly dynamic, with broken symmetry. PTCH1 is a loosely-bound dimer, and the relative mobility of the protomers poses a challenge to conventional single-particle reconstruction. In both cases, the multi-body approach delivers dramatic local resolution improvements. Moreover, the application of a new method for kernel principal component analysis of rigid-body motions to multi-body refinement data permits modeling of protein motion at the single molecule level. The method accounts for the non-linearity of the **SE**(3) manifold using exact geodesic kernels for translation and rotation separately, or through a heuristic metric based on the concept of screw trajectories that enables low-dimensional embedding of arbitrary rigid-body pose observations. Multi-body analysis is finally used to animate molecular movies that reveal orthogonal modes of motion exhibited by TRPM4 and PTCH1 particles.

Principles of multi-body refinement

Single-particle cryo-EM relies on the remarkable phenomenon that many species of biological molecules are so uniform as to be precisely superimposable within Angstroms [1].

However, macromolecules with complex functions or comprising multiple domains, and multimeric complexes of such molecules, usually exhibit significant conformational heterogeneity that is also often critical to biological function. These globally flexible, but locally ordered, particles are frequently non-superimposable (**Figure 1.1A**), and resulting misalignment and symmetry breaking can severely limit the resolution attainable through 3D reconstruction. Multi-body refinement (MBR) attempts to resolve this difficulty by applying carefully focused 3D refinement to each independently mobile component of a flexible particle (**Figure 1.1B**) [2]. The so-called 'consensus refinement' treating the particle as rigid is used to define target bodies (subparticles) believed to have local order. The particle alignment parameters determined during the consensus refinement are then transformed to shift the center of rotation to each target subparticle. In the case of pseudosymmetric homomultimers, or other particles with local symmetry (similar to non-crystallographic symmetry), this operation results in an expansion of the number of particles. The transformed alignment parameters are used to initialize a local search for improved alignment parameters that deliver a higher resolution structure of the subparticle. The local refinement may be enhanced using a focused map, and potentially partial signal subtraction (**Figure 1C**), in order to utilize only the signal derived from a particular region. The key principles of MBR are thus 1) focused refinement with accurate initial parameters, 2) elimination of interdomain constraints on particle alignment, and 3) expansion of local symmetry increasing the effective number of particles.

An illustrative example can be found in the homotetrameric cation channel TRPM4, which regulates cardiac muscle synchronization [3, p. 4]. The cryo-EM structure of TRPM4 with imposed C4 symmetry (**Figure 1.1D**) has heterogeneous resolution, ranging from ~2.6 Å in the transmembrane domain (TMD) (blue box in **Figure 1.1D-E**), to worse than 8 Å in the distal

Melastatin Homology Region domains MHR1-2 (red box in **Figure 1.1D-E**). Interestingly, the MHR1-2 domains undergo significant conformational changes following Ca^{2+} binding, which is in turn required for channel densitization. The relatively low resolution and broken symmetry in MHR1-2 is likely related to functional dynamics that link the Ca^{2+}-free and Ca^{2+} bound states of the protein. MBR using the four MHR1-2 subparticles within each TRPM4 tetramer relieves the impossibility of alignment particles with four flexible components and increases the number of particles from ~84,000 to ~336,000. The result is a striking improvement in the resolution of MHR1-2 (**Figure 1.E**, right), which now approaches 3 Å. In contrast, applying MBR to the TMD results in no change in local resolution, suggesting the TMD is in fact nearly C4 symmetric in most TRPM4 particles (**n.s.**). The enhanced resolution of the MBR structure of MHR1-2 was essential in completing an atomic model of TRPM4.

A second example is given by the cryo-EM structure of a flexible dimer of PTCH1, a eukaryotic RND-like protein that is an essential regulator of the Hedgehog signaling pathway [4]. This pathway is implicated in the development of virtually all animals, and is essential for ubiquitous morphological features such as bilateral symmetry or appendages with differentiated digitation. Aberrant inhibition of the Hedgehog pathway during mammalian development leads to severe physical abnormalities such as cyclopia and holoprosencephaly, while aberrant activation in adults leads to basal cell carncinomas. In normal function, PTCH1 regulates Hedgehog signaling by enzymatically depleting cholesterol in the cytoplasmic leaflet of the plasma membrane, and enriching it in the extracellular leaflet. Cholesterol in the outer leaflet inactivates the pathway receptor Smoothened, until the morphogen Sonic hedgehog becomes prevalent in the extracellular milieu. Hedgehog then binds to PTCH1 and inhibits its cholesterol transport activity [5], [6], relieving the inhibition of Smoothened and activating the pathway's downstream effectors.

Much about the function of PTCH1 remains mysterious, including the purpose of PTCH1 dimerization. The relatively small interface in the distal extracellular domain (ECD) permits a great deal of flexibility, and allows displacements of the TMDs from a position implying a high degree of membrane curvature to a near-parallel configuration [7]. The flexibility hampers alignment and results in very low resolution in the TMD (**Figure 1.1A**, **Figure 1.1F**). MBR of PTCH1 monomers increases the effective number of particles from ~225,000 to ~450,000, and eliminates type of misalignment shown in **Figure 1.1A**. The resulting structure has dramatically improved resolution, which not only allowed complete modeling of PTCH1, but also revealed sterol-like densities bound at three distinct sites within the protein and thus confirmed that PTCH1 is active as a RND-like transporter most likely powered by a Na^+ antiport mechanism.

Multi-body phenomenology of flexible structures

MBR produces great improvements in resolution based on determination of new alignment parameters for the targeted regions of a particle. A key insight is that these changes in pose directly inform the state of individual particles, by relating the positions of their components to the overall object. The residuals between the consensus and subparticle alignment parameters capture these changes in particle state (**Figure 1.2A-B**). It is clear that careful analysis of residual alignments might yield great insight into macromolecular motions and thermodynamics. Statistical treatment of the residuals requires comparing the rigid-body motions they represent. These motions are known to belong to the Lie group **SE**(3). The group is a composition of subgroups **SO**(3), comprising 3D rotation, and $\mathbb{R}^3$ comprising simple 3D translations (1).

$$SE(3) = SO(3) + \mathbb{R}^3 \quad (1)$$

Both **SO**(3) and **SE**(3) are non-linear manifolds. Taking a maze as an ersatz for a manifold, it is clear that comparing points from the manifold may not be straightforward (**Figure 1.2C**). A

consequence is that previous approaches to statistical analysis of MBR data using principal component analysis (PCA) of residual alignments expressed using Euler angles [8] are only valid for very small movements over which the manifold is approximately flat [9]. The simplest approach valid for PCA in non-linear subspaces such as **SE**(3) is kernel PCA (kPCA), which replaces the implied Euclidean distance in standard PCA with any kernel function related to a valid inner product space [10]. These geometrically correct kernels represent geodesics, the shortest distance between points on the manifold. What kernels should be adopted for PCA of MBR data?

As discussed, MBR data lie on **SE**(3). Unfortunately, no analytic geodesics exist for SE(3), owing to intrinsic differences of the **SO**(3) and $\mathbb{R}^3$ subgroups [11]. In other words, rotation and translation are incommensurable. The subgroups themselves, however, both admit simple geodesics (**Figure 1.2D**), and we can begin to investigate motions within biological particles considering translation and rotation separately. The shortest distance between two points *i* and *j* in $\mathbb{R}^3$ is the familiar Euclidean distance, which is simply the length of the straight line between the points (4). Residual translations can thus be subjected directly to standard PCA.

$$d_{ij} = \sqrt{\sum_k^3 \left(x_{ik} - x_{jk}\right)^2} \tag{2}$$

In contrast, the geodesic on **SO**(3) is the total *angle* of the shortest, smooth rotation separating two points on the manifold [12]. If these points (rotations) are represented as standard rotation matrices $\mathbf{R}_i$ and $\mathbf{R}_j$, this separating angle ω_{ij} is given by (5). Similar to other Lie groups, **SO**(3) represents the exponentiation of its tangent Lie algebra $\mathfrak{so}(3)$, and this fundamental non-linearity is reflected by the matrix logarithm in the expression.

$$\omega_{ij} = \left\| log_M \boldsymbol{R}_i \boldsymbol{R}_j^T \right\|_2 \tag{3}$$

The appropriate kernel for **SO**(3) is generated by converting distance measurements between all pairs of observations into inner products using a straightforward algebraic relationship. As noted above, kPCA depends on creating a Gram matrix containing the inner products for all pairs of observations. The Gram matrix elements g_{ij} are calculated directly from the pairwise distance matrix elements ω_{ij} following (4); this operation is known as "double centering." The eigenvalue decomposition (5)-(6) of the Gram matrix then yields non-linear PCs. Note that the kPCA approach avoids any kind of *ad-hoc* parameter scaling, and simply represents rigid-body rotations as they are.

$$g_{ij} = -\frac{1}{2}\left(\omega_{ij}^2 - \frac{1}{n}\sum_i^n \omega_{ij}^2 - \frac{1}{n}\sum_j^n \omega_{ij}^2 + \frac{1}{n^2}\sum_i^n\sum_j^n \omega_{ij}^2\right) \quad (4)$$

$$\boldsymbol{G\nu} = \boldsymbol{\lambda\nu} \quad (5)$$

$$X_k = \sqrt{\lambda_k}\nu_k \; \forall \; k \in \; \|\boldsymbol{\nu}\| \quad (6)$$

We now apply kPCA to MBR of TRPM4 MHR1-2. The Gram matrix for *N* particles is a *N*x*N* matrix, and there are hundreds of thousands of MHR1-2 subparticles. A sample kPCA approach is therefore required in order to the fit the Gram matrix in computer memory. Results using 10,000 randomly sampled MHR1-2 subparticles for kPCA are shown in **Figure 1.3A-B**. The remaining MHR1-2 residual rotations can be projected into the subspace of the sample using out-of-sample kPCA [13]. A scatter plot of out-of-sample kPCA scores for the entire MHR1-2 dataset is shown in **Figure 1.3C**. These results fundamentally represent particle motions, which are best interpreted through explicit visualization in molecular movies. However, in kPCA the Gram matrix and PCA scores cannot be used to directly sample orientation observations. An alternative is to cluster subparticles in the kPCA latent space, and use the orientations corresponding to the cluster centroids as 'key-frame' or anchor structures for constructing an animated trajectory. This approach is illustrated in **Figure 1.3D-E**. The particles are clustered by their score on a single PC,

while only considering particles that fall within given percentile bounds for PC loading and distance from the PC axis. This allowed cylindrical region forces the clusters to have constant diameter, regardless of their location within the ellipsoidal distribution of latent coordinates. Without a constraint such as this, clusters become very wide at the tails of the latent distribution, which results in misleading influence of orthogonal PCs (**Figure 1.3F**). For more details, see **Methodology**.

Animations of TRPM4 MHR1-2 in the tetramer according to its first three PCs are shown in **Movie 1.1** and **Movie 1.2**, and animations of the PTCH1 dimer are shown in **Movie 1.3**. The power of the geometrically correct PCA kernel is reflected in the clean separation of orthogonal modes of motion within the particles.

Despite the apparent success of kPCA taking into account rotation only, there are certainly many flexible particles with significant translational components. Is there a natural way to include translation information via an appropriate kernel function? As noted earlier, there is no true geodesic in **SE**(3). However, we may be able to identify approximate or heuristic metrics that capture most of the properties of an exact geodesic. For example, all rigid-body motions may be represented as *screws*, which consist of a translation along an axis and a rotation about that axis (**Figure 1.3G**). The path swept out by an object subjected to a screw is a helical trajectory with length dependent on both the rotational and translation component of the rigid-body motion [14]. This screw distance d_{sc} is an intuitive (heuristic) metric on SE(3); slightly different motions will produce a small d_{sc}, while very different motions will lead to large d_{sc}. The screw distance has been proposed as an approximate SE(3) metric for motion planning and smooth blending of motion trajectories [15], but has not previously been used as a kernel for PCA of object pose data. A scatter plot showing the screw kPCA approach applied to the PTCH1 MBR data is shown in **Figure 1.3H**.

The plot reveals significant internal structure and potential clusters that are not evident in kPCA of orientations alone. Further research is required to build on this promising early result by numerically verifying the geodesic-like properties of the screw distance, and using screw kPCA to produce animated motion trajectories for structural interpretation.

Methodology

Quaternion toolbox for MBR

While a Gram matrix computed using the distance formula in (4) permits geometrically correct PCA of 3D rotation data, the matrix logarithm requires a prohibitively expensive power series expansion (using matrix exponentiation). Orientation distance calculations can be greatly accelerated by taking advantage of the isomorphism between **SO**(3) and **su**(2). Elements of **su**(2) are four-vectors, called *quaternions*. Quaternions are conceptually similar to rotation matrices, and act in just the same way provided the conventions defined in (7)-(9).

$$q = w + ix + jy + kz = w + \vec{q} \tag{7}$$

$$i^2 = j^2 = k^2 = ijk = -1 \tag{8}$$

$$q_1 q_2 = w_1 w_2 - \vec{q}_1 \cdot \vec{q}_2 + w_1 \vec{q}_2 + w_2 \vec{q}_1 + \vec{q}_1 \times \vec{q}_2 \tag{9}$$

Distances between orientations represented as quaternions, or as rotation matrices, have the simple relationship given by (10).

$$\omega_{ij} = 2\,cos^{-1}\left|q_i \cdot q_j\right| = \left\| log_M \boldsymbol{R}_i \boldsymbol{R}_j^T \right\|_2 \tag{10}$$

Cryo-EM images are projections, therefore the orientation distance must be clamped to account for an "opposite" view delivering an identical image (11). Symmetry can be dealt with by applying the symmetry operators to one of the observations and always taking the minimum distance.

$$\omega' = \begin{cases} \pi - \omega, & \omega > \frac{\pi}{2} \\ \omega, & \omega \leq \frac{\pi}{2} \end{cases} \tag{11}$$

Dot-products between four-vectors and the inverse cosine are computationally inexpensive operations and yield dramatic performance improvements for kPCA. For typical multi-body refinement data comprising hundreds of thousands of observations, the wall time for computing a pairwise distance matrix is improved from hours or days to just seconds.

Beyond performant orientation distances, quaternions have a number of other advantages that are indispensable for analyzing orientation data. Mean rotation, for example, is needed for correct orientation PCA [9], for relating whole orientation datasets, and for averaging of noisy single-particle observations. While a number of approximate methods using rotation matrices have been proposed, quaternions provide the *only* exact method for calculating average rotation [16]. First, quaternions are converted to an equivalent form called the *attitude matrix* as in (12).

$$\boldsymbol{A} = \boldsymbol{Q}^T \boldsymbol{Q} \tag{12}$$

The attitude matrices for a set of N quaternions are then summed directly per (13)

$$A_{jk} = \sum_{i}^{N} \sum_{j}^{4} \sum_{k}^{4} q_{ij} q_{ik} \tag{13}$$

The largest eigenvector of the summed attitude matrix $\boldsymbol{A}$ is itself a quaternion, which exactly represents the mean orientation. Solving the eigenvalue problem (14) and sorting the eigenvalues (15) for a single 4x4 matrix are simple and computationally inexpensive.

$$\boldsymbol{\lambda X} = \boldsymbol{AX} \tag{14}$$

$$\hat{q} \equiv x_{\underset{i}{\operatorname{argmax}} \lambda_i} \tag{15}$$

Indeed, the attitude approach for exact average rotation is significantly simpler and more efficient than the proposed approximations. In UCSF pyem [17], the algorithm is implemented by

a single line of code in `pyem.quat.meanq()` as part of a complete library for creating and manipulating quaternions.

Spherical linear interpolation

Animations that use only a small number of anchor structures are difficult to interpret because the relatively large motions between anchors lead to a "jerky" or "choppy" appearance. Increasing the number of clusters can ameliorate this effect, but different animations with different numbers of total frames are difficult to synchronize and compare. Ideal trajectories for animation would have a large, arbitrary number of steps of consistent size. For trajectories in Euclidean space, this result could be achieved using linear interpolation to add additional steps between the anchors. There is an intuitive connection between geodesic distances and interpolation, and just as linear interpolation between positions in Euclidean space follow a straight line, interpolation between points on a manifold should also follow geodesics of that manifold. Animation trajectories can thus be augmented by through additional points that lie on the shortest arcs of rotation between subsequent anchor structures. This method is called Spherical Linear Interpolation (SLERP), and can be efficiently implemented using quaternion exponentiation as in (16). SLERP can also be used to calculate the "square-root" of quaternion q_j, representing a rotation in the same direction but of half magnitude, by setting q_i to the identity and t to one-half.

$$\mathrm{SLERP}(q_i, q_j, t) = q_i(q_i^{-1}q_j)^t,\ t \in [0,1] \tag{16}$$

We can now create smooth animations with an arbitrary number of frames using SLERP to interpolate additional steps between anchors such as shown in **Figure 3.3E**. This technique is used in **Movies 1.1-1.3** via the implementation in UCSF pyem [17].

Dual quaternions for screw distance

The screw distance is an intuitive heuristic metric in **SE**(3). Efficient screw distance calculation requires the introduction of the dual quaternion, a new mathematical construct for representing rigid-body motions [18]. Much as quaternions are inspired by complex numbers, dual quaternions are drawn from the field of dual numbers. Much like complex numbers, dual numbers have a real part, and dual part which multiplies the dual unit ε. Like the imaginary unit i, the dual unit has special composition rules given by (17)-(18).

$$\varepsilon \neq 0 \tag{17}$$

$$\varepsilon^2 = 0 \tag{18}$$

These rules state that ε is not zero, but the square of ε is zero. Although this property is counterintuitive, it causes ε to behave very similarly to i, except that where multiplication by i^2 introduces a factor of -1, ε^2 introduces a factor of 0 that cancels a term. A dual quaternion is defined as the composition of a regular (or real) quaternion q_r, representing rotation, and a dual quaternion q_d representing translation.

$$\hat{q} = q_r + \varepsilon \cdot q_d \tag{19}$$

The dual quaternion is constructed directly from a translation vector per (20).

$$q_d = 0 + i\frac{x}{2} + j\frac{y}{2} + k\frac{z}{2} \tag{20}$$

Remarkably, if dual quaternions are multiplied using the same formula from (9) used for standard quaternions, they correctly assess arbitrary rigid-body transformations. Furthermore, dual quaternions have a direct interpretation in terms of screws, with readily obtainable Plücker coordinates (21)-(25) that can be used to efficiently compute the screw distance d_{sc} per (26). In (21)-(24) w and v refer to the scalar and vector parts of a quaternion, and subscript r and d indicate the real or dual quaternion.

$$\theta = 2\cos^{-1} w_r \tag{21}$$

$$\vec{s} = \frac{\vec{v}_r}{\|\vec{v}_r\|} \tag{22}$$

$$d = -2\frac{w_d}{\|\vec{v}_r\|} \tag{23}$$

$$\vec{m} = \left(\vec{v}_d - \vec{s}\frac{dw_r}{2}\right) \cdot \frac{1}{\|\vec{v}_r\|} \tag{24}$$

$$\vec{a} = \vec{s} \times \vec{m} \tag{25}$$

$$d_{sc} = \sqrt{d^2 + \theta^2 \|a\|^2} \tag{26}$$

UCSF pyem [17] includes a full set of library functions for creating and manipulating dual quaternions, and performant calculation of screw distances.

Figures

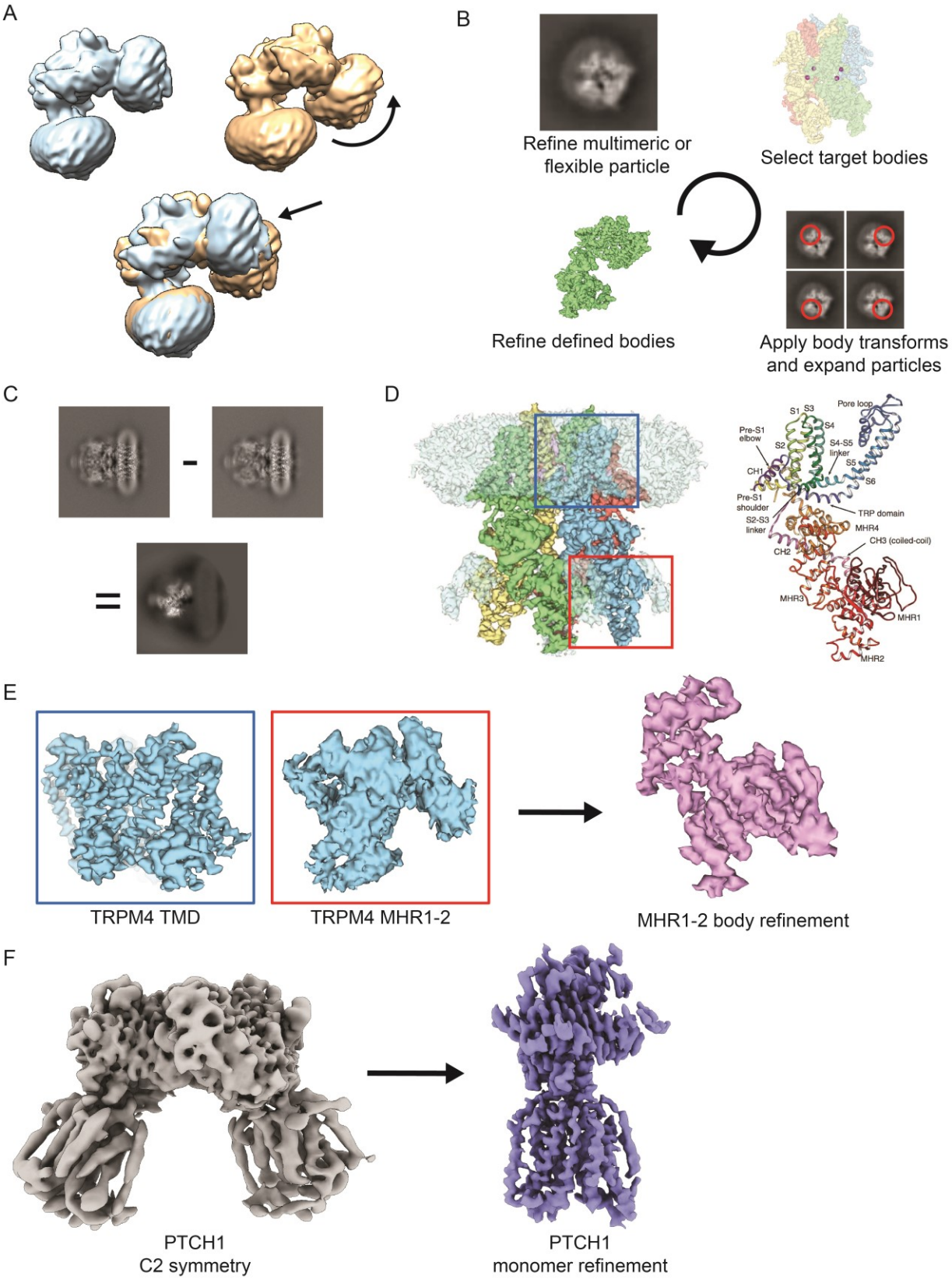

A
B
Refine multimeric or flexible particle
Select target bodies
Refine defined bodies
Apply body transforms and expand particles
C
=
D
S1
S3
Pore loop
Pre-S1 elbow
S2
S4
S4-S5 linker
S5
CH1
S6
Pre-S1 shoulder
TRP domain
S2-S3 linker
MHR4
CH3 (coiled-coil)
CH2
MHR3
MHR1
MHR2
E
TRPM4 TMD
TRPM4 MHR1-2
MHR1-2 body refinement
F
PTCH1
C2 symmetry
PTCH1
monomer refinement

Figure 1.1 Principles of multi-body refinement
(A) Alignment of PTCH1 dimer cryo-EM envelopes in different conformations demonstrate that flexible, but locally ordered, structures may not be superimposible via any single transformation. The relative motion of one PTCH1 monomer is indicated in the top right panel by a curved arrow. The overlay of the two structures in the bottom panel reveals the necessary mis-alignment of one monomer of or the other, indicated by a small, straight arrow.
(B) The basic cycle of multi-body refinement. The structure of a flexible multi-domain or multimeric particle (top left) is used to identify target bodies, such as individual protomers (top right). The particle images are duplicated according to the number of target bodies, and alignment parameters for the expanded particles are transformed to represent the center of the targets (bottom right). The new alignment parameters can be used for local refinement of each target body alone, as for an isolated, rigid particle (bottom left).
(C) Refinement of a single target body may be improved if simulated (reprojected) images of the rest of the particle are subtracted from the raw data. If a projection corresponding to all but one domain (top right) is subtracted away from the complete particle (top left), the resulting image has signal only for the region of interest (bottom).
(D) The cryo-EM structure of the tetrameric ion channel TRPM4 with applied C4 symmetry (left), and the ribbon diagram for a single TRPM4 monomer (right). One transmembrane domain is highlighted by a blue box, and one MHR1-2 region is indicated by a red box.
(E) Close-ups of boxed areas in (D) show the resolution of the transmembrane domain (left, blue box) is significantly higher than that of MHR1-2 (right, red box). However, multi-body refinement of MHR1-2 (far right) yields a dramatic improvement in resolution, comparable to the transmembrane domain in the original structure. The improved resolution was critical for building the complete atomic model of TRPM4.
(F) Local resolution in the cryo-EM structure of the PTCH1 dimer progressively declines with distance from the C2 symmetry axis (left). As for TRPM4 in (E), multi-body refinement of the PTCH1 monomer (right) eliminates the effects of pseudo-symmetry and dramatically improves the resolution. These improvements were critical to modeling PTCH1, and in identifying sterol transport substrates bound within the protein.

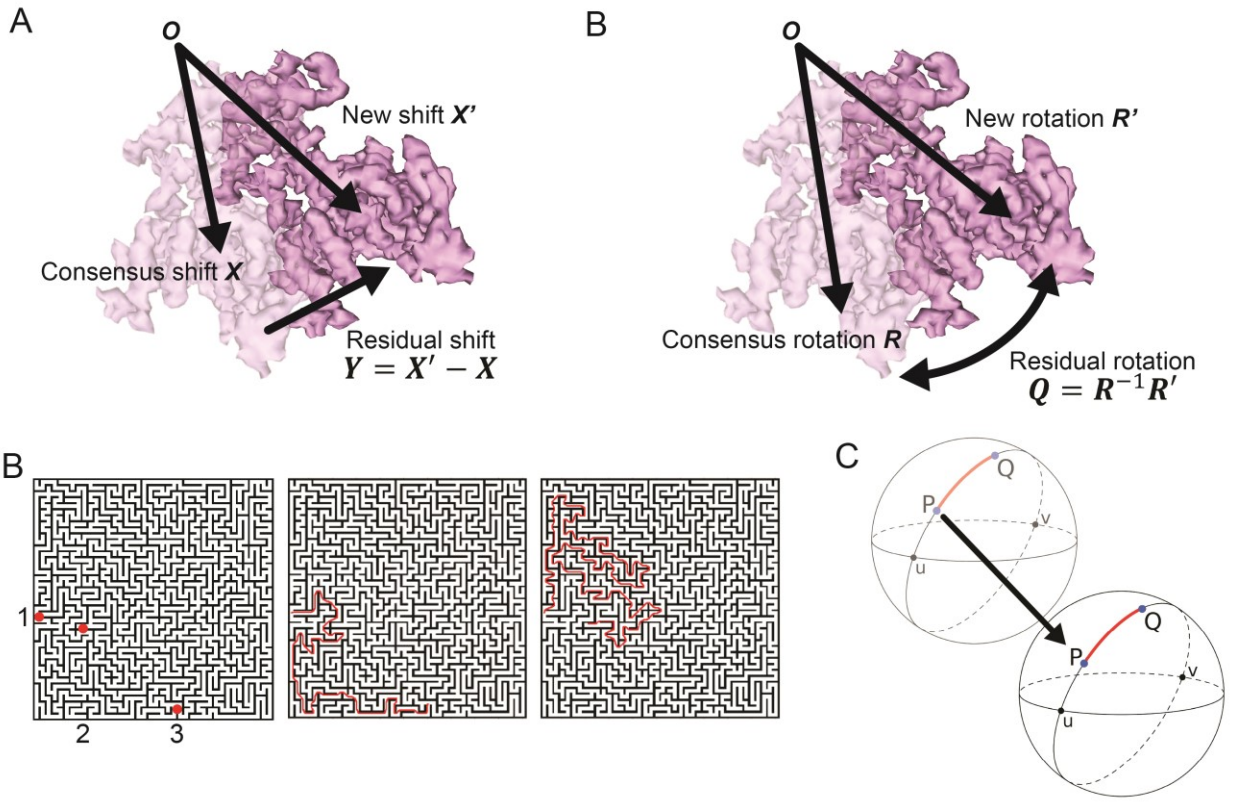


Figure 1.2 Residual alignments lie on the non-linear manifold SE(3)

(A) The localized alignment parameters determined through multi-body refinement can be used to extract the residual shifts and rotations representing the state of each particle, in terms of the relative positions of the target bodies.

(B) An illustration of the difficulty in comparing observations drawn from a non-linear manifold such as **SE**(3). Such a manifold might be represented as the hallways of a maze (left). Within the maze, the shortest distances between points are no longer obvious. For example, while points 1 and 2 are closer than points 1 and 3 in terms of straight-line distance, the shortest allowed path from point 1 to 3 (center) is actually much shorter than that between points 1 and 2 (right).

(E) Rigid-body motions lie on the manifold **SE**(3), which is the composition of 3D translation and 3D rotation. The shortest distance between translation vectors is simply the magnitude of their difference vector (straight arrow). In contrast, the geodesic between rotations is the angle subtended by the shortest spherical arc between them (red arc).

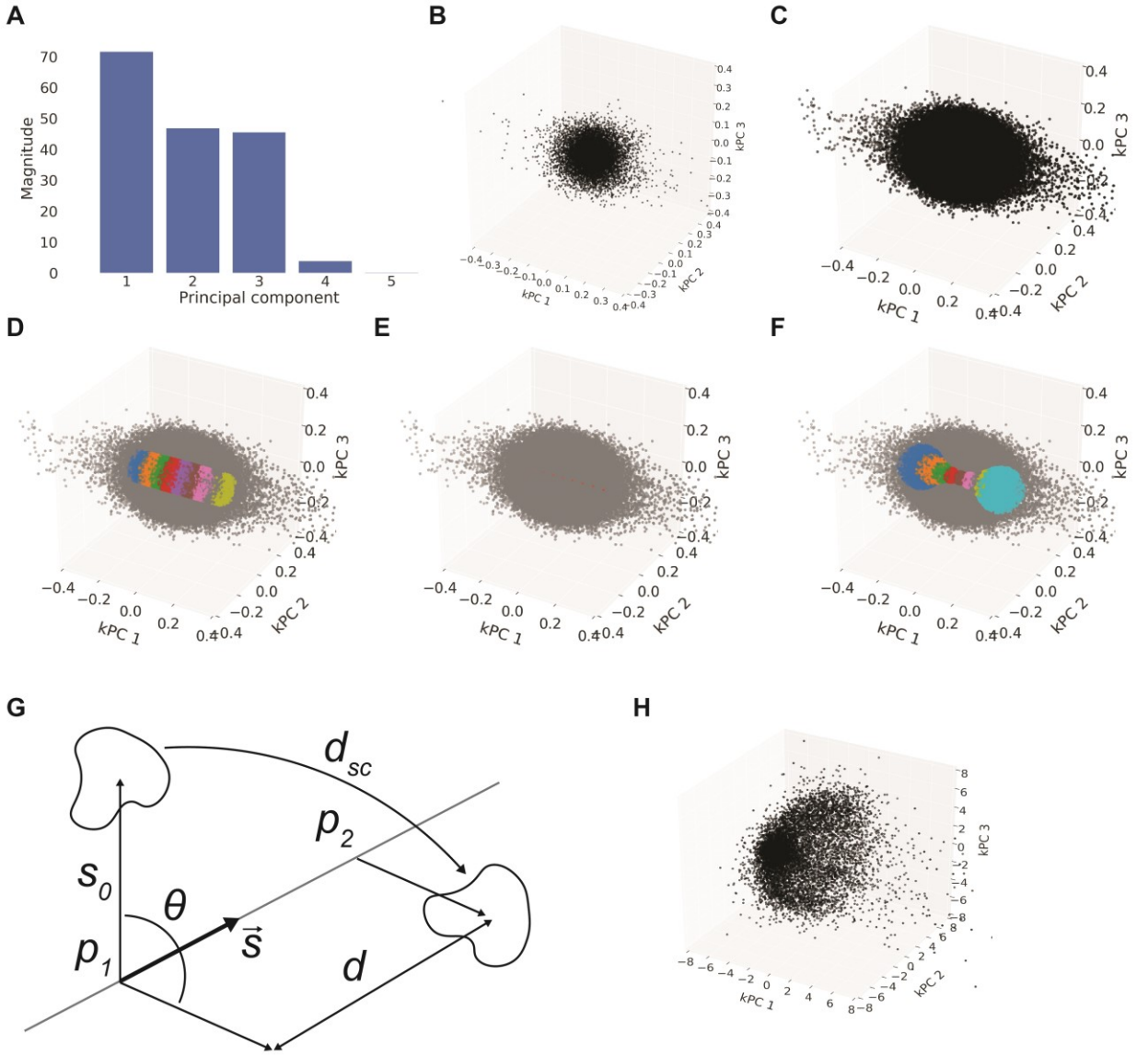


Figure 1.3 Kernel PCA of TRPM4 MHR1-2 multi-body refinement data

(A) Magnitude of eigenvalues from kPCA of multi-body refinement data for 10,000 randomly sampled TRMP4 MHR1-2 subparticles. As expected, there are no more than four non-zero eigenvalues.

(B) Scatter plot of scores for the 10,000 subparticle sample in the top three PCs. Each point represents a single MHR1-2 domain, and the position of the point in the plot corresponds to some particular position of the domain with respect to the whole TRPM4 particle.

(C) Scatter plot based on out-of-sample kPCA for the complete TRPM4 MHR1-2 data with ~336,000 particles.

(D) Particles may be clustered according to their position along each PC axis. The clusters should be constrained to a cylinder centered on the PC axis, in order to reduce the impact of noisy outliers, and to aid interpretation by limiting animations to the effects of that PC alone.

(E) Cluster centroids (red stars) for the clusters in (D). The centroids represent anchor structures for animation, which are determined by averaging the orientations within each cluster.

(F) If clustering is performed along one PC axis without outlier rejection as in (D), the decreasing density of observations away from the origin leads to increased dispersion along the other PCs. If these clusters are used for animation, the beginning and end of the animation will reflect multiple PCs, which can be confusing and hinder interpretation.
(G) Any rigid-body motion can be represented as a screw, which is a translation along an axis, and a rotation about that axis. The length of the corresponding helical trajectory d_{sc} is an intuitive heuristic metric for $\mathbf{SE}(3)$.
(H) PTCH1 dimer kPCA plot using the screw distance kernel to account for rotation and translation simultaneously. Interestingly, the screw kPCA scatter plot seems to have more internal structure than quaternion kPCA, which does not account for translation.

Supplemental Files

Movie 1.1 Overview of TRPM4 structure and first component of MHR1-2 motion
The complete structure of the TRPM4 tetramer is shown and rotated, before zooming in for a close-up of the MHR1-2 region from a single monomer. The consensus refinement is faded out and replaced by the dramatically improved MHR1-2 multi-body refinement, and MHR1-2 is animated according to a trajectory along the first PC from kPCA of the MHR1-2 subparticle refinement data.

Movie 1.2 kPCA animations of TRPM4 MHR1-2
Montage of animations for the first three PCs from kPCA of the TRMP4 MHR1-2 subparticle refinement data. The first PC is on the left, second PC at center, and third PC on the right. Note the smoothness of the animations and that each PC captures a unique mode of motion for the MHR1-2 region.

Movie 1.3 kPCA animation of PTCH1 dimer
Montage of kPCA animations derived from multi-body refinement of PTCH1. The true orthogonality of the components identified using by the geodesic quaternion kernel is clear, with PC1 evincing outwards motions that separate two the monomers, PC2 representing symmetric rotations on the transverse axis, and PC3 representing antisymmetric rotations around that same transverse axis.

Acknowledgements

I would like to thank Drs. Eugene J. Palovcak and Evan M. Green our many fruitful discussions about advanced cryo-EM processing techniques, which were of great assistance during development of my approach to multi-body refinement. I would also like to thank Drs. David P. Bulkley and Yunxiao Zhang for introducing me to Hedgehog pathway structural biology through our work on PTCH1. Finally, Drs. Henriette E. Autzen and Alexsander Myasnikov were extremely supportive of my starting efforts towards symmetry expansion and multi-body refinement, and brought me my first significant cryo-EM processing project in the form of the ion channel TRPM4.

References

[1] F. J. Sigworth, "Principles of cryo-EM single-particle image processing," *Microscopy*, vol. 65, no. 1, pp. 57–67, Feb. 2016, doi: 10.1093/jmicro/dfv370.

[2] S. H. W. Scheres, "Processing of Structurally Heterogeneous Cryo-EM Data in RELION," B.-M. in Enzymology, Ed. Academic Press. Accessed: Jul. 28, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0076687916300301

[3] H. E. Autzen, A. G. Myasnikov, M. G. Campbell, D. Asarnow, D. Julius, and Y. Cheng, "Structure of the human TRPM4 ion channel in a lipid nanodisc," *Science*, vol. 359, no. 6372, pp. 228–232, Jan. 2018, doi: 10.1126/science.aar4510.

[4] Y. Zhang *et al.*, "Structural Basis for Cholesterol Transport-like Activity of the Hedgehog Receptor Patched," *Cell*, vol. 175, no. 5, pp. 1352-1364.e14, Nov. 2018, doi: 10.1016/j.cell.2018.10.026.

[5] X. Gong *et al.*, "Structural basis for the recognition of Sonic Hedgehog by human Patched1," *Science*, p. eaas8935, Jun. 2018, doi: 10.1126/science.aas8935.

[6] X. Qi, P. Schmiege, E. Coutavas, J. Wang, and X. Li, "Structures of human Patched and its complex with native palmitoylated sonic hedgehog," *Nature*, p. 1, Jul. 2018, doi: 10.1038/s41586-018-0308-7.

[7] X. Qi, P. Schmiege, E. Coutavas, and X. Li, "Two Patched molecules engage distinct sites on Hedgehog yielding a signaling-competent complex," *Science*, p. eaas8843, Aug. 2018, doi: 10.1126/science.aas8843.

[8] T. Nakane, D. Kimanius, E. Lindahl, and S. H. W. Scheres, "Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION," *bioRxiv*, p. 286856, Mar. 2018, doi: 10.1101/286856.

[9] M. P. Johnson, “Exploiting quaternions to support expressive interactive character motion,” Thesis, Massachusetts Institute of Technology, 2003. Accessed: Apr. 22, 2018. [Online]. Available: http://dspace.mit.edu/handle/1721.1/8031

[10] B. Schölkopf, A. Smola, and K.-R. Müller, “Nonlinear Component Analysis as a Kernel Eigenvalue Problem,” *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998, doi: 10.1162/089976698300017467.

[11] F. C. Park, “Distance Metrics on the Rigid-Body Motions with Applications to Mechanism Design,” *J. Mech. Des.*, vol. 117, no. 1, pp. 48–54, Mar. 1995, doi: 10.1115/1.2826116.

[12] D. Q. Huynh, “Metrics for 3D Rotations: Comparison and Analysis,” *J. Math. Imaging Vis.*, vol. 35, no. 2, pp. 155–164, Oct. 2009, doi: 10.1007/s10851-009-0161-2.

[13] B. Scholkopf, C. Burges, and A. Smola, “Kernel Principal Component Analysis,” in *Advances in Kernel Methods–Support Vector Learning*, MIT Press, Cambridge, MA, 1999.

[14] M. Žefran, V. Kumar, and C. Croke, “Metrics and Connections for Rigid-Body Kinematics,” *Int. J. Robot. Res.*, vol. 18, no. 2, pp. 242–1, Feb. 1999, doi: 10.1177/027836499901800208.

[15] A. Torsello, E. Rodolà, and A. Albarelli, “Multiview registration via graph diffusion of dual quaternions,” in *CVPR 2011*, Jun. 2011, pp. 2441–2448. doi: 10.1109/CVPR.2011.5995565.

[16] F. L. Markley, “Averaging Quaternions.”

[17] D. Asarnow, E. Palovcak, and Y. Cheng, *asarnow/pyem: UCSF pyem v0.5*. Zenodo, 2019. doi: 10.5281/ZENODO.3576630.

[18] B. Kenwright, “A beginners guide to dual-quaternions: What they are, how they work, and how to use them for 3D character hierarchies,” in *20th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, 2012, pp. 1–10. Accessed: Sep. 02, 2021. [Online]. Available:

https://researchportal.hw.ac.uk/en/publications/a-beginners-guide-to-dual-quaternions-what-they-are-how-they-work

Chapter 2

Structural insight into SARS-CoV-2 neutralizing antibodies and modulation of syncytia

Contributing Authors

Daniel Asarnow[1,3,13], Bei Wang[2,13], Wen-Hsin Lee[2,14], Yuanyu Hu[2,14], Ching-Wen Huang[2,14], Bryan Faust[1,3,14], Patricia Miang Lon Ng[2,14], Eve Zi Xian Ngoh[2,14], Markus Bohn[3,4,14], David Bulkley[1,3,14], Andrés Pizzorno[5,14], Beatrice Ary[3,4], Hwee Ching Tan[2], Chia Yin Lee[2], Rabiatul Adawiyah Minhat[2], Olivier Terrier[5], Mun Kuen Soh[2], Frannie Jiuyi Teo[2], Yvonne Yee Chin Yeap[2], Shirley Gek Kheng Seah[6], Conrad En Zuo Chan[6], Emily Connelly[4], Nicholas J. Young[4], Sebastian Maurer-Stroh[7,8,9], Laurent Renia[2,8], Brendon John Hanson[6], Manuel Rosa-Calatrava[5,10], Aashish Manglik[3,4,11,*], Yifan Cheng[1,3,12,*], Charles S. Craik[3,4,*], Cheng-I Wang[2,15,*]

[1]Department of Biochemistry and Biophysics, University of California San Francisco (UCSF) School of Medicine, San Francisco, CA, USA.

[2]Singapore Immunology Network, Agency for Science, Technology and Research (A*STAR), 8A Biomedical Grove, Immunos, Singapore 138648, Singapore.

[3]QBI COVID-19 Research Group (QCRG), San Francisco, CA, USA

[4]Department of Pharmaceutical Chemistry, University of California San Francisco (UCSF), San Francisco, CA, USA.

[5]CIRI, Centre International de Recherche en Infectiologie, (Team VirPath), Université de Lyon, Inserm U1111, Université Claude Bernard Lyon 1, CNRS, UMR5308, ENS de Lyon, 69007 Lyon, France.

[6]Biological Defence Program, DSO National Laboratories, 27 Medical Drive, Singapore 117510, Singapore.

[7]Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, Matrix, Singapore 138671, Singapore.

[8]Infectious Diseases Laboratories (ID Labs), Agency for Science, Technology and Research (A*STAR), 8A Biomedical Grove, Immunos, Singapore 138648, Singapore

[9]Department of Biological Sciences, Faculty of Science, National University of Singapore, Singapore 117558, Singapore.

[10]VirNext, Faculté de Médecine RTH Laennec, Université Claude Bernard Lyon 1, Université de Lyon, Lyon, France.

[11]Department of Anesthesia and Perioperative Care, UCSF, San Francisco, CA, USA

[12]Howard Hughes Medical Institute, UCSF, San Francisco, CA, USA.

[13]These authors contributed equally to this work.

[14] These authors contributed equally to this work.

Abstract

Infection by SARS-CoV-2 is initiated by binding of viral Spike protein to host receptor angiotensin-converting enzyme 2 (ACE2), followed by fusion of viral and host membranes. While antibodies that block this interaction are in emergency use as early COVID-19 therapies, precise determinants of neutralization potency remain unknown. We discovered a series of antibodies that all potently block ACE2 binding, yet exhibit divergent neutralization efficacy against live virus. Strikingly, these neutralizing antibodies can either inhibit or enhance Spike-mediated membrane fusion and formation of syncytia, which are associated with chronic tissue damage in COVID-19 patients. Multiple cryogenic electron microscopy structures of Spike-antibody complexes reveal distinct binding modes that not only block ACE2 binding, but also alter the Spike protein conformational cycle triggered by ACE2 binding. We show that stabilization of different Spike conformations leads to modulation of Spike-mediated membrane fusion, with profound implications in COVID-19 pathology and immunity.

Introduction

The first step of viral infection by coronaviruses such as SARS-CoV and SARS-CoV-2 is the binding of a Spike protein on the virion to a specific receptor in the membrane of a host cell [1]. The virus enters cells by fusion of the viral envelope with cellular plasma membranes and alternatively by endocytosis and subsequent fusion of the viral envelope with endosomal membranes. The SARS-CoV-2 Spike protein, similar to that of other coronaviruses, comprises two subunits, S1 and S2, and is responsible for target recognition and mediating viral entry [1]. Upon binding to the host cell receptor through the receptor binding domain (RBD) at the tip of the S1 subunit, the Spike protein undergoes dramatic conformational changes and proteolytic processing. Further shedding of the S1 subunit exposes the S2 subunit fusion peptide, which inserts

into the host cell membrane and induces viral fusion [2]. Both SARS-CoV-2 and SARS-CoV use the angiotensin converting enzyme 2 (ACE2) as the entry receptor to infect host cells [3]–[5]. The RBD binds to ACE2 via the receptor binding motif (RBM), a small patch made up of about 20 amino acids [6, p. 2], [7].

In addition to mediating viral entry, excess Spike protein in the membranes of coronavirus-infected cells drives neighboring cells expressing ACE2 to fuse and form syncytia (multinucleated giant cells) through a pH-independent mechanism [8], [9]. Syncytia are associated with lung tissue damage in SARS-CoV and MERS-CoV infections, and have been widely observed in autopsies of patients afflicted with severe COVID-19 [10], [11]. Preliminary reports also implicate syncytia in chronic cardiovascular injury due to COVID-19 [12].

A common therapeutic strategy against COVID-19 and other coronavirus-related illnesses is blocking of the RBD/ACE2 interaction by antibodies, nanobodies, or soluble ACE2 as a decoy [13]. Since the first report of COVID-19, many SARS-CoV-2 neutralizing monoclonal antibodies (NAbs) have been discovered from immunized animals [14]–[16] and COVID-19 convalescent patients [14], [17]–[25]. Most SARS-CoV-2 NAbs recognize epitopes within the ACE2 binding site, imposing direct competition between the virus-ACE2 interaction [14], [17], [18], [23], [23], [25], while others bind outside the RBM [14], [15], [26] or even the RBD [17], [19]. A number of NAbs in late clinical trials have already been used as experimental treatments for COVID-19 [14], with promising efficacy in patients with early-stage infections. Competitive inhibition of receptor binding, denaturation of native Spike conformations [27], and pre-fusion trapping [15] may all contribute to neutralization, but the molecular mechanisms governing neutralization potency of NAbs remain poorly understood, and it is unknown whether receptor-blocking NAbs can also influence Spike mediated cell-cell fusion and syncytia formation.

Neutralizing antibodies against pathogens can also be isolated by *in vitro* selection from highly diverse combinatorial human libraries [28]–[30], and may be distinct from those found in natural infection or by immunization [31]. We screened a naïve combinatorial human Fab library for antibodies that target the SARS-CoV-2 Spike protein RBD and competitively block ACE2 binding. We discovered a series of antibodies that all exhibit effective receptor blockade, but have strikingly different neutralization potencies against SARS-CoV-2. Taking these antibodies as mechanistic probes, we show that bivalent binding and receptor blockade are not sole determinants of potent neutralization. In addition to blocking ACE2, these antibodies either inhibit or enhance syncytia formation in Vero E6 cells, suggesting that potentiation of cell-cell fusion by antibodies may compromise the effectiveness of viral neutralization in treatment of severe COVID-19. One potently neutralizing and potentially therapeutic antibody, designated 5A6, uniquely inhibits cell-cell fusion and syncytia formation and blocks receptor binding. High resolution cryogenic electron microscopy (cryo-EM) structures of multiple Spike-antibody complexes provide insight into determinants of viral neutralization potency, and reveal that 5A6 recognizes a cryptic, quaternary epitope that conveys receptor blockade and inhibits syncytia by trapping the pre-fusion state.

Results

Isolation of SARS-CoV-2 receptor-blocking antibodies from a naïve human library

We identified six antibodies that block the RBD/ACE2 interaction with nanomolar EC_{50} values by phage display of a naïve combinatorial human Fab library comprising 3 x 10^{10} random heavy and light IgG chain pairs drawn from 22 healthy donors [32] (**Figure 2.1A**, **Figure 2.7A**). Their germlines and degree of hypermutation are given in **Table 3.1**. Whereas the source Fabs had moderate intrinsic affinities for immobilized RBD, as high as 1.6 nM for 3D11 and 7.6 nM for 5A6, clones reformatted as IgGs showed sub-picomolar to picomolar binding avidity (**Figure**

2.1B, **Figure 2.7B-D**). Biolayer interferometry (BLI) of free Fabs and their equivalent IgGs indicates that the greatly enhanced binding of the IgGs is primarily due to slower dissociation, likely due to bivalent binding of both Fab arms to their epitopes (**Figure 2.1C**, **Table 3.2**). Stepwise binding BLI assays show that Fab fragments 5A6 and 3D11 have non-overlapping footprints on the RBD, while 5A6 shares at least partially overlapping epitopes with the other four antibodies (**Figure 2.1D**).

SARS-CoV-2 neutralization by receptor blocking antibodies

We evaluated the six receptor-blocking antibodies for neutralizing activity against SARS-CoV-2 pseudovirus in CHO-ACE2 cells with a luciferase reporter (**Figure 2.2A**), and against live SARS-CoV-2 [33] in Vero E6 cells by cell viability (**Figure 2.2B**). The antibodies neutralize pseudovirus with IC_{50} values ranging from 75.5 to 428.3 ng/mL, but while neutralization of live virus is 11- to 20-fold less potent for other antibodies, 5A6 retains similar potency with IC_{50} of 140.7 ng/ml (<2-fold weaker). We speculate that this difference in neutralization potency in live virus and pseudovirus assays is most likely due to the non-replicating nature of pseudoviruses, which are thereby more sensitive to blockade of viral entry. We validated live virus neutralization of a subset of antibodies (2H4, 3D11 and 5A6) using an RT-qPCR to quantify viral replication (**Figure 2.8A**), and observed similar trends as obtained from cell viability assays (**Figure 2.2B**). To more accurately assess the therapeutic potential of 5A6, we studied its neutralizing potency in SARS-CoV-2 infection of human airway epithelia (HAE) [34]. SARS-CoV-2 replication in HAE was reduced 1000-fold by 5A6 at 75 ng/mL and 10,000-fold at 150 ng/mL, and 5A6 also helped maintain epithelium integrity (represented by trans-epithelia electrical resistance), supporting its activity in a physiologically relevant *in vitro* model (**Figure 2.2C**).

All IgGs effectively block ACE2-RBD binding, with IC_{50} values below 50 nanomolar (**Figure 2.7A**). Therefore, the relative viral neutralization potencies of these antibodies cannot be ascribed to competitive receptor blocking alone. In order to interrogate other determinants of neutralization, we next compared the potency of each IgG antibody to their respective monomeric Fab fragments. A bivalent ACE2-Fc fusion protein was included as a reference for multivalent receptor blockade. All antibodies show dramatically increased potency against live virus compared to Fab, which is consistent with bivalent engagement of the Spike trimer by the IgG compared to the monovalent Fab fragment. The affinity or avidity for RBD is generally predictive of viral neutralization IC_{50}, with two striking exceptions (**Figure 2.2D**, **Table 3.2**). Antibody 5A6 exhibits far greater viral neutralization potency than other antibodies with superior avidity. Conversely, 3D11 is among the least potent in viral neutralization despite displaying the strongest binding.

We asked if the discrepancy could arise from the differences in the structural arrangement of IgGs bound to the RBDs of intact, trimeric Spike on the virion. We used surface plasmon resonance (SPR) to measure antibody binding to immobilized Spike trimers, as opposed to immobilized RBD (**Figure 2.8B-D**, **Table 3.2**). Although kinetics of binding to trimer were largely similar, we noted that 5A6 IgG binds somewhat more tightly (3.6x) to the intact trimer than to the flexible Fc-RBD construct used for BLI, while 3D11 IgG binds much (18.7x) more weakly (but still apparently bivalently, with 21.7x tighter binding than 3D11 Fab). To further investigate the relationship of the antibodies to intact Spike assemblies, we purified SARS-CoV-2 pseudovirus by gradient centrifugation and immobilized the viral particles on ELISA plates. The higher optical signal at saturation in concentration-dependent binding curves reveals that 5A6 IgG likely packs with higher density on the viral surface than the other four tested IgG antibodies or 5A6 Fab (**Figure 2.2E**). It has been proposed that effective viral neutralization requires antibody packing

density exceeding a critical threshold [35]–[37] and 5A6 may possess a unique binding mode that accommodates a denser structural arrangement on the viral surface. Notably, 3D11 IgG exhibits a similarly high signal at saturation, but with lower affinity for the pseudoviral particles, despite having higher affinity than 5A6 for immobilized RBD or Spike trimer (**Figure 2.2E**, **Figure 2.8**). Finally, 2H4 and 1F4 IgGs saturate immobilized pseudovirus at 1/3 the density of 5A6 or 3D11, while neutralizing live virus slightly more effectively than 3D11. These results suggest at least three different classes of receptor-blocking antibodies with distinct structural relationships to RBDs on viral particles.

Neutralizing antibodies inhibit or enhance Spike-mediated cell fusion

It is widely appreciated that viral proteins often possess multiple critical functions. With receptor-blocking activity and avid binding eliminated as sole determinants of neutralization by prior experiments, we reasoned that 5A6 might interfere with additional functions of the SARS-CoV-2 Spike. The most prominent example is induction of fusion of infected cells with neighboring cells, leading to formation of syncytia (multinucleated giant cells), a phenomenon known to hasten disease progression in respiratory syncytial virus [38] and human immunodeficiency virus [39], [40], and now widely observed in late-stage COVID-19 [10]. We therefore assessed whether antibodies discovered in our campaign inhibit Spike-mediated syncytia formation. To directly examine syncytia formation by Spike alone, we expressed Spike protein with a C-terminal fluorescent tag in Vero E6 cells. Addition of trypsin as an exogenous Spike-processing enzyme resulted in cells with a diffuse fluorescent signal and multiple nuclei, indicative of syncytia formation. (**Figure 2.3A**). We assayed the impact of receptor-blocking antibodies on this trypsin-induced cell-cell fusion, using antibodies 2H4, 5A6 and 3D11, which represent different modes of viral neutralization. The 5A6 IgG has a dose-dependent inhibitory effect on

syncytia (**Figure 2.3B**). By contrast 2H4 IgG has no significant effect. Surprisingly, 3D11 potentiates cell-cell fusion. Furthermore, 5A6 Fabs also enhanced syncytial fusion, albeit weakly (**Figure 2.3A-B**). We conclude that 5A6 IgG directly inhibits Spike-mediated fusion, while other receptor blocking antibodies fail to inhibit or even accelerate this process.

Structures of Spike-Fab complexes

The Spike trimer exists in equilibrium between the closed conformation, with all RBDs nestled closely around the S2 subunit, and "receptor seeking" states featuring one or more open RBDs that become erect and disengage from the S2 [4], [41], [42]. To provide structural insights into how antibodies targeting different RBD epitopes divergently modulate Spike protein function, we determined structures of the trimeric Spike protein alone (**Figure 2.9A-B**), and in complex with Fab fragments of 3D11 (**Figure 2.9C-D**), 2H4 (**Figure 2.9E**), and 5A6 (**Figure 2.9F-G**). Consistent with epitope binning and functional characterization, these Fabs explore different regions of the RBD surface, are compatible with different (open or closed) RBD states in the Spike trimer, and bind with distinct geometries relative to the RBD and ACE2 interface (**Figure 2.4**). This form of three-dimensional epitope mapping provides atomic level understanding of the determinants of viral inhibition.

2H4 is an orthosteric receptor-mimetic antibody

We determined multiple structures of 2H4 Fab bound to Spike with resolution sufficient for unambiguously docking a model of 2H4, but precluding precise modelling of the epitope and complementarity determining regions (CDRs) (**Figure 2.9E**). Three major conformational states were identified by 3D classification, revealing either one, two, or three 2H4 Fabs bound to the Spike trimer. The receptor blocking activity of 2H4 is straightforward, as it recognizes an epitope that overlaps much of the ACE2 interface (**Figure 2.4A**). Binding of the 2H4 Fab is compatible

with both major RBD conformations, and the structures are drawn from an ensemble of quaternary states reminiscent of those that follow ACE2 binding, and lead to S1 shedding and spike-mediated membrane fusion [2] (**Figure 2.4B**). The first of three predominant states features 2H4 bound to one open RBD, and the other two RBDs closed. The second state adds a second copy of 2H4, on a closed RBD counter-clockwise from the first. Density inspection and 3D variability analysis (3DVA) [43] reveal that the third RBD is primarily open, and a trajectory of opening states correlate with binding the second Fab (**Movie 2.2**). The final state features three Fabs bound, and a strictly open third RBD. This restricted ensemble arises because when the bound RBD is closed, the Fab incurs clashes with the counter-clockwise adjacent RBD that can only be relieved by opening of that RBD and NTD, even beyond the degree of opening in the triple ACE2 complex with three open RBDs [2] (**Figure 2.4C**). These observations suggest 2H4 directly blocks receptor binding, but also acts as a receptor-mimetic that admits the same cycle of Spike conformations as does ACE2. A neutralizing antibody against SARS-CoV has also been reported to engage in orthosteric receptor mimicry [44], suggesting activation of fusion-associated conformational changes may be an intrinsic consequence of direct receptor interface binding in betacoronaviruses.

3D11 allosterically blocks ACE2 binding and triggers Spike opening

The Spike:3D11 complex is relatively homogeneous, with only one major state (**Figure 2.4D**), and we determined its structure to ~3.0 Å resolution (**Figure 2.9C**). All three RBDs are bound to 3D11 Fab in the open conformation, with the Fab making a right angle to the long axis of the RBD, via an epitope exposed only in the open state and outside the RBM (**Figure 2.4A**). The epitope partly overlaps those of some other antibodies that bind outside the RBM [45], [46], but is distinct from those [26], [47] that bind freely to a closed RBD (**Figure 2.10D**). Clashes between 3D11 and Spike NTDs also prevent 3D11 binding to closed RBDs, indicating 3D11 binds

only to open RBDs. This restriction of binding to the subset of Spike conformations with open RBDs might account for lessened avidity of 3D11 IgG for the intact Spike trimer, as compared with RBD (**Figure 2.8D**). 3D classification reveals outward motions of the NTD and variation in Fab occupancy, but only open RBDs are observed (**Figure 2.10A**). As for 2H4, the 3D11 bound RBDs are "more open" (displaced further outward) than those in the triple ACE2 complex (**Figure 2.4E**).

Although its epitope does not significantly overlap with the ACE2-RBD interface, 3D11 nevertheless effectively blocks ACE2 binding and stabilizes a quaternary state of the Spike, with three open RBDs and NTDs, that closely resembles the penultimate stage of ACE2-induced Spike opening [2]. We therefore term 3D11 an allosteric receptor-mimetic antibody, which does not directly target the ACE2 interface, yet prevents ACE2 binding and enhances Spike-mediated fusion by rapidly advancing the Spike conformational cycle to its final stages.

5A6 traps a pre-fusion conformation to inhibit spike-mediated fusion

Multiple states of the Spike:5A6 complex were resolved to better than 3.0 Å, with local resolution sufficient for accurate modelling of the Fab-RBD interface (**Figure 2.9G**). 5A6 recognizes surface loops near the tip of the RBD, which are solvent exposed even when all RBDs are closed. The binding geometry is permissive of any trimer configuration and any stoichiometry, without steric constraints from Spike or Fab. Yet despite this complete conformational freedom, all 5A6 complexes feature at least two 5A6 Fabs bound to an open RBD counter-clockwise adjacent in turn from a closed RBD (**Figure 2.4F**). The hallmark of these states is a cryptic, quaternary epitope in which a region of the Fab V_L domain makes a second interaction with an adjacent, open RBD (**Figure 2.4G**). This new interaction is not possible with a closed RBD, and requires an adjustment away from the average positions of open RBDs in other structures (**Figure**

2.10B). The closed RBD bound via the main 5A6 interface is also displaced from other closed conformations.

CDR loops H1, H2, H3, and L1 engage the canonical epitope with a buried surface area of 850 $Å^2$ (**Figure 2.5A**). Partial overlap of the RBM and Fab interface, and a clash induced between ACE2 and the Fab V_L domain, are likely sufficient to exclude ACE2 from the RBD. The second interaction contributes an additional 363 $Å^2$ (**Figure 2.5B**), and the C_{H1} and C_L domains of 5A6 at the cryptic quaternary epitope induce an even more severe clash with ACE2 (**Figure 2.4G**). Two Fabs thus act synergistically to block ACE2 binding, while one Fab is capable of blocking ACE2 at two RBDs simultaneously. The secondary interface must be released in order for the bound RBD to open, and we hypothesize that 5A6 at its quaternary epitope locks one RBD closed, thereby arresting the trimer in its pre-fusion state. Precise conservation of binding mode and the cryptic quaternary epitope from free Fab to IgG is confirmed by a structure of 5A6 IgG complexed with Spike trimer at ~15 Å resolution (**Figure 2.5C**). Although the Fc domain is not well resolved due to the flexibility of the hinge region, the structures suggest 5A6 IgG may bind to two RBDs from the same trimer (**Figure 2.10C**), and shows that no steric effects preclude binding of three IgGs at sufficient concentration. Noting the weak potentiation of syncytia formation by 5A6 Fab, we deduce that the cryptic epitope likely appears following initial Fab binding, and leads to cooperative action against SARS-CoV-2 by imbuing a second binding event with enhanced affinity and receptor blockade. We also conclude that the geometry of the quaternary epitope and avidity of 5A6 IgG drive robust pre-fusion conformational trapping and potent inhibition of Spike-mediated fusion and syncytia formation.

Discussion

Receptor engagement to a Spike RBD locks it in the open conformation and triggers a cooperative process in which the Spike conformational ensemble is driven towards further opening by successive rounds of receptor binding [2]. The process culminates in unsheathing of the S2 subunit, and following proteolytic cleavage, shedding of the S1-ACE2 subcomplex. S1 shedding in turn facilitates the post-fusion state transition, leading to membrane fusion and viral entry. Spike protein on the surface of infected cells can also mediate ACE2-dependent fusion of neighboring cells to form multinucleated giant cells, presumably through the same cycle of proteolysis and conformational transitions. Genetic variation in Spike, inside and outside the RBM, can influence function by altering the conformational equilibrium of the trimer, as seen for the D614G variant which more readily populates states with multiple open RBDs [42] and exhibits heightened infectivity [48].

We identified six receptor-blocking antibodies that exhibit differences in avidity, binding mode, and neutralization of live virus, despite the fact that all of them have similarly high affinities for the RBD of the viral Spike protein. Strikingly, one highly potent, neutralizing antibody inhibits Spike-mediated syncytia formation, while another more weakly neutralizing antibody actually potentiates cell-cell fusion. These findings suggest the potential effectiveness of a neutralizing antibody is influenced by a number of possibly countervailing factors, highlighting a complex basis for viral neutralization potency, *contra* a simplistic view in which receptor blocking and affinity enhancement through avidity are the sole determinants of viral neutralization. Cryo-EM structures reveal the conformational landscapes of three Spike-Fab complexes, and provide 3D mapping of antibody-Spike epitopes, from which we can draw mechanistic insights that explain their distinct bioactivities in terms of the known intermediates in Spike opening (**Figure 2.6**).

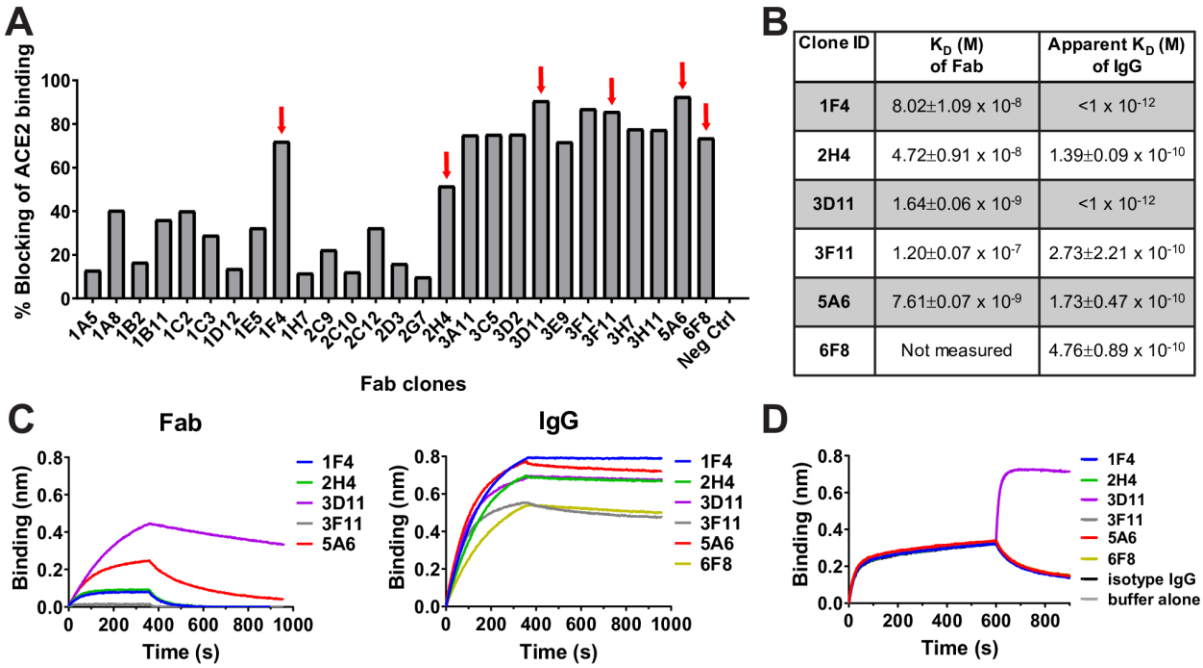
Certain receptor-blocking antibodies, such as 2H4, may function as orthosteric receptor mimetics conducive to the same cooperative processes as receptor binding. Another class of antibodies, exemplified by 3D11, act as allosteric effectors that advances the Spike directly to the final stages of S2 unsheathing. In contrast, 5A6 possesses a unique binding mode that stabilizes a Spike conformation that prohibits S1 shedding and traps the pre-fusion state. We hypothesize that synergy between receptor blockade and pre-fusion trapping allows 5A6 to prevent both targeted viral fusion and Spike-mediated cell-cell fusion.

Directly comparing the monovalent Fabs and bivalent IgGs also provides insight into the nature of avidity. The magnitude of binding and neutralization enhancement in the IgG format supports bivalent binding for all six antibodies [49]. This result may be counterintuitive because modelling studies suggest the hinge linking IgG Fc and Fab domains may have difficulty bridging the gaps seen in our structures of Spike:Fab complexes (**Figure 2.10C**). On the other hand, our structure of the 5A6 IgG complex confirms that 5A6 Fab and IgG forms do bind with congruent geometries, and a morph between 5A6 Fab bound to closed and open RBDs shows that shorter distances between Fabs do obtain for intermediate RBD conformations (Movie 2.3). These results may imply that high-affinity, bivalent binding to intermediate RBD conformations is replaced by high-density, monovalent binding as the concentration of 5A6 IgG increases. In contrast to 5A6, 3D11 and 2H4 Spike:IgG complexes were not tractable for single-particle cryo-EM (**Figure 2.10E**). Qualitative image analysis suggests these species do not trap defined conformational states of the Spike trimer, and that 3D11 in particular may achieve viral neutralization by destabilizing the Spike, as does the CR3022 antibody with a similar epitope [27]. Intriguingly, we found that 3D11 has greatly reduced potency against pseudovirus bearing D614G Spike, while that of 5A6 is slightly improved, though position 614 is outside the RBD and far from the epitope of either

antibody in the RBD (**Figure 2.12**). The D614G mutant Spike is known to occupy states with multiple open RBDs [42], and has been found to shed the S1 subunit less readily than the original SARS-CoV-2 Spike protein [50]. We can thus understand altered neutralization of D614G pseudovirus by 3D11 and 5A6 in terms of the model of **Figure 2.6**, because the effects of both antibodies are mediated by open RBD conformations that represent immediately available binding sites for 3D11, and present the full quaternary epitope of 5A6. The reduced S1 shedding of the more stable D614G Spike may also assist the pre-fusion trapping activity of 5A6, while conveying resistance against trimer denaturation by 3D11.

The quaternary epitope recognized by 5A6 conveys cooperative binding as well as avidity, and both aspects may hinder viral escape via mutations in Spike protein [47]. In summary, our work establishes a platform for exploring allosteric control surfaces across the SARS-CoV-2 Spike protein. While we employ syncytia formation as a general model for Spike protein function, pre-fusion trapping by 5A6 may convey greater efficacy against late-stage or severe illness due to the role of syncytia in COVID-19 pathology [10]. Notably, cell-cell fusion during respiratory syncytial virus (RSV) infection also leads to severe tissue damage, and the therapeutic antibody Palivizumab is known to trap a pre-fusion state of the fusogenic F protein of RSV [51]. Uncovering the particular modes of action and functional consequences of Spike binders in general may deepen our understanding of avidity, viral neutralization, and disruption of syncytia formation and play a role in the development of therapeutics or vaccines during the on-going COVID-19 pandemic.

Figures



Clone ID	K_D (M) of Fab	Apparent K_D (M) of IgG
1F4	$8.02\pm1.09 \times 10^{-8}$	$<1 \times 10^{-12}$
2H4	$4.72\pm0.91 \times 10^{-8}$	$1.39\pm0.09 \times 10^{-10}$
3D11	$1.64\pm0.06 \times 10^{-9}$	$<1 \times 10^{-12}$
3F11	$1.20\pm0.07 \times 10^{-7}$	$2.73\pm2.21 \times 10^{-10}$
5A6	$7.61\pm0.07 \times 10^{-9}$	$1.73\pm0.47 \times 10^{-10}$
6F8	Not measured	$4.76\pm0.89 \times 10^{-10}$

Figure 2.1 Isolation of SARS-CoV-2 receptor-blocking antibodies from a naïve human library

(A) Blocking of ACE2/RBD (SARS-CoV-2) interactions by 27 Fab clones, tested by competition ELISA. The samples used in the assays were unpurified Fab from bacterial supernatant, hence the percentages of blocking were not indicative of their true potency. Red arrows indicate the 6 clones in subsequent studies.

(B) K_D of Fab based on 1:1 Langmuir fitting and apparent K_D of IgG based on 1:2 bivalent analyte fitting of BLI sensorgrams for immobilized Fc-RBD. Values are the mean and standard deviation of two independent experiments.

(C) Binding and dissociation of Fabs and IgGs to and from immobilized Fc-RBD by BLI. The concentration of Fabs and IgGs shown is 12.5 nM.

(D) Epitope binning of 5A6 by BLI analysis. 5A6 is immobilized as the ligand. Tagless RBD is introduced as the first analyte. The second antibody is introduced as the second analyte. As controls, buffer alone, an isotype IgG and 5A6 IgG were included as the second analyte. See also **Figure 2.7**, **Table 3.1**, and **Table 3.2**.

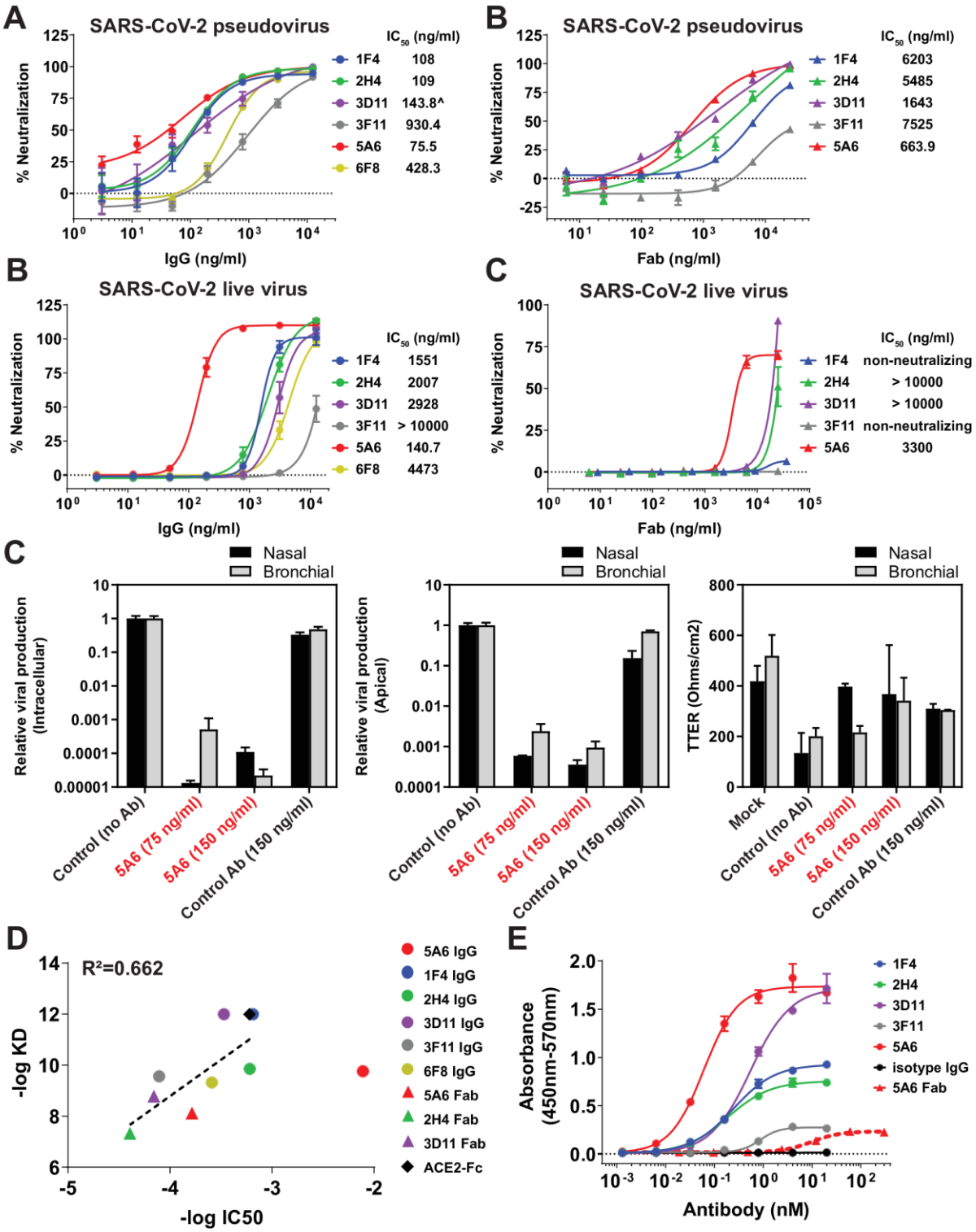


Figure 2.2 SARS-CoV-2 neutralization by receptor blocking antibodies

(A) Infection of CHO-ACE2 cells by SARS-CoV-2 pseudovirus were determined in the presence of receptor blocking IgGs (left panel) or Fabs (right panel). Luciferase activities in the CHO-ACE2 cells were measured, and the percent neutralization was calculated. Data are presented as mean ± SEM in triplicates and are representative of two independent experiments. The IC_{50} was calculated

by a variable slope four parameter non-linear regression model using either Graphpad PRISM 7 Software, or ^Quest Graph™ IC_{50} Calculator from AAT Bioquest, Inc (https://www.aatbio.com/tools/ic50-calculator) with top and bottom constraints set at 100% and 0% respectively.
(B) Infection of Vero E6 C1008 cells by SARS-CoV-2 live virus (isolated from a nasopharyngeal swab of a patient in Singapore) were determined in the presence of receptor blocking IgGs (left panel) or Fabs (right panel). Infection induced cytopathic effect was determined by detecting the amount of ATP present in the uninfected live cells from which the percent neutralization was calculated. Data are presented as mean ± SEM in triplicates and are representative of two independent experiments. The IC_{50} was calculated by a variable slope four parameter non-linear regression model in Graphpad PRISM 7 Software. Pseudo- and live virus neutralization assays were not performed for 6F8 Fab due to 6F8 IgG's low and similar potency to other clones.
(C) Evaluation of antiviral activity of 5A6 in a model of reconstituted human airway epithelia (HAE). Viral genome quantification was performed using RT-qPCR, and results are expressed in relative viral production (intracellular or apical) compared to control. Bars represent the mean ± SD in duplicates. ****P <0.001*, ** *P<0.01 and *P <0.05* compared to the control (no Ab) by one-way ANOVA. The trans-epithelial resistance (TEER in Ω/cm^2) was measures at 48hpi.
(D) Correlation curve of affinity/avidity for RBD and live virus neutralization potency (IC_{50}) of receptor blocking IgG antibodies (circles), Fab antibodies (triangles) and ACE2-Fc (black diamond). The IC_{50} values were calculated using a four parameter logistic regression model in the Quest Graph™ IC_{50} Calculator from AAT Bioquest, Inc.
(E) Binding of the IgG (solid lines, circles) and 5A6 Fab (dashed line, red triangle) to the purified SARS-CoV-2 pseudovirus. See also **Figure 2.8** and **Table 3.2**.

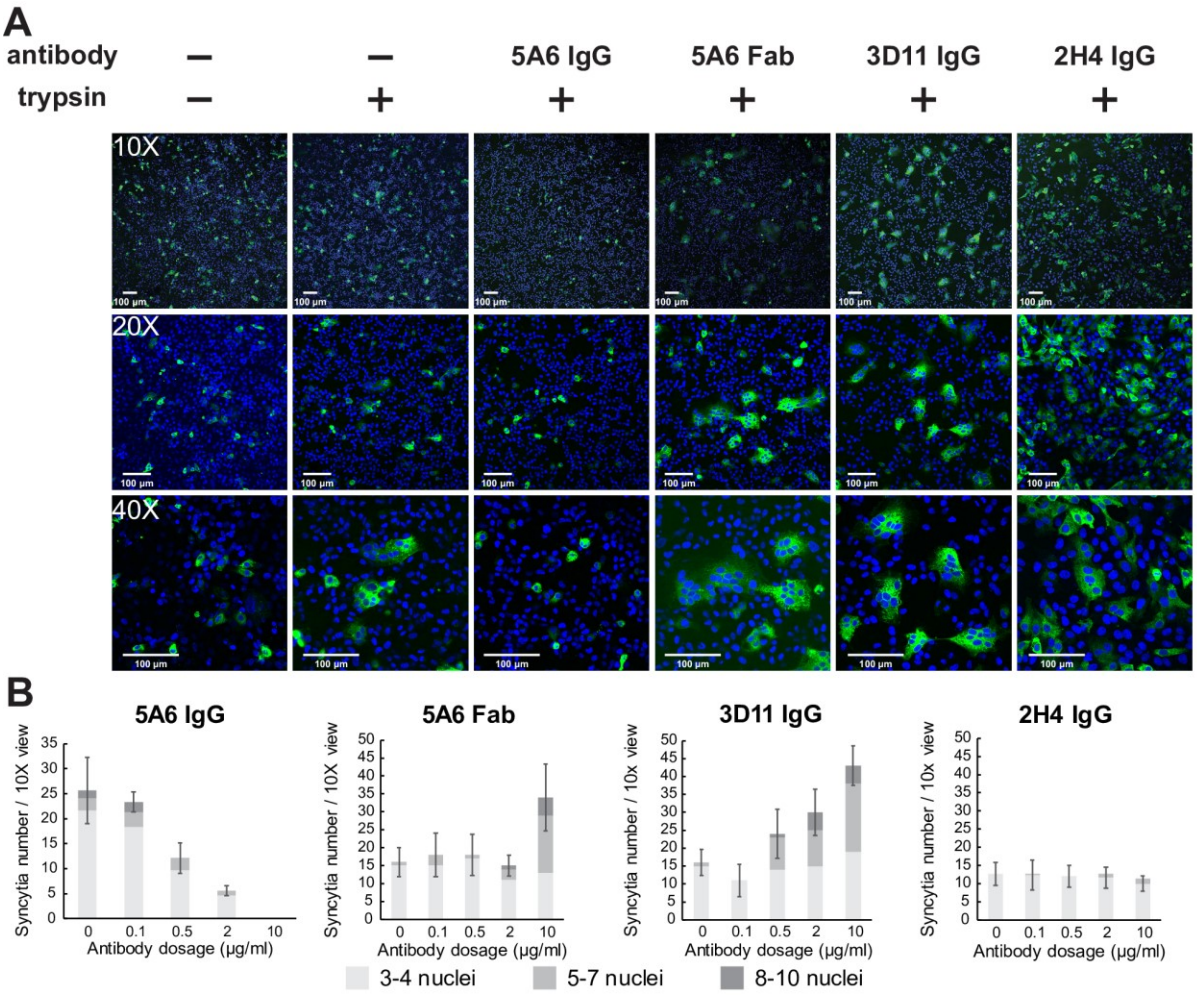


Figure 2.3 Anti-SARS-CoV-2 Spike RBD IgG antibodies affect trypsin induced cell syncytia formation

Vero E6 cells were transfected with furin recognition mutation of SARS-CoV-2 S-protein (R682RAR to A682AAR)-GFP. After 48 hours, the cell culture medium was changed to DMEM (no serum) and treated with or without antibodies and incubated for 1 hour at 37°C. The cells were then treated with or without trypsin at 15 µg/ml for 2 hours at 37°C. Cells were fixed with 4%PFA and stained with DAPI.

(A) 5A6 IgG (20 µg/ml), 5A6 Fab (20 µg/ml), 3D11 IgG (20 µg/ml) and 2H4 IgG (10 µg/ml) treated S-protein expressing Vero E6 cell images in 10x, 20x and 40x objective view. Images were taken by Olympus confocal microscope.

(B) Dosage response of 5A6 IgG, 5A6 Fab, 3D11 IgG and 2H4 IgG. S-protein expressing Vero E6 were treated with 0, 0.1, 0.5, 2, 10 µg/ml of each antibody. $2x10^5$ cells/sample were used for transfection of 5A6 IgG dosage response. $1.6x10^5$ cells/sample were used for tranfection of 5A6 Fab, 2H4 IgG and 3D11 IgG. Data quantification were calculated on syncytia numbers and nuclei number presented in each syncytium.

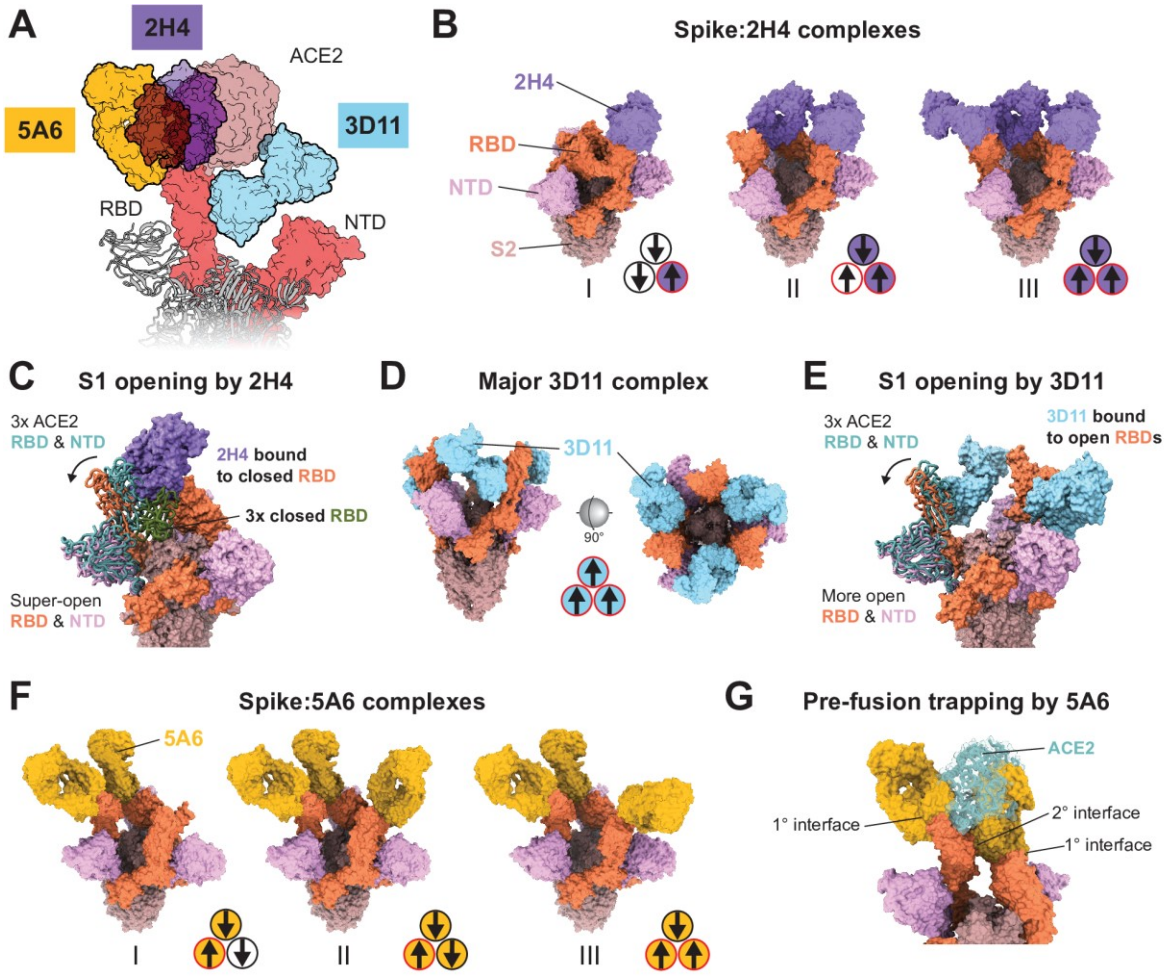


Figure 2.4 Structures of Spike-Fab complexes

(A) Schematic representation of three Fabs, 5A6 (goldenrod), 2H4 (purple), and 3D11 (sky blue), bound to the SARS-CoV-2 Spike protein. All Fabs are shown in relation to the complex formed by an open RBD (red) and the extracellular domain of ACE2 (rose brown).

(B) Spike:2H4 complexes depicted as surface models, with 2H4 in purple, RBDs in coral, NTDs in plum, and the S2 core in rose brown. A triad of small circles to the lower left of each complex figure represent the three RBDs, with an inset arrow indicating the open (up) or closed (down) conformation, and purple fill indicated a bound Fab. The ensemble of Spike:2H4 complexes is reminescent of several intermediates in the conformational cycle triggered by serial binding of ACE2.

(C) A cut-away of Spike:2H4 complex II highlighting the steric effect of 2H4 bound to a closed RBD on the counter-clockwise adjacent RBD and NTD. The 2H4-bound RBD clashes with both RBDs from either the fully closed trimer conformation (PDB: 6zgi; dark green) and the fully opened conformation with three ACE2 molecules (PDB: 7a98; cadet blue).

(D) The single major Spike:3D11 complex, colors as in B (3D11 in sky blue). All RBDs are open, resembling the open trimer bound to three copies of ACE2. At right, the extracelluar view along the Spike trimer axis reveals its S2 core essentially unsheathed.

(E) A cut-away of the Spike:3D11 complex showing its effect on the neigboring RBD is similar to that of 2H4, but less pronounced. Colors as in C., 3D11 in sky blue.

(F) Spike:5A6 complexes, colors as in B. with 5A6 in goldenrod. Each complex exhibits the same quaternary state across two RBDs, with 5A6 bound between one closed RBD and a counter-clockwise adjacent open RBD, also bearing a copy of 5A6. The third RBD is usually open and may also bind 5A6.
(G) The quaternary epitope seen in all Spike:5A6 complexes involves two distinct interfaces, and appears to trap the pre-fusion Spike by locking closed one RBD. Binding to either interface results in steric occlusion of ACE2 (PDB: 7a94; cadet blue) from the RBD. Notably, both Fabs synergistically block ACE2 binding to the open RBD, while the Fab at the quaternary epitope blocks ACE2 binding to two RBDs simultaneously. See also **Figure 2.9-3.11**, **Table 3.3** and **Movie 2.1-3.3**.

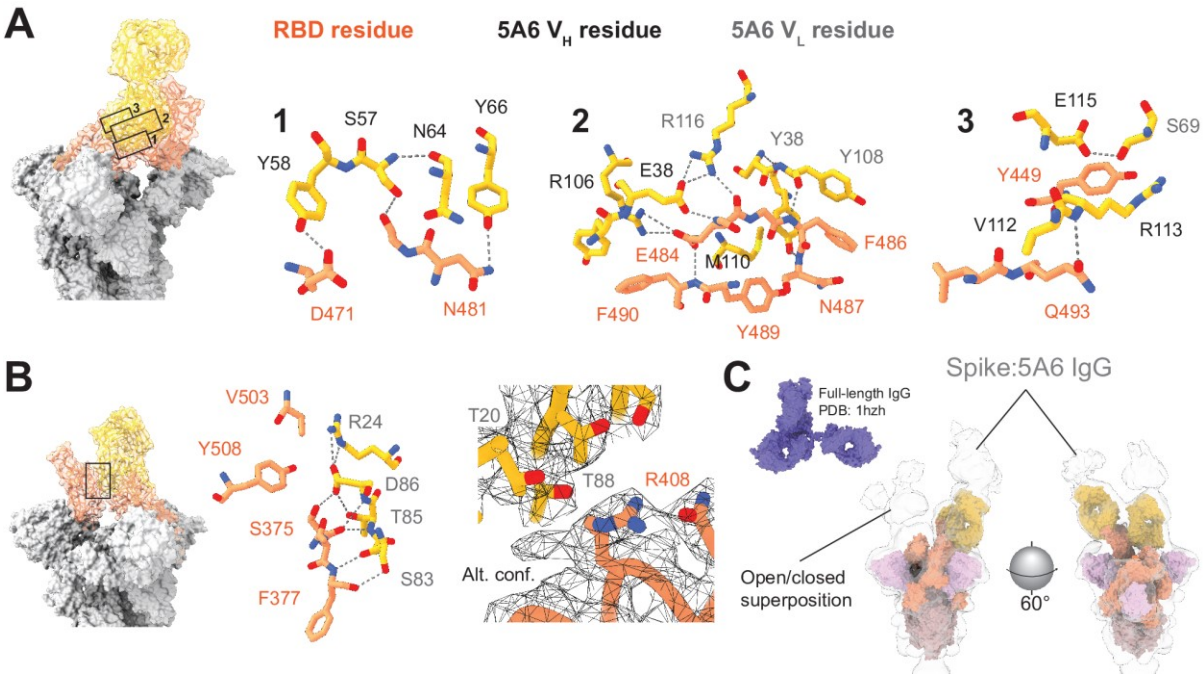


Figure 2.5 Binding mode and epitope of 5A6

(A) A tour through the primary interface between Spike RBD and 5A6, using three immediately adjacent cross-sections along the viewing axis. Spike residues and labels are colored in coral or by heteroatom, and 5A6 residues are colored goldenrod, with labels for V_H residues in black and V_L residues in gray. Fab residue labels use the IMGT numbering system (Lefranc et al., 2003). Predicted hydrogen bonds are shown as dashed gray lines. The interface features extensive hydrogen bonding, numerous hydrophobic contacts, and multiple salt bridges. A aromatic cluster formed by 5A6 V_L Y38 and Y108, and RBD F486, a salt bridge between 5A6 V_H R106 and RBD E484, and a cation-pi interaction between Fab V_H R112 and RBD Y449 are particularly notable.

(B) The secondary interface between 5A6 and the neighboring open RBD comprises mostly hydrogen bonds, many of which involve main-chain atoms. Atom and label colors are as in (A). An interesting feature is stabilization of an alternate conformation of RBD R408 by 5A6 V_L T20 and T88.

(C) Two views of the Spike:5A6 IgG complex cryo-EM map (translucent gray) overlaid on the surface model of Spike:5A6 Fab complex I (colors as in Figure 4). The docked Fab complex model and the IgG complex cryo-EM density reveal a congruent epitope binding geometry in both formats, with two RBDs in the characteristic open/closed trapped conformation. Fc domains are visible as unmodeled blobs to the back of each Fab, and are like a superposition of multiple possible stoichiometries. The third RBD seems to be a superposition of open and closed RBDs with IgG bound. The surface model of a full-length human IgG x-ray crystal structure (PDB: 1hzh) is shown for scale (comparison is facilitated by use of orthographic projection).

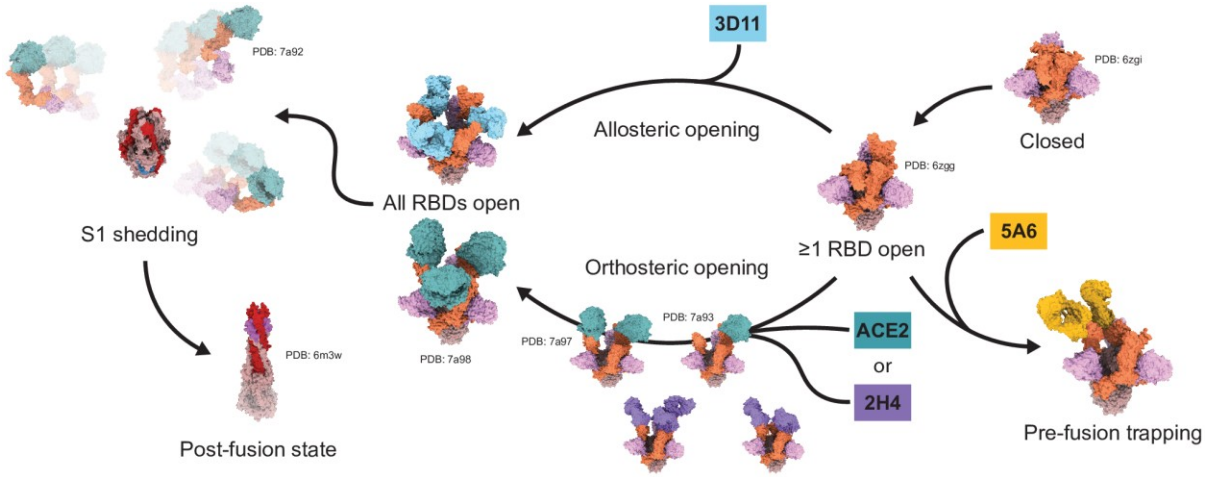


Figure 2.6 Spike functional modulation by receptor-blocking antibodies
A schematic model representing the possible effects of receptor-blocking antibodies, and ACE2 itself, on the conformational cycle of the SARS-CoV-2 Spike trimer. On the viral surface, the Spike is found predominantly in either the closed conformation, or a "receptor seeking" conformation with one RBD open. When serially bound by ACE2 or an orthosteric mimetic antibody like 2H4, the Spike trimer passes through a series of conformations that eventually permit S1 shedding and the S2 post-fusion transition that mediates membrane fusion. Alternatively, allosteric antibodies such as 3D11 can advance the trimer directly to the end of the opening process, potentiating formation of syncytia through fusion of neighboring cells. Allosteric opening most likely contributes to lower potency in a high-affinity receptor-blocking antibody, and might even suggest the possibility antibody-dependent enhancement of infection. Finally, the Spike might instead be recognized by 5A6, which inhibits membrane fusion and syncytia formation by preventing S1 shedding and trapping the pre-fusion trimer. By enjoining the exposure and conformational transition of the S2 subunit, the 5A6 complex represents an unproductive dead-end for the Spike trimer. See also **Figure 2.12**.

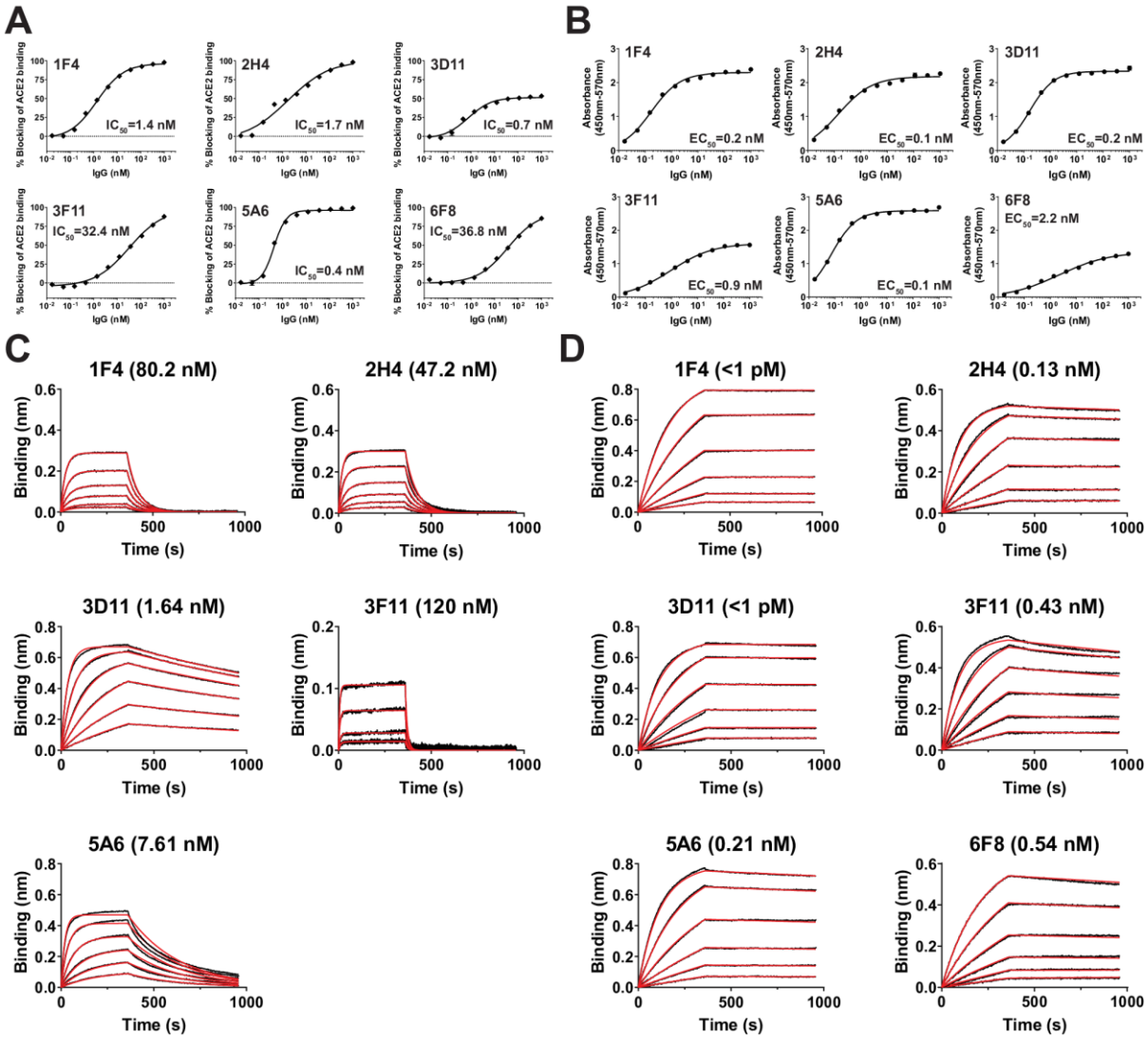


Figure 2.7 Characterizations of SARS-CoV-2 receptor-blocking antibodies

(A) Blocking of ACE2/SARS-CoV-2 RBD interaction by 1F4, 2H4, 3D11, 3F11, 5A6 and 6F8 IgGs tested by competition ELISA. Data are presented as mean ± SD in triplicates and are representative of two independent experiments.

(B) Binding avidity of 1F4, 2H4, 3D11, 3F11, 5A6 and 6F8 IgG antibodies to SARS-CoV-2 Spike RBD proteins tested by ELISA. Data are presented as mean ± SD in triplicates and are representative of two independent experiments.

(C) Binding affinity of five Fab clones to SARS-CoV-2 Spike RBD protein measured by biolayer interferometry. Fab binding to immobilized Fc-RBD was tested using a range of Fab concentrations from 100 nM to 3.125 nM (in 2-fold dilution). A representative set of measurements from two independent experiments is shown with sensorgrams in black and curve fittings in red.

(D) Binding avidity of six IgGs to the RBD by biolayer interferometry. IgG Binding to immobilized Fc-RBD was tested using a range of IgG concentrations from 12.5 nM to 0.39 nM (in 2-fold dilutions). The anti-Fc sensor chip was quenched with excess irrelevant, same-isotype IgG to prevent confounding from antibody binding directly to the chip. A representative set of measurements from two independent experiments is shown with sensorgrams in black and curve fittings in red.

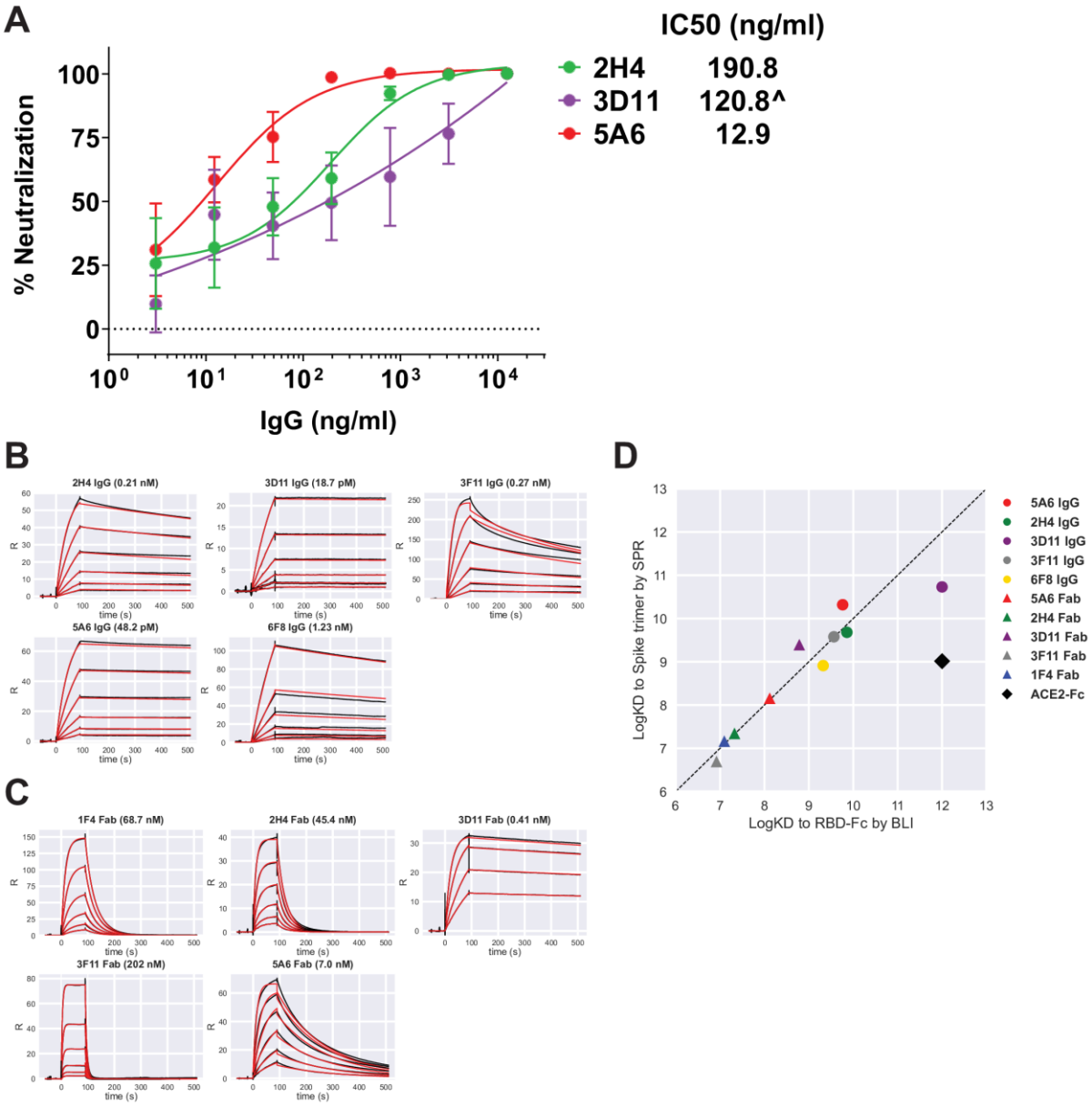


Figure 2.8 SARS-CoV-2 live virus neutralization and antibody binding to Spike trimer measured by surface plasmon resonance

(A) The potency of 2H4, 3D11 and 5A6 IgG antibodies in neutralizing live SARS-CoV-2 virus assays determined by measuring the viral genome copy number (GCN). Infection of Vero E6 C1008 cells by SARS-CoV-2 live virus (isolated from a nasopharyngeal swab of a patient in Singapore) were determined in the presence of receptor blocking IgGs 2H4, 3D11 and 5A6. 48 hours post infection, culture supernatant was harvested and viral GCN was determined by RT-qPCR targeting the N gene and GCN values were determined by comparing Ct values against a logGCN standard curve. The GCN values were then converted to percent neutralization and plotted with a non-linear regression curve fit using PRISM. IC_{50} was calculated by a variable slope four

parameter non-linear regression model in Graphpad PRISM 7 without or ^with top and bottom constraints set at 100% and 0% respectively. Data are presented as mean ± SEM from 6 replicates.
(B) Binding avidity of IgG clones 2H4, 3D11, 3F11, 5A6, and 6F8 for intact Spike trimer measured by surface plasmon resonance (SPR). A range of IgG concentrations from 12.5 nM to 0.39 nM (in 2-fold serial dilution) are shown, with sensorgrams in black and curve fits in red.
(C) Binding affinity of Fab clones 1F4, 2H4, 3D11, 3D11, and 5A6 for intact Spike trimer measured by surface plasmon resonance (SPR). A range of Fab concentrations from 100 nM to 3.125 nM (in 2-fold serial dilution) are shown, with sensorgrams in black and curve fits in red.
(D) Log-log scatter plot comparing antibody binding constants for Fc-RBD (X axis) to those for Spike trimer (Y axis). Affinity or avidity of antibodies or Fab fragments for the flexible Fc-RBD construct represent binding without geometric constraints, while measurements using immobilized Spike trimers represent binding with the specific geometries afforded by the Spike:antibody complexes. For most species, SPR and BLI measurements are similar, however 3D11 IgG and ACE2-Fc bind significantly more weakly to relatively unrestricted Fc-RBD than to Spike trimer (note that the 3D11 IgG binds >10x more tightly than 3D11 Fab in both sets of experiments, indicating avid binding). 5A6 IgG binds somewhat more tightly to Spike trimer than to Fc-RBD, perhaps indicating that RBDs within a Spike trimer have particularly favorable geometries for binding.

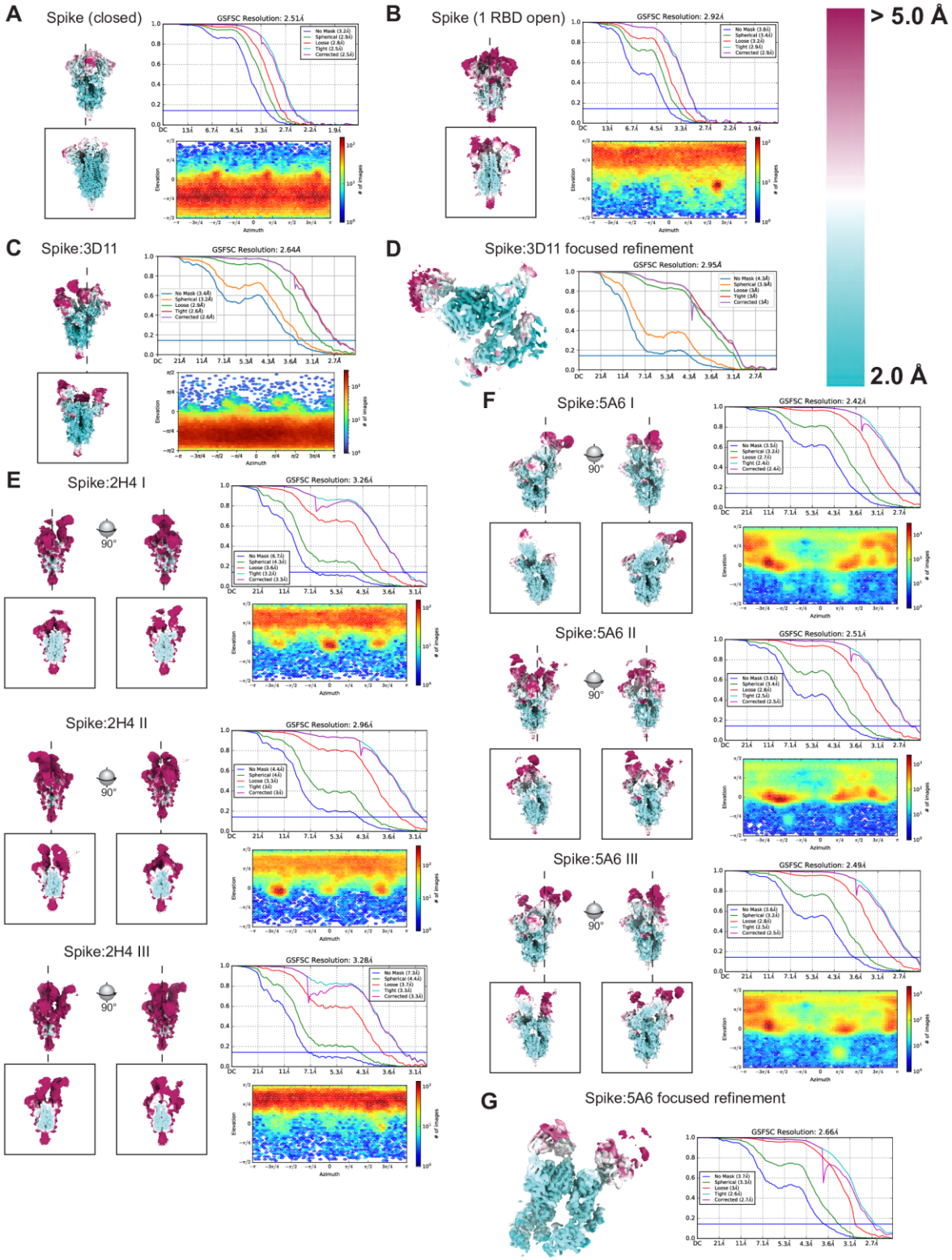


Figure 2.9 Cryo-EM densities and resolution estimation
Density maps colored by local resolution, Fourier shell correlation curves, and particle orientation distributions for the structures reported in this work. All maps use the same local resolution scale, shown at the top right of the figure.

(A) The apo Spike, with all RBDs closed.
(B) The apo Spike, with one RBD open.
(C) The Spike:3D11 complex.
(D) Refinement of Spike:3D11, focused on the Fab variable domains and RBD epitope.
(E) Spike:2H4 complexes with one, two, or three Fabs bound.
(F) Spike:5A6 complexes.
(G) Refinment of Spike:5A6 complex I, focused on the Fab variable domains and quatenary epitope involving two RBDs (one open, with the clockwise adjacent RBD trapped closed).

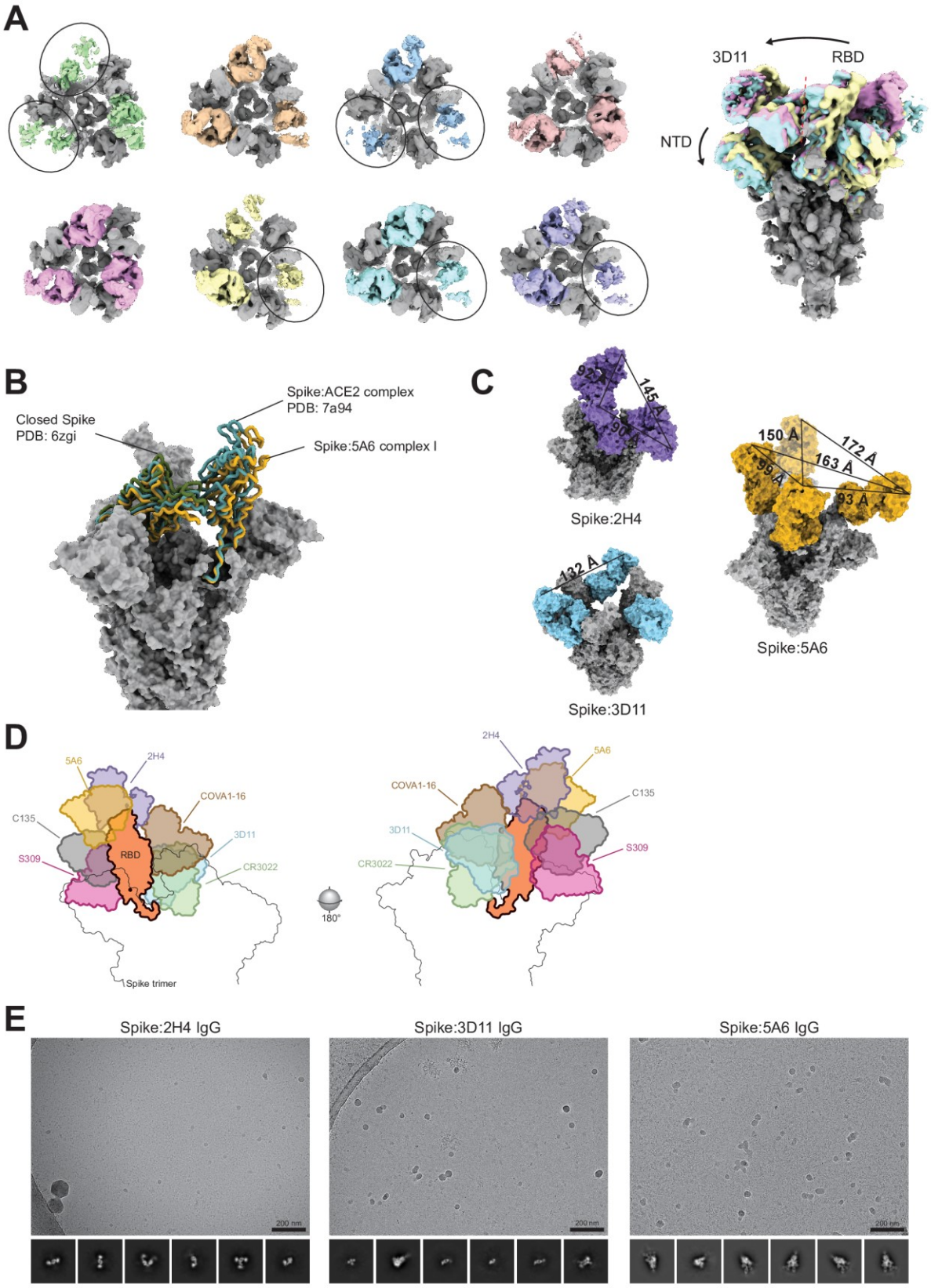


Figure 2.10 Additional structural details

(A) Eight subclasses of the Spike:3D11 complex, determined using symmetry relaxation in Relion 3.1. At left, top views show that these classes vary in the occupancy of Fab at each RBD. Only Fabs are colored, with missing or weak Fab densities are indicated by black ellipses. At right, two extremum classes and one intermediate show relative motion of the RBDs and NTDs, relaxations which likely contribute to the S2 unsheathing that eventually permits Spike-mediated membrane fusion. Fabs, RBDs, and NTDs are colored, and a dashed line delineates Fab and RBD. The general direction of S2-opening movements, as observed in the classes, are indicated by black arrows.
(B) The quaternary epitope bound by 5A6 is cryptic because the Fab stabilizes unique conformations of RBDs, observed only the the complex, that contribute to the epitope. RBDs from Spike:5A6 complex I (goldenrod) are shown relative to the fully closed Spike (PDB: 6zgi, forest green) and to the Spike with one RBD open and bound to ACE2 (PDB: 7a94, cadet blue). Interestingly, the open RBD bound to 5A6 is more open than that bound to ACE2, yet the neighboring RBD trapped closed by 5A6 is more closed than in the singular ACE2 complex or the fully closed Spike.
(C) The hinge regions connecting the Fc domain of an IgG antibody to each of its two Fabs are 23 residues long, approximately 10 of which are flexible due to disulfide bonds. Assuming a standard polypeptide length of about 3.5 Å per residue, each hinge might extend as far as 35 Å, allowing for some 70 Å separation between the two Fabs. As shown, the shortest gaps between Fabs in the Spike:2H4 and Spike:5A6 complexes are ~90 Å, however variation of the elbow angles between Fab V and C domains could reduce the effective separation. Different Fab clones have elbow angles across just over 90° (Stanfield et al., 2006), and changes as great as 37° have been observed between multiple structures of the same clone (Wilson and Stanfield, 1994). For example, a 15° elbow bend might reduce separation by 10 Å at each Fab (20 Å total), to about the maximum length of the hinge. Bivalent, IgG-bound states thus likely differ in Fab elbow angle and feature some relaxation of the RBDs, in order to support the avid binding of IgG antibodies to Spike trimer observed in our experiments.
(D) Comparison of 2H4, 3D11, and 5A6 to non-receptor binding motif antibodies reported in literature. The 2D representation of molecular surface models, with superposed RBDs, shows that C135 and S309 bind to the outward face of the RBD, which is exposed in the closed state, while our antibodies 5A6 and 2H4 bind to the tip of the RBD at epitopes that partially overlap the RBS. Finally, COVA1-16, CR3022, and our 3D11 clone all bind to the inward face of the RBD, which is hidden in the closed state. These last three have overlapping epitopes, but different binding geometries such that, unlike COVA1-16 and 3D11, CR3022 actually clashes severely with the NTD even when the bound RBD is open.
(E) Example cryo-EM images and selected 2D class averages for Spike incubated with 2H4 IgG (left), 3D11 IgG (center), or 5A6 IgG (right). Images were collected on a Talos Arctica at 200 keV and low-pass filtered to 20 Å. The Spike:2H4 IgG sample contains numerous, relatively small particles, and some of its 2D class averages resemble monomeric Spike (classes 4 and 6) or two Spike monomers crosslinked by antibody (class 3). In the Spike:3D11 IgG sample, much of the protein is contained within stereotypical aggregates approximately 150 nm in size. Notable 2D averages resemble a Fab bound to Spike RBD (classes 3-5), rare Spike trimers in the closed conformation (class 2), and a potential Spike dimer or pair of monomers crosslinked by antibody (class 6). In contrast to the others, Spike:5A6 IgG is a well behaved sample (despite the crowded micrograph). The 2D class averages are easily recognizable as intact Spike trimer, as is the unique binding mode of 5A6.

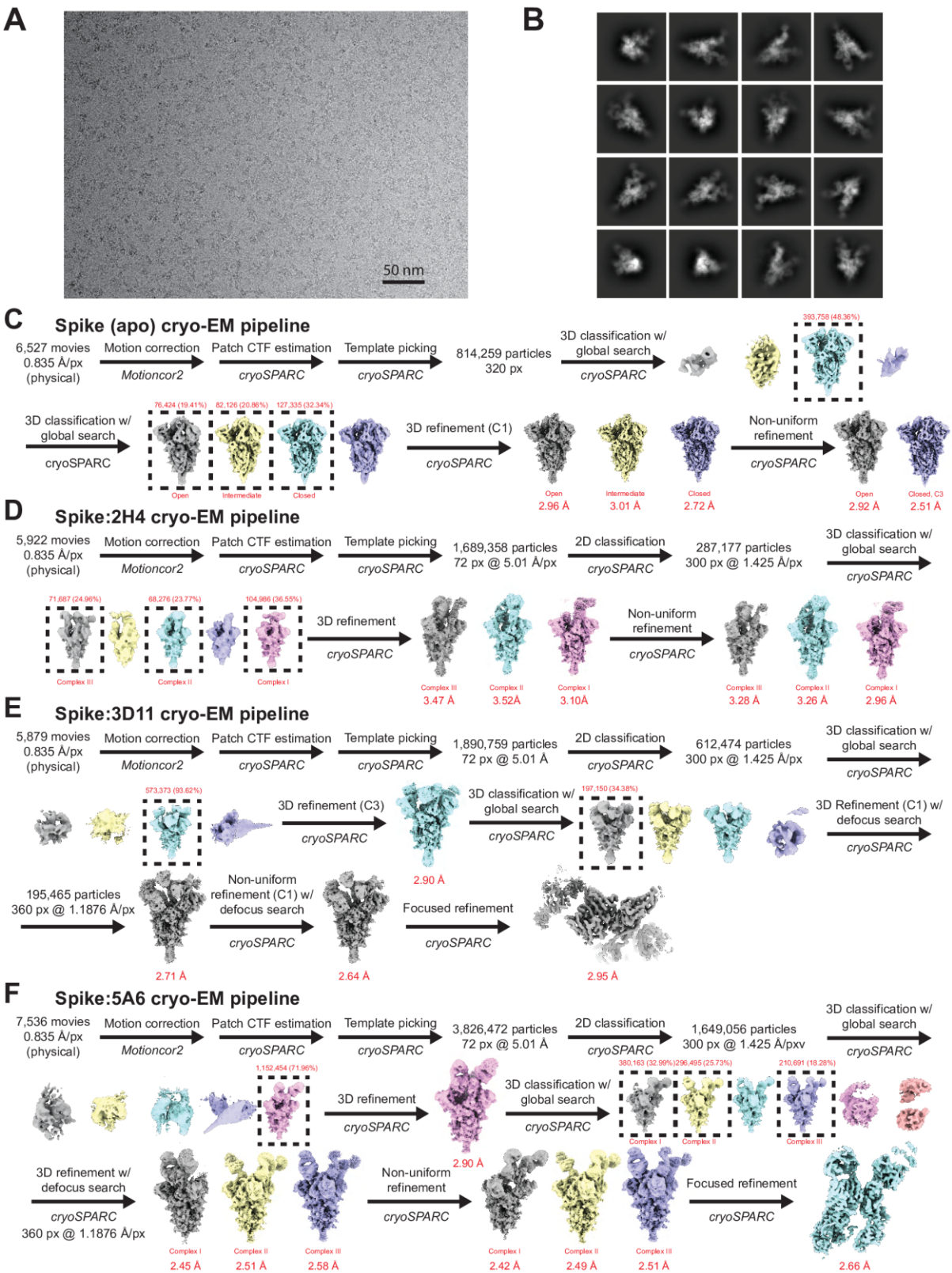


Figure 2.11 Cryo-EM processing

(A) A micrograph drawn from the Spike:5A6 complex dataset, representative of those obtained for the Fab complexes.

(B) Selected 2D class averages of Spike:5A6 particles, evincing clear secondary structure and multiple Fabs bound to the RBDs.
(C) Cryo-EM image processing workflow for Spike alone, leading to structures of the trimer with all RBDs closed, and with one RBD open or in an intermediate state.
(D) Processing workflow for the Spike:2H4 complex, resulting in structures with one, two, or three Fabs bound, as presented in the text.
(E) Processing workflow for the Spike:3D11 complex, culminating in a high-resolution structure of the Fab and RBD epitope.
(F) Processing workflow for the Spike:5A6 complex, resulting in the complexes presented in the text. Note the fourth class found during the second 3D classification step is a lower resolution duplicate of Spike:5A6 complex III. Processing culminates in a high-resolution of the Fab and its quaternary epitope involving two RBDs.

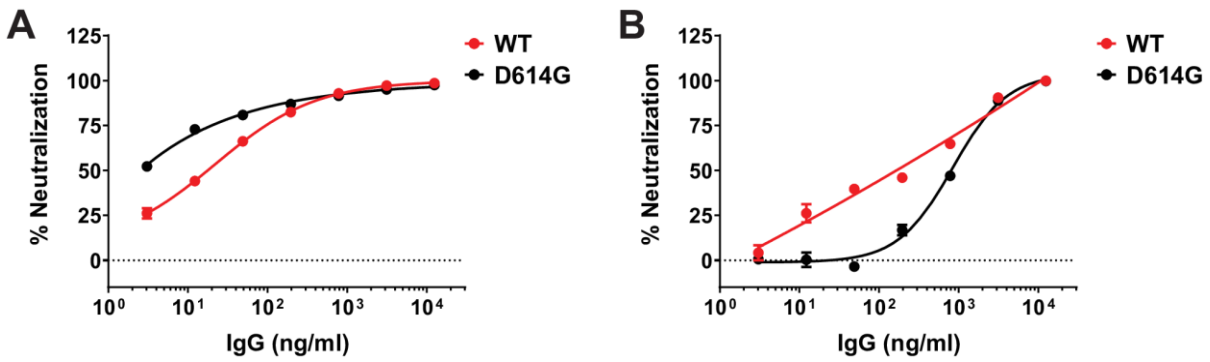


Figure 2.12 Neutralization of the SARS-CoV-2 pseudovirus with Spike mutant D614G

(A) Neutralization by 5A6 of SARS-CoV-2 pseudovirus bearing either wild-type Spike protein (red), or Spike with the D614G mutation (black). The efficacy of 5A6 IgG is against D614G mutant Spike is improved over wild-type.

(B) Neutralization by 3D11 of SARS-CoV-2 pseudovirus bearing either wild-type Spike protein (red), or Spike with the D614G mutation (black). 3D11 IgG suffers a severe loss of efficacy against the mutant pseudovirus (more than 5-fold weaker IC_{50}). Data are presented as mean ± SEM in triplicates and are representative of two independent experiments. IC50 was calculated by variable slope four parameter non-linear regression model using Graphpad PRISM 7 Software without or ^with top and bottom constraints set at 100% and 0% respectively.

Tables

Table 2.1 Percent framework identity to the germline sequences

mAb	VH (% identity)	VL (% identity)
1F4	IGHV6-1*01 (100.0)	IGLV1-47*01 (87.2)
2H4	IGHV1-2*06 (88.8)	IGKV5-2*01 (82.3)
3D11	IGHV4-34*01 (100.0)	IGLV6-57*02 (97.5)
3F11	IGHV1-18*01 (92.5)	IGKV3-20*01 (86.1)
5A6	IGHV3-30*01 (92.5)	IGKV1-39*01 (97.5)
6F8	IGHV3-66*01 (98.8)	IGKV3-20*01 (94.9)

VH and VL % identity refers to variable region compared to germline (IMGT).

Table 2.2 Binding kinetics of antibodies and Fabs measured against Fc-RBD by BLI or Spike trimer by SPR

mAb		k_a RBD ($M^{-1}s^{-1}$)	k_d RBD (s^{-1})	K_D RBD	k_a trimer ($M^{-1}s^{-1}$)	k_d trimer (s^{-1})	K_D trimer
1F4	Fab	2.14E+05	1.70E-02	8.02E-08	4.432E+05	3.046E-02	6.874E-08
	IgG	2.50E+05	<1.0E-07	<1.00E-12	-	-	-
2H4	Fab	3.58E+05	1.66E-02	4.72E-08	7.130E+05	3.340E-02	4.536E-08
	IgG	4.38E+05	6.10E-05	1.39E-10	2.042E+06	4.264E-04	2.088E-10
3D11	Fab	2.99E+05	4.90E-04	1.64E-09	5.098E+05	2.063E-04	4.047E-10
	IgG	3.44E+05	<1.0E-07	<1.00E-12	1.164E+06	2.173E-05	1.867E-11
3F11	Fab	7.68E+05	9.20E-02	1.20E-07	8.776E+05	1.775E-01	2.023E-07
	IgG	4.78E+05	1.23E-04	2.73E-10	8.487E+06	2.250E-03	2.651E-10
5A6	Fab	4.69E+05	3.57E-03	7.61E-09	8.557E+05	6.011E-03	7.025E-09
	IgG	3.39E+05	5.92E-05	1.73E-10	2.029E+06	9.772E-05	4.816E-11
6F8	Fab	-	-	-	-	-	-
	IgG	1.83E+05	8.76E-05	4.76E-10	3.617E+05	4.441E-04	1.228E-09
ACE2-Fc		1.43E+05	<1.0E-07	1.00E-12	3.751E+05	3.637E-04	9.699E-10

Table 2.3 Cryo-EM data collection, refinement and validation statistics

	Spike:5A6 complex I (EMDB-22993) (PDB 7KQB)	Spike:5A6 Fab and RBDs only	Spike:3D11 complex (EMD-22997) (PDB 7KQE)	Spike:3D11 Fab and RBD only	Spike:2H4 complex Rigid body docking (EMDB-22994)
Data collection and processing					
Magnification	EFTEM 105,000		EFTEM 105,000		EFTEM 105,000
Voltage (kV)	300		300		300
Electron exposure (e–/Å^2)	66.7		66.7		66.7
Defocus range (μm)	-0.5-1.5		-0.5-1.5		-0.5-1.5
Pixel size (Å)	0.835		0.835		0.835
Symmetry imposed	C1		C1		C1
Initial particle images (no.)	3,826,472		1,890,759		1,689,358
Final particle images (no.)	380,163		195,465		104,986
Map resolution (Å) FSC threshold 0.143	2.42	2.66	2.64	2.95	3.38
Map resolution range (Å)					
Refinement					
Initial model used (PDB code)	7a94	7a94	7a98	7a98	7a95
Model resolution (Å) FSC threshold 0.5	3.9	3.0	8.3	3.1	
Model resolution range (Å)					
Map sharpening *B* factor (Å^2)	-74.9	-87.7	-76.3	-57.1	-58.0
Model composition					
Non-hydrogen atoms	32,217	6,590	34,626	4,870	34,977
Protein residues	4,105	852	4,467	635	4511
Ligands	0	0	0	0	0
B factors (Å^2)					
Protein (min/max/mean)	1.84/281.53/94.83	1.84/115.96/2	0/609.22/243.17	0/609.22/185	0/754.58/163.41
Ligand	-/-/-	5.03 -/-/-	-/-/-	.68 -/-/-	-/-/-
R.m.s. deviations					
Bond lengths (Å)	0.012	0.012	0.012	0.012	0.017
Bond angles (°)	1.916	2.069	1.805	2.039	2.003
Validation					
MolProbity score	0.87	1.23	1.03	1.02	2.61
Clashscore	0.09	0.77	0.2	0.21	27.30
Poor rotamers (%)	0.98	1.22	1.08	1.28	2.04
Ramachandran plot					
Favored (%)	95.27	93.72	93.59	94.75	93.02
Allowed (%)	4.65	6.16	6.3	4.93	6.08
Disallowed (%)	0.07	0	0.11	0.32	0.90

Table 2.4 Key Resources

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
1A5	Laboratory of Cheng-I Wang	N/A
1A8	Laboratory of Cheng-I Wang	N/A
1B2	Laboratory of Cheng-I Wang	N/A
1B11	Laboratory of Cheng-I Wang	N/A
1C2	Laboratory of Cheng-I Wang	N/A
1C3	Laboratory of Cheng-I Wang	N/A
1D12	Laboratory of Cheng-I Wang	N/A
1E5	Laboratory of Cheng-I Wang	N/A
1F4	Laboratory of Cheng-I Wang	N/A
1H7	Laboratory of Cheng-I Wang	N/A
2C9	Laboratory of Cheng-I Wang	N/A
2C10	Laboratory of Cheng-I Wang	N/A
2C12	Laboratory of Cheng-I Wang	N/A
2D3	Laboratory of Cheng-I Wang	N/A
2G7	Laboratory of Cheng-I Wang	N/A
2H4	Laboratory of Cheng-I Wang	N/A
3A11	Laboratory of Cheng-I Wang	N/A
3C5	Laboratory of Cheng-I Wang	N/A
3D2	Laboratory of Cheng-I Wang	N/A
3D11	Laboratory of Cheng-I Wang	N/A
3E9	Laboratory of Cheng-I Wang	N/A
3F1	Laboratory of Cheng-I Wang	N/A
3F11	Laboratory of Cheng-I Wang	N/A
3H7	Laboratory of Cheng-I Wang	N/A
3H11	Laboratory of Cheng-I Wang	N/A
5A6	Laboratory of Cheng-I Wang	N/A
6F8	Laboratory of Cheng-I Wang	N/A
HRP conjugated anti-human Fc antibody	JACKSON ImmunoResearch	Cat#109-036-097
HRP conjugated anti-human Fab antibody	JACKSON ImmunoResearch	Cat#109-036-097
Bacterial and virus strains		
SARS-CoV-2 hCoV-19/Singapore/3/2020	Laboratory of Brendon John Hanson	EPI_ISL_407988
SARS-CoV-2 BetaCoV/France/IDF0571/2020	Laboratory of Dr. Manuel Rosa-Calatrava; Pizzorno et al. 2020	EPI_ISL_411218
Chemicals, peptides, and recombinant proteins		
SARS-CoV-2 Spike protein RBD-mFc	Sino Biological	Cat#40592-V05H
human ACE2-Fc	This paper	N/A
Critical commercial assays		
Lenti-X p24 rapid titer kit	Takara Bio	Cat#632500
Luciferase Assay System	Promega	Cat#E1510
Viral ToxGlo™ Assay	Promega	Cat#G8941
TMB Substrate	Surmodics, BioFX®	Cat#TMBW-1000-01

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: cell lines		
ExpiCHO-S™	Thermo Fisher Scientific	Cat#A29127
293T	ATCC	Cat#CRL-3216
Vero E6 C1008	ATCC	Cat#CRL-1586
CHO-ACE2	Laboratory of Yee-Joo Tan	N/A
MucilAir Nasal HAE	Epithelix SARL	Ref: EP02MP
MucilAir Bronchial HAE	Epithelix SARL	Ref: EP01MD
Oligonucleotides		
Forward primer HKU-ORF1b-nsp14F: 50-TGGGGYTTTACRGGTAACCT-30	Eurogentec	N/A
Reverse primer HKU-ORF1b-nsp14R: 50-AACRCGCTTAACAAAGCACTC-30	Eurogentec	N/A
Probe HKU-ORF1b-nsp141P: 50-FAMTAGTTGTGATGCWATCATGACTAG-TAMRA-30	Eurogentec	N/A
Recombinant DNA		
Plasmid: pMDLg/pRRE	Laboratory of Didier Trono	Addgene ID: 12251
Plasmid: pRSV-Rev	Laboratory of Didier Trono	Addgene ID: 12253
Plasmid: pHIV-luc-ZsGreen	Laboratory of Bryan Welm	Addgene ID: 39196
Plasmid: pTT5LnX-CoV-SP (codon optimized S gene of SARS-CoV-2, GenBank: YP_009724390.1)	Laboratory of Brendon John Hanson	N/A
Plasmid: pTT5LnX-CoV-SP-D614G	This paper	N/A
Software and algorithms		
GraphPad Prism version 7.03	GraphPad	https://www.graphpad.com/scientific-software/prism/
Quest Graph IC50 calculator AAT	AAT Bioquest	https://www.aatbio.com/tools/ic50-calculator
Octet System Data Acquisition Software 9.0.0.4.	ForteBio	N/A
Biacore T200 Evaluation Software	Cytiva Life Sciences	N/A
GraphPad Prism version 7.03	GraphPad	https://www.graphpad.com/scientific-software/prism/
cryoSPARC v2.7-3.1	Structura Biotechnology	https://cryosparc.com/
Relion 3.1	Laboratory of Molecular Biology, Medical Research Council	https://github.com/3dem/relion
UCSF ChimeraX 1.1	UCSF	https://www.cgl.ucsf.edu/chimerax/
ISOLDE 1.1.0	Cambridge Institute for Medical Research	https://isolde.cimr.cam.ac.uk/
Coot 0.9.4.1	Laboratory of Molecular Biology, Medical Research Council	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Rosetta 3.12	RosettaCommons	https://www.rosettacommons.org/

Supplemental Files

Movie 2.1 Tour of Spike:5A6 Complex I
First, the density map for Spike:5A6 complex I is displayed using a low iso-surface threshold, highlighting the binding geometry. The Spike is shown with RBDs in coral, NTDs in plum, the S2 core in rose brown, and 5A6 Fabs is shown in goldenrod. Next, the map is shown in transparency overlaid on a colored copy of the map with a higher density threshold, revealing high resolution details. The high threshold map is replaced by a ribbon diagram of the complex, employing the same color scheme. Finally, the ribbon diagram is shown embedded in the low resolution Spike:5A6 IgG density.

Movie 2.2 3D Variability Analysis of Spike:2H4 complex
3DVA resolves principal components of the variability in single-particle EM data. The movie illustrates the first 3DVA component obtained when data for all Spike:2H4 complexes are processed together. At one extreme of this axis of variability, corresponding to the beginning of the movie, the RBD farthest from the camera is closed, with no Fab density. Its counter-clockwise neighbor (to the left) has a superposition of density for both closed and open conformations. A scan across the 3DVA component then shows that as a 2H4 Fab binds to the closed, far RBD, the neighboring RBD is forced into a super-open conformation, and density for the closed state disappears entirely. Simultaneously, the NTD adjacent to this RBD (left side of frame) relaxes outwards and partially reaveals the S2 subunit.

Movie 2.3 Morph between 5A6 complexed with closed or open RBDs
The structure of Spike:5A6 Complex I was used to model the transition from closed to open RBD conformations when 5A6 is bound to two adjacent Spike monomers. Monitoring the distance between the C-terminii of the two 5A6 heavy chains throughout the animation reveals that the Fabs are closest in a putative intermediate RBD conformation. The nearest approach is approximately 80 Å, a gap that can likely be spanned via the flexibility of Fab elbows and the IgG hinge region.

Methods

Experimental models and subject details

Cells

The human embryonic kidney epithelial cell 293T (ATCC, Cat#CRL-3216) and Vero E6 cells (ATCC CRL-1586) were cultured in Dulbecco's modified Eagle's medium (Hyclone, Cat#SH30022.01) supplemented with 10% heat-inactivated FBS (Gibco, Cat#10270-106) and 100 U/ml Penicillin-Streptomycin (Gibco, Cat#15140-122). A stable cell line expressing human ACE2, CHO-ACE2 (a kind gift from Professor Yee-Joo Tan, IMCB, A*Star) [52] was maintained in Dulbecco's modified Eagle's medium supplemented with 10% heat-inactivated FBS, 1% MEM Non-Essential Amino Acids Solution (Gibco, Cat#11140-050) and 0.5 mg/ml of Geneticin™ Selective Antibiotic (Gibco, Cat#10131-027). All cells were maintained at 37°C and 5% CO_2 and were passaged by dissociating the cells with StemPro™ Accutase™ Cell Dissociation Reagent (Gibco, Cat#A1110501) every 2 to 3 days.

HAE

MucilAir™ HAE (human airway epithelia) reconstituted from human primary cells obtained from nasal or bronchial biopsies were provided by Epithelix SARL (Geneva, Switzerland) and maintained in air-liquid interphase with specific culture medium in Costar Transwell inserts (Corning, NY, USA) according to the manufacturer's instructions.

Virus

The Singapore strain of SARS-CoV-2 live virus used in this study was isolated from a nasopharyngeal swab of a patient in Singapore [33]. The French strain of SARS-CoV-2 live virus used in this study was isolated from one of the first COVID-19 cases confirmed in France: a 47-year old female patient hospitalized in January 2020 in the Department of Infectious and Tropical

Diseases, Bichat Claude Bernard Hospital, Paris [53]. The complete viral genome sequence was obtained using Illumina MiSeq sequencing technology, was then deposited after assembly on the GISAID EpiCoV platform (Accession ID EPI_ISL_411218) under the name BetaCoV/France/IDF0571/2020.

Antibody discovery from phage display library

Anti-SARS-CoV-2 Spike RBD antibodies were isolated from an HX02 human Fab phage display library (Humanyx Pte Ltd) via *in vitro* selection. Briefly, biopanning was performed using SARS-CoV-2 RBD (YP_009724390.1) (Arg319-Phe541) with a mouse Fc tag (Sino Biological, Cat#40592-V05H) biotinylated using the EZ-Link NHS-PEG4-Biotin labelling kit (Thermo Fisher Scientific, Cat#A39259). In both rounds of biopanning, biotinylated SARS-CoV-2 RBD-mFc protein (Sino Biological, Cat#40592-V05H) was immobilized on M280 streptavidin-coated magnetic beads (Life Technologies, Cat#11205D); 3.5 x 10^{12} cfu phage in 1ml 1% casein-PBS blocking buffer was used in the first round, and 1.64 x 10^{11} cfu phage were used in the second round. During the biopanning process, binders to mouse Fc were removed by pre-incubation of phage with 2 μM mouse IgG before mixing with the RBD-mFc antigen. After two rounds of biopanning, the Fabs of selected clones were expressed in *E. coli* HB2151 cells (Stratagene) to screen for RBD binders by ELISA. Unique clones were identified by DNA sequencing.

IgG expression and purification

Fabs were reformatted into human IgG in the pTT5 vector (National Research Council of Canada) and the IgG antibodies were expressed using ExpiCHO expression system (Thermo Fisher Scientific) by transient co-transfection of plasmids expressing the heavy and light chain of each antibody clone. Eight days after transfection, ExpiCHO-S cell suspension was centrifuged

for 10 min at 2000 rpm and filtered with 0.22 μm filter to remove the cells and debris. Antibodies were then purified from the culture supernatant using Protein G Agarose (Merck Millipore, Cat#16-266) following the manufacturer's instructions. After elution, the purified antibodies were dialyzed at 4°C for 4-20 hours against 1x PBS, for 3 times and concentrated to 1-2 mg/ml using 10MWCO Vivaspin 20 (Sartorius, Cat#VS2001).

Fab production and purification

The tag-less Fab fragments were produced using the ExpiCHO transient expression system. Eight days after transfection, ExpiCHO-S cell suspension was centrifuged and filtered; and Fab was purified from the filtered culture supernatant using cation exchange chromatography (CIEX) on AKTA FPLC System (GE Healthcare). In brief, the supernatant was concentrated to 2 ml using 10MWCO Vivaspin 20 (Sartorius), diluted 1:20 in Buffer A (20 mM Sodium Acetate, pH 5.2), filtered through 0.22 μm filter, and loaded onto Mono-S 5/50 GL column at a flow rate of 1ml/min. Fab fragments were eluted in Buffer B (20 mM Sodium Acetate, pH 5.2 with 1 M Sodium Chloride) with a sequential linear gradient of 0% to 5% in 5 min, 5% to 15% in 30 min, and 15% to 100% in 20 min of Buffer B injection at a flow rate of 1 ml/min. The resulting purified Fab fragments were dialyzed at 4°C for 4-20 hours against 4 liters of 20 mM Histidine, 150 mM NaCl, pH 6.6, for 3 times and concentrated to 1-2 mg/ml using 10MWCO Vivaspin 6 (Sartorius, Cat#VS0601).

Avidity binding ELISA to RBD proteins

Anti-SARS-CoV-2 Spike RBD IgG antibodies were tested in an ELISA against biotinylated recombinant SARS-CoV-2 Spike protein RBD-mFc (Sino Biological, Cat#40592-V05H) or SARS-CoV Spike protein RBD-His (Sino Biological, Cat#40150-V08B2) to assess binding avidity for the target. In brief, NeutrAvidin protein (Thermo Fisher Scientific, Cat#31000)

was coated at 5 μg/ml onto 96-well ELISA plates in coating buffer (8.4 g/L $NaHCO_3$, 3.56 g/L Na_2CO_3, pH 9.5) overnight at 4°C. After blocking with 1% Casein (Thermo Fisher Scientific, Cat#A37528) for two hours, biotinylated antigen at 0.2 μg/ml was added to the plates and captured by NeutrAvidin during one-hour incubation at room temperature. After washing with 0.05% PBST for 5 times, the IgG antibodies were added at different concentrations with 3-fold dilutions in triplicate and incubated for one hour. The wells were then washed again with 0.05% PBST, followed by addition of HRP conjugated anti-human Fc antibody (1:3000, JACKSON ImmunoResearch, Cat#109-036-098). Finally, the wells were washed and the HRP activity was measured at 450 nm with addition of 3,3',5,5'-tetramethylbenzidine (TMB) substrate (Surmodics, BioFX®, Cat#TMBW-1000-01).

Avidity binding ELISA to purified pseudovirus

Anti-SARS-CoV-2 Spike RBD IgG or Fab antibodies were tested in an ELISA against iodixanol-gradient-purified SARS-CoV-2 pseudoviruses with an isotype IgG used as a negative control antibody. In brief, 1 μg/ml of pseudoviral particles were coated in coating buffer onto 96-well ELISA plates overnight at 4°C. After blocking with 1% Casein (Thermo Fisher Scientific, #A37528) for two hours, serially diluted IgG or Fab antibodies starting from 20 nM (IgG) or 300 nM (Fab) with five-fold dilutions were added to the plates and incubated for an hour at room temperature. The wells were then washed again with 0.05% PBST, followed by addition of HRP conjugated anti-human Fc antibody (JACKSON ImmunoResearch, Cat#109-036-098) or HRP conjugated anti-human Fab antibody (JACKSON ImmunoResearch, Cat#109-036-097) for one-hour incubation before HRP activity was measured at 450 nm with addition of TMB substrate (Surmodics, BioFX®, Cat#TMBW-1000-01).

Competition ELISA

Anti-SARS-CoV-2 Spike RBD IgG antibodies were tested in a competition ELISA to assess their ability to block the Spike protein RBD from binding to human ACE2 protein. In brief, the recombinant human ACE2 protein with a human Fc tag (ACE2-Fc) was coated onto the 96-well ELISA plates in coating buffer overnight at 4°C and blocked with 1% Casein. Then different concentrations of anti-SARS-CoV-2 Spike RBD IgG antibodies were pre-incubated with 0.5 nM biotinylated Spike protein RBD-mFc (Sino Biological, Cat#40592-V05H) for one hour at room temperature before they were added to the ELISA plates coated with ACE2-Fc. After one-hour incubation, the wells were washed with 0.05% PBST for five times and HRP conjugated streptavidin (Biolegend, Cat#405210) was added at a dilution of 1:3000, and incubated for another one hour before HRP activity was measured at 450 nm with addition of TMB substrate (Surmodics, BioFX®, Cat#TMBW-1000-01).

Conversion of IgG antibodies to Fab fragments

Digestion reaction for each IgG was prepared using immobilized FabALACTICA microspin columns (Genovis). 100 μL of IgG at 5 mg/mL concentration in digestion buffer (150 mM sodium phosphate, pH 7.0) were incubated overnight on each column. Digested sample was further purified using a HiTrap Protein L column followed by size-exclusion chromatography (Superdex 75 10/300 GL) using an Äkta Pure FPLC (GE Healthcare).

Soluble SARS-CoV-2 Spike production

The expression plasmid containing the prefusion S ectodomain as used in Wrapp, et al. [41] was kindly provided by Prof. Jason McLellan (University of Texas at Austin). This construct was used to transiently transfect high-density Chinese Hamster Ovary (ExpiCHO) cells with ExpiFectamine per the “Max Titer” protocol provided (Thermo Fisher). Six days post-transfection,

0.2 µm-filtered supernatant was collected and incubated with Ni-Sepharose Excel (Cytiva Life Sciences) for batch purification. Eluate was collected, concentrated in a 50 MWCO Amicon Ultra-15 centrifugal filter unit (Millipore Sigma), and injected onto a Superose6 10/300 GL column equilibrated in 10 mM HEPES, 200 mM NaCl, pH 8.0 to isolate trimeric, monodisperse material for Fab/IgG complexing.

Cryo-EM sample preparation and imaging

2.5 µL of Spike-Fab complex at a concentration of 0.4 mg/mL was applied to a 300 mesh gold Quantifoil 1.2/1.3 holey carbon grid that was glow discharged for 30 seconds at 15 mA immediately before sample application. Grids were blotted using Whatman #1 filter paper for 8 or 10 seconds at a blot force of 0 at 4°C and 100% humidity using a Mark IV Vitrobot (Thermo Fisher Scientific) and plunge frozen into liquid ethane. Samples were loaded onto a Titan Krios transmission electron microscope (Thermo Fisher Scientific) equipped with a Gatan K3 direct electron detector (Gatan) and a Quantum GIF energy filter (Gatan) operated with a 20 eV slit width during image acquisition. The K3 camera was operated in CDS mode using super resolution. A nominal magnification of 105,000x was used, for a pixel size of 0.835 Å (0.4175 Å super resolution pixel size) at the sample. A dose rate of 8 $e^-/(pix \cdot sec)$, or 11.5 $e^-/(Å^2 \cdot sec)$, and a frame rate of 0.05 sec/frame was used with a total exposure time of 5.9 sec, for a total dose of 67.7 $e^-/Å^2$. Automated data collection was performed using SerialEM [54].

Image processing

Dose-weighted, motion-corrected sums down-sampled to the physical pixel size were obtained from the super-resolution DED movies using UCSF Motioncor2 [55]. For the Spike trimer, CTF estimation was performed in cryoSPARC [56] followed by blob-based particle picking, 2D classification, *ab initio* modelling, 3D classification, and 3D refinement. For images

of antibody complexes, particles were instead picked using templates generated from the apo trimer structure, and the apo trimer was likewise used as an initial model in 3D classification. The resolution of the interface between the Spike RBD and the 5A6 Fab was further improved using naïve focused refinements. Additional 3D classification of the Spike:3D11 complex was performed in Relion 3.1 [57]. Processing details are given in **Table 3.3** and **Figure 2.11**.

Molecular modelling

For each Spike:Fab complex, a previously determined structure of the SARS-CoV-2 Spike protein, along with a full-length Fab homology model computed by MODELLER [58], were simultaneously docked into cryo-EM density using UCSF Chimera [59]. Spike with one open RBD and one copy of ACE2 bound (PDB: 7a94) was used with 5A6, Spike with three open RBDs and three copies of ACE2 bound (PDB: 7a98) were used with 3D11, and Spike with two open RBDs and one copy of ACE2 bound (PDB: 7a95) was used with 2H4. Missing segments and side chains in the RBDs were built using Coot. Interactive, density-restrained molecular dynamics simulations in ChimeraX [60] and ISOLDE [61] were used to finalize the models, and atomic b-factors were calculated using Rosetta [62]. Models for the Spike:5A6 and Spike:3D11 complexes were first built into density maps from whole-particle cryo-EM reconstructions, and then further refined using maps from focused refinements of the Fab and Spike RBD. Maps of Spike:2H4 complexes were of lower resolution and model building was terminated after the docking step described above. Model statistics and density fit information are presented in **Table 3.3** and **Figure 2.9**.

Generation of pseudovirus particles

Pseudotyped viral particles expressing SARS-CoV-2 Spike protein were produced by transfecting of 30 million 293T cells with 12 μg pMDLg/pRRE (a gift from Didier Trono, Addgene

#12251), 6 µg pRSV-Rev (a gift from Didier Trono, Addgene #12253), 24 µg pHIV-Luc-ZsGreen (a gift from Bryan Welm, Addgene #39196) and 12 µg pTT5LnX-CoV-SP (expressing SARS-CoV-2 Spike protein, Genbank: YP_009724390.1, a kind gift from DSO National Laboratories) using Lipofectamine 2000 transfection reagent (Invitrogen, Cat#11668-019). The transfected cells were cultured at 37°C incubator for 3 days. Viral supernatant was harvested, centrifuged at 700 g for 10min to remove cell debris and filtered through a 0.45 µm filter unit (Sartorius, Cat#16555). Lenti-X p24 rapid titer kit (Takara Bio, Cat#632200) was used to quantify the viral titres following the manufacturer instructions. pTT5LnX-CoV-SP plasmid with D614G mutation was generated using QuickChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent, Cat#210513) and was used to generate mutant pseudovirus expressing SARS-CoV-2 Spike protein carrying D614G mutation.

Purification of pseudovirus particles

To concentrate and purify the pseudovirus particles expressing the SARS-CoV-2 Spike glycoproteins, pre-cleared 40 mL viral supernatant was concentrated by 20% sucrose gradient centrifugation at 10,000 g for 4 hours at 4°C in an SW41 Ti rotor with no brake. Upon removal of supernatant, 1mL of PBS was added to the virus pellet and left at 4°C overnight. Concentrated virus was further purified by an OptiPrep (60% [wt/vol] iodixanol, STEMCELL Technologies, Cat#07820;) velocity gradient. Iodixanol gradients were prepared in PBS in 1.2% increments ranging from 6 to 18%. Pseudoviruses were layered onto the top of the gradient and centrifuged for 1.5 hours at 200,000 g in an SW41 Ti rotor. Gradient fraction that contained pseudovirus pellet was collected.

Pseudovirus neutralization assay

CHO-ACE2 cells were seeded at a density of 3.2 x 10^4 cells in 100 µL of complete medium without Geneticin in 96-well Flat Clear Bottom Black Polystyrene TC-treated Microplates (Corning, Cat#3904). Serially diluted IgG or Fab antibodies were incubated in a 96-well flat-bottom cell culture plate (Costar, Cat#3596) with an equal volume of pseudovirus (12 ng of p24) at the final volume of 50 µL at 37°C for one hour, and the mixture was added to the monolayer of pre-seeded CHO-ACE2 cells in triplicate. After one hour of pseudovirus infection at 37°C, 150 µl of culture medium was added to each well and the cells were further incubated for another 48 hours. Upon removal of culture medium, cells were washed twice with sterile PBS, and then lysed in 20 µL of 1x Passive lysis buffer (Promega, Cat#E1941) with gentle shaking at 37°C for 30 minutes. Luciferase activity was then assessed using a Luciferase Assay System (Promega, Cat#E1510) on a Promega GloMax Luminometer. The relative luciferase units (RLU) were converted to percent neutralization and plotted with a non-linear regression curve fit using GraphPad PRISM.

Live virus neutralization assay in Vero E6 cells

The potency of the IgG or Fab antibodies were determined in neutralizing live SARS-CoV-2 virus assays. In brief, 25 µl of 100 $TCID_{50}$ of SARS-CoV-2 live virus (hCoV-19/Singapore/3/2020) was mixed with an equal volume of serially diluted IgG or Fab antibodies and incubated at 37°C for one hour before the mixture was added to 50 µl of Vero E6 C1008 cells in suspension. The infected cells were incubated at 37°C incubator for four days and the cell viability was determined using Viral ToxGlo™ Assay (Promega, Cat#G8941). The potency of 2H4, 3D11 and 5A6 IgG antibodies in neutralizing live SARS-CoV-2 virus assays was also determined by measuring the viral genome copy number (GCN). 25 µl of 100 TCID50 of SARS-

CoV-2 live virus (hCoV-19/Singapore/3/2020) was mixed with an equal volume of serially diluted 2H4, 3D11 or 5A6 IgG antibodies and incubated at 37°C for one hour before the mixture was added to 50 µl of $4x10^5$ Vero E6 C1008 cells in suspension. The infected cells were incubated at 37°C incubator for 48 hrs after which supernatant was harvested and viral GCN was determined by subsequent RT-qPCR targeting the E gene using the RESOLUTE 2.0 kit as per manufacturer's instructions. Briefly, 2.5 µl of supernatant was diluted with 2.5 µl of Milli-Q water and added to 20 µl of RT-PCR master mix. PCR was carried out as follows: reverse transcription at 55°C for 15min, inactivation at 95°C for 4min, followed by 45 cycles of amplification consisting of denaturation at 95°C for 3 sec and annealing/extension at 62°C for 30 sec and GCN values determined by comparing Ct values against a logGCN standard curve.

Live virus neutralization assay in HAE

The potency of 5A6 IgG was tested in neutralizing a live virus strain (BetaCoV/France/IDF0571/2020) using MucilAir^{TM} HAE (human airway epithelia) model. Briefly, the apical poles of HAE were gently washed twice with warm Opti-MEM medium (Gibco, Thermo Fisher Scientific) and then infected directly with a 150 µl dilution of live SARS-CoV-2 virus strain (BetaCoV/France/IDF0571/2020) in Opti-MEM medium, at a multiplicity of infection (MOI) of 0.1. Viral suspensions were pre-incubated for 60 min with antibody 5A6 IgG (75 ng/ml or 150 ng/ml) or an anti-Ebola glycoprotein control antibody (150 ng/ml) before infection. A control infection was performed in absence of antibody. For mock infection, the same procedure was followed using Opti-MEM as inoculum. Viral replication was quantified as the measured copy number of the viral genomes inside, and at the apical poles of, nasal and bronchial HAE. Samples collected from apical washes at 48 hours post-infection were separated into 2 tubes: one for TCID50 viral titration (stored at -80°C) and one for RT-qPCR. HAE cells were harvested in RLT

buffer (Qiagen) and total RNA was extracted using the RNeasy Mini Kit (Qiagen) for subsequent RT-qPCR. Variations in trans epithelial electrical resistance (Δ TEER) were measured using a dedicated volt-ohm meter (EVOM2, Epithelial Volt/Ohm Meter for TEER) and expressed as Ohm/cm^2.

Cell-cell fusion assay

Vero E6 cells were transfected with S protein bearing furin recognition mutation (R682RAR to A682AAR) with C-terminal GFP tag by Lipofectamin 2000 (Invitrogen, Cat#11668-019) and were cultured on μ-Slide 8 well chamber slides (ibidi,). The transfection efficiency was monitored by percentage of GFP positive cells and optimized within 15-30% to achieve the best signal-to-noise ratio in the following cell-cell fusion assay. After 48 hours, cells were treated with various antibodies diluted in DMEM without FBS for 1 hour at 37°C. Cells were then treated with 15 μg/ml trypsin and incubated at 37°C for another 2 hours. After trypsin treatment, cells were fixed with 4% PFA at room temperature for 15 mins and the cell nuclei were stained with DAPI. Images were taken by Olympus confocal microscope.

Fab affinity measurement by BLI

Binding affinity of purified Fab to RBD was measured on the Octet96Red system (ForteBio). Anti-human IgG Fc (AHC) sensors were first loaded with 1 μg/ml of Fc-RBD for 10 min, followed by kinetics buffer (phosphate-buffered saline buffer supplemented with 0.1% Tween-20 and 0.1% BSA) for 5 min to establish a stable baseline. The sensors were then dipped into different concentrations of each Fab from 100 nM to 3.125 nM in two-fold dilutions for 6 min, and then in kinetics buffer again for 10 min to measure association and dissociation. Assays were run at 25°C and data was analysed on the Octet System Data Acquisition Software version 9.0.0.4. using the 1:1 Langmuir binding model.

Avidity binding by BLI

Avidity of anti-SARS-CoV-2 Spike RBD IgG antibodies for RBD was measured on the Octet96Red system. Anti-hIgG Fc capture (AHC) sensors were used. The sensors were loaded with 1 μg/ml of Fc-RBD (made in-house) in assay buffer (phosphate-buffered saline buffer supplemented with 0.1% Tween-20 and 0.1% BSA) for 10 min, quenched in 0.5 mg/ml of isotype IgG in assay buffer for 10 min, then dipped in assay buffer for 12 min for the system to stabilize. To measure the association of 5A6, the sensors were dipped in a range of 5A6 IgG concentrations (25-0.39 nM in 2-fold serial dilutions) in assay buffer for 6 min. To measure dissociation, the sensors were dipped in assay buffer for 10 min. The experiment was conducted at 25°C. Data analysis was done in the Octet System Data Acquisition Software version 9.0.0.4. using the 1:2 bivalent model.

Epitope binning by BLI

Epitope binning was done using a classical sandwich assay. The AR2G sensor tips (ForteBio) were activated in freshly prepared 20mM EDC (1-ethyl-3-[3-dimethylaminopropyl]-carbodiimide hydrochloride), 10mM NHS (N-hydroxysuccinimide) solution and the 5A6 antibody was immobilized to the sensor tips using a concentration of 7.5 μg/ml of 5A6 in 10 mM sodium acetate pH 6 buffer. After quenching in 1M ethanolamine, the 5A6-immobilized sensor tips were dipped in 5 μg/ml of tagless RBD for 600 seconds, then in 10μg/ml of the second antibody for 300 sec. The assay was run at 25°C. Sensor tips were regenerated in 10 mM glycine at pH 2.7 and neutralized in PBS with 0.1% Tween-20 before another cycle of sandwich assay was performed. Each sensor tip was used in a total of 3 cycles. Data analysis was done in the Octet System Data Acquisition Software version 9.0.0.4.

IgG and Fab affinity for Spike trimer by SPR

StreptagII-tagged prefusion S ectodomain, diluted to 10 µg/mL in 10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.05% PS-20, pH 7.4, was captured on a StreptactinXT-immobilized (Iba Life Sciences) CM5 Series S sensor chip at an average level of 224 or 347 RU (response units) for IgG and Fab kinetics measurements, respectively using a Biacore T200 (Cytiva Life Sciences). Two-fold serial dilutions of purified IgG from 12.5 nM to 0.39 nM or Fab from 100 nM to 3.125 nM were flowed over the captured prefusion S ectodomain at 30 µL/minute for 90 seconds followed by 420 seconds of dissociation flow. Following each cycle, the chip surface was regenerated with 3 M guanidine hydrochloride. The resulting reference flow cell and blank-injection subtracted sensorgrams were fit to a 1:1 Langmuir binding model using the Biacore T200 Evaluation Software (Cytiva Life Sciences).

Statistical analysis.

Data were analysed using GraphPad Prism version 7.03. Statistical tests are indicated in the figure legends. EC_{50} values were calculated by non-linear regression analysis on the binding curves using GraphPad Prism and IC_{50} values were calculated either using the [Inhibitor] vs response variable slope four parameter non-linear regression model of GraphPad Prism, or the four parameter logistic regression model in the Quest Graph™ IC_{50} Calculator from AAT Bioquest, Inc (https://www.aatbio.com/tools/ic50-calculator). One-way analysis of variance (ANOVA) was used to compare differences between groups. Differences were considered statistically significant at confidence levels $*P < 0.05$ or $**P < 0.01$, $***P < 0.001$.

Acknowledgements

We thank Professor Yee-Joo Tan (Department of Microbiology, NUS; Institute of Molecular and Cell Biology, A*STAR) who kindly provided CHO-ACE2 cells. We also thank Wong Pui San and Chye De Ho in the BSL 3 facility (DSO National Laboratories), as well as Dr. Pei Yin Lim and Ms. Siti Nazihah Mohd Salleh (Singapore Immunology Network) for their technical assistance. We thank Professor Jason McLellan (University of Texas at Austin) for the prefusion S ectodomain. We thank Professor Joseph DeRisi and Professor Robert Fletterick (UCSF) and Dr. John Pak (CZI Biohub, San Francisco) for helpful discussions. This work was supported by core research grants provided to Singapore Immunology Network of the Biomedical Research Council (BMRC), the A*CCLERATE GAP fund (ACCL/19-GAP064-R20H-I) from the Agency of Science, Technology and Research (A*STAR, Singapore). S.M.S. was supported by NMRC grant COVID19RF-004. Support was also provided by the National Institutes of Health Grant P50AI150476 (C.S.C., Y.C.), NIH Grants S10 S10OD020054 (Y.C.) and S10OD021741 (Y.C.). This work was supported by the UCSF COVID-19 Response Fund, a grant from Allen & Company, and supporters of the UCSF Program for Breakthrough Biomedical Research (PBBR), which was established with support from the Sandler Foundation (C.S.C., Y.C., A.M.). Additional support was provided by a Fast Grant from Emergent Ventures at the Mercatus Center, George Mason University (C.S.C.). Y.C. is an Investigator of the Howard Hughes Medical Institute. Additional support was provided by INSERM REACTing (REsearch & ACtion emergING infectious diseases), CNRS and Institut Mérieux (Mérieux Research grant) (A.P., O.T., M.R.C.).

Author Contributions

B.W., D.A., W.H.L., Y.H., C.W.H., B.F., P.M.L.N., E.Z.X.N., M.B., D.B., S.M.S., M.R.C., A.M., Y.C., C.S.C., and C.I.W. conceptualized and designed the study. L.R., B.J.H.,

M.R.C., A.M., Y.C., C.S.C., and C.I.W. coordinated the study. B.W., D.A., W.H.L., Y.H., C.W.H., B.F., P.M.L.N., E.Z.X.N., M.B., D.B., A.P., B.A., H.C.T., C.Y.L., R.A.M., A.P., B.A., O.T., M.K.S., F.J.T., Y.Y.C.Y., S.G.K.S., C.E.Z.C., E.C., and N.J.Y. conducted the experiments. B.W., D.A., W.H.L., Y.H., C.W.H., B.F., P.M.L.N., E.Z.X.N., M.B., D.B., H.C.T., C.Y.L., R.A.M., A.P., B.A., H.C.T., C.Y.L., R.A.M., O.T., S.G.K.S., C.E.Z.C., S.M.S, L.R., M.R.C., A.M., Y.C., C.S.C., and C.I.W. analysed and interpreted the data. D.A., B.W., W.H.L., Y.H., C.W.H., B.F., P.M.L.N., E.Z.X.N., M.B., D.B., A.P., O.T., C.E.Z.C., S.M.S., L.R., B.J.H., M.R.C., A.M., Y.C., C.S.C., and C.I.W. wrote and revised the manuscript. All authors read and approved the final manuscript.

References

[1] M. A. Tortorici and D. Veesler, “Chapter Four - Structural insights into coronavirus entry,” in *Advances in Virus Research*, vol. 105, F. A. Rey, Ed. Academic Press, 2019, pp. 93–116. doi: 10.1016/bs.aivir.2019.08.002.

[2] D. J. Benton *et al.*, “Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion,” *Nature*, vol. 588, no. 7837, Art. no. 7837, Dec. 2020, doi: 10.1038/s41586-020-2772-0.

[3] M. Hoffmann *et al.*, “SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor,” *Cell*, vol. 181, no. 2, pp. 271-280.e8, Apr. 2020, doi: 10.1016/j.cell.2020.02.052.

[4] A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veesler, “Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein,” *Cell*, vol. 181, no. 2, pp. 281-292.e6, Apr. 2020, doi: 10.1016/j.cell.2020.02.058.

[5] P. Zhou *et al.*, “A pneumonia outbreak associated with a new coronavirus of probable bat origin,” *Nature*, vol. 579, no. 7798, Art. no. 7798, Mar. 2020, doi: 10.1038/s41586-020-2012-7.

[6] J. Lan *et al.*, “Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor,” *Nature*, vol. 581, no. 7807, Art. no. 7807, May 2020, doi: 10.1038/s41586-020-2180-5.

[7] W. Tai *et al.*, “Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine,” *Cell. Mol. Immunol.*, vol. 17, no. 6, Art. no. 6, Jun. 2020, doi: 10.1038/s41423-020-0400-4.

[8] F. Musarrat *et al.*, "The anti-HIV drug nelfinavir mesylate (Viracept) is a potent inhibitor of cell fusion caused by the SARSCoV-2 spike (S) glycoprotein warranting further evaluation as an antiviral against COVID-19 infections," *J. Med. Virol.*, vol. 92, no. 10, pp. 2087–2095, 2020, doi: 10.1002/jmv.25985.

[9] A. Shulla, T. Heald-Sargent, G. Subramanya, J. Zhao, S. Perlman, and T. Gallagher, "A Transmembrane Serine Protease Is Linked to the Severe Acute Respiratory Syndrome Coronavirus Receptor and Activates Virus Entry," *J. Virol.*, vol. 85, no. 2, pp. 873–882, Jan. 2011, doi: 10.1128/JVI.02062-10.

[10] R. Bussani *et al.*, "Persistence of viral RNA, pneumocyte syncytia and thrombosis are hallmarks of advanced COVID-19 pathology," *EBioMedicine*, vol. 61, p. 103104, Nov. 2020, doi: 10.1016/j.ebiom.2020.103104.

[11] Z. Xu *et al.*, "Pathological findings of COVID-19 associated with acute respiratory distress syndrome," *Lancet Respir. Med.*, vol. 8, no. 4, pp. 420–422, Apr. 2020, doi: 10.1016/S2213-2600(20)30076-X.

[12] J. Schneider *et al.*, "SARS-CoV-2 direct cardiac damage through spike-mediated cardiomyocyte fusion," In Review, preprint, Oct. 2020. doi: 10.21203/rs.3.rs-95587/v1.

[13] G. Zhou and Q. Zhao, "Perspectives on therapeutic neutralizing antibodies against the Novel Coronavirus SARS-CoV-2," *Int. J. Biol. Sci.*, vol. 16, no. 10, pp. 1718–1723, Mar. 2020, doi: 10.7150/ijbs.45123.

[14] J. Hansen *et al.*, "Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail," *Science*, vol. 369, pp. 1010–1014, Jun. 2020, doi: 10.1126/science.abd0827.

[15] C. Wang *et al.*, "A human monoclonal antibody blocking SARS-CoV-2 infection," *Nat. Commun.*, vol. 11, no. 1, Art. no. 1, May 2020, doi: 10.1038/s41467-020-16256-y.

[16] D. Wrapp *et al.*, "Structural Basis for Potent Neutralization of Betacoronaviruses by Single-Domain Camelid Antibodies," *Cell*, vol. 181, no. 5, pp. 1004-1015.e15, May 2020, doi: 10.1016/j.cell.2020.04.031.

[17] P. J. M. Brouwer *et al.*, "Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability," *Science*, vol. 369, no. 6504, pp. 643–650, Aug. 2020, doi: 10.1126/science.abc5902.

[18] Y. Cao *et al.*, "Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-Throughput Single-Cell Sequencing of Convalescent Patients' B Cells," *Cell*, vol. 182, no. 1, pp. 73-84.e16, Jul. 2020, doi: 10.1016/j.cell.2020.05.025.

[19] X. Chi *et al.*, "A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2," *Science*, vol. 369, no. 6504, pp. 650–655, Aug. 2020, doi: 10.1126/science.abc6952.

[20] B. Ju *et al.*, "Human neutralizing antibodies elicited by SARS-CoV-2 infection," *Nature*, vol. 584, no. 7819, Art. no. 7819, Aug. 2020, doi: 10.1038/s41586-020-2380-z.

[21] Y. Lou *et al.*, "Cross-neutralization antibodies against SARS-CoV-2 and RBD mutations from convalescent patient antibody libraries," *bioRxiv*, p. 2020.06.06.137513, Jun. 2020, doi: 10.1101/2020.06.06.137513.

[22] T. F. Rogers *et al.*, "Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model," *Science*, vol. 369, no. 6506, pp. 956–963, Aug. 2020, doi: 10.1126/science.abc7520.

[23] R. Shi *et al.*, "A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2," *Nature*, vol. 584, no. 7819, Art. no. 7819, Aug. 2020, doi: 10.1038/s41586-020-2381-y.

[24] J. Wan *et al.*, "Human-IgG-Neutralizing Monoclonal Antibodies Block the SARS-CoV-2 Infection," *Cell Rep.*, vol. 32, no. 3, p. 107918, Jul. 2020, doi: 10.1016/j.celrep.2020.107918.

[25] Y. Wu *et al.*, "A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2," *Science*, vol. 368, no. 6496, pp. 1274–1278, Jun. 2020, doi: 10.1126/science.abc2241.

[26] D. Pinto *et al.*, "Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody," *Nature*, vol. 583, no. 7815, Art. no. 7815, Jul. 2020, doi: 10.1038/s41586-020-2349-y.

[27] J. Huo *et al.*, "Neutralization of SARS-CoV-2 by Destruction of the Prefusion Spike," *Cell Host Microbe*, vol. 28, no. 3, pp. 445-454.e6, Sep. 2020, doi: 10.1016/j.chom.2020.06.010.

[28] A. Frenzel *et al.*, "Designing Human Antibodies by Phage Display," *Transfus. Med. Hemotherapy*, vol. 44, no. 5, pp. 312–318, Sep. 2017, doi: 10.1159/000479633.

[29] R.-M. Lu *et al.*, "Development of therapeutic antibodies for the treatment of diseases," *J. Biomed. Sci.*, vol. 27, no. 1, p. 1, Jan. 2020, doi: 10.1186/s12929-019-0592-z.

[30] T. J. Vaughan, J. K. Osbourn, and P. R. Tempest, "Human antibodies by design," *Nat. Biotechnol.*, vol. 16, no. 6, Art. no. 6, Jun. 1998, doi: 10.1038/nbt0698-535.

[31] E. O. Saphire *et al.*, "Systematic analysis of monoclonal antibodies against Ebola virus GP defines features that contribute to protection," *Cell*, vol. 174, no. 4, pp. 938-952.e13, Aug. 2018, doi: 10.1016/j.cell.2018.07.033.

[32] A. X. Goh *et al.*, "A novel human anti-interleukin-1β neutralizing monoclonal antibody showing in vivo efficacy," *mAbs*, vol. 6, no. 3, pp. 764–772, May 2014, doi: 10.4161/mabs.28614.

[33] B. E. Young *et al.*, "Epidemiologic Features and Clinical Course of Patients Infected With SARS-CoV-2 in Singapore," *JAMA*, vol. 323, no. 15, p. 1488, Apr. 2020, doi: 10.1001/jama.2020.3204.

[34] A. Pizzorno *et al.*, "Characterization and Treatment of SARS-CoV-2 in Nasal and Bronchial Human Airway Epithelia," *Cell Rep. Med.*, vol. 1, no. 4, Jul. 2020, doi: 10.1016/j.xcrm.2020.100059.

[35] D. R. Burton, E. O. Saphire, and P. W. Parren, "A model for neutralization of viruses based on antibody coating of the virion surface," *Curr. Top. Microbiol. Immunol.*, vol. 260, pp. 109–143, 2001, doi: 10.1007/978-3-662-05783-4_7.

[36] K. A. Dowd and T. C. Pierson, "Antibody-mediated neutralization of flaviviruses: a reductionist view," *Virology*, vol. 411, no. 2, pp. 306–315, Mar. 2011, doi: 10.1016/j.virol.2010.12.020.

[37] A. Flamand, H. Raux, Y. Gaudin, and R. W. H. Ruigrok, "Mechanisms of Rabies Virus Neutralization," *Virology*, vol. 194, no. 1, pp. 302–313, May 1993, doi: 10.1006/viro.1993.1261.

[38] P. S. McNamara and R. L. Smyth, "The pathogenesis of respiratory syncytial virus disease in childhood," *Br. Med. Bull.*, vol. 61, no. 1, pp. 13–28, Mar. 2002, doi: 10.1093/bmb/61.1.13.

[39] M. Koot *et al.*, "Prognostic Value of HIV-1 Syncytium-Inducing Phenotype for Rate of CD4+ Cell Depletion and Progression to AIDS," *Ann. Intern. Med.*, vol. 118, no. 9, pp. 681–688, May 1993, doi: 10.7326/0003-4819-118-9-199305010-00004.

[40] A. Sylwester, S. Murphy, D. Shutt, and D. R. Soll, “HIV-induced T cell syncytia are self-perpetuating and the primary cause of T cell death in culture.,” *J. Immunol.*, vol. 158, no. 8, pp. 3996–4007, Apr. 1997.

[41] D. Wrapp *et al.*, “Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation,” *Science*, vol. 367, no. 6483, pp. 1260–1263, Mar. 2020, doi: 10.1126/science.abb2507.

[42] L. Yurkovetskiy *et al.*, “Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant,” *Cell*, vol. 183, no. 3, pp. 739-751.e8, Oct. 2020, doi: 10.1016/j.cell.2020.09.032.

[43] A. Punjani and D. J. Fleet, “3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM,” *J. Struct. Biol.*, vol. 213, no. 2, p. 107702, Jun. 2021, doi: 10.1016/j.jsb.2021.107702.

[44] A. C. Walls *et al.*, “Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion,” *Cell*, vol. 176, no. 5, pp. 1026-1039.e15, Feb. 2019, doi: 10.1016/j.cell.2018.12.028.

[45] H. Liu *et al.*, “Cross-Neutralization of a SARS-CoV-2 Antibody to a Functionally Conserved Site Is Mediated by Avidity,” *Immunity*, vol. 53, no. 6, pp. 1272-1280.e5, Dec. 2020, doi: 10.1016/j.immuni.2020.10.023.

[46] M. Yuan *et al.*, “A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV,” *Science*, vol. 368, no. 6491, pp. 630–633, May 2020, doi: 10.1126/science.abb7269.

[47] C. O. Barnes *et al.*, “SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies,” *Nature*, vol. 588, no. 7839, Art. no. 7839, Dec. 2020, doi: 10.1038/s41586-020-2852-1.

[48] B. Korber *et al.*, "Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus," *Cell*, vol. 182, no. 4, pp. 812-827.e19, Aug. 2020, doi: 10.1016/j.cell.2020.06.043.

[49] G. Vauquelin and S. J. Charlton, "Exploring avidity: understanding the potential gains in functional affinity and target residence time of bivalent and heterobivalent ligands," *Br. J. Pharmacol.*, vol. 168, no. 8, pp. 1771–1785, 2013, doi: https://doi.org/10.1111/bph.12106.

[50] L. Zhang *et al.*, "SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity," *Nat. Commun.*, vol. 11, no. 1, Art. no. 1, Nov. 2020, doi: 10.1038/s41467-020-19808-4.

[51] K. Huang, L. Incognito, X. Cheng, N. D. Ulbrandt, and H. Wu, "Respiratory Syncytial Virus-Neutralizing Monoclonal Antibodies Motavizumab and Palivizumab Inhibit Fusion," *J. Virol.*, vol. 84, no. 16, pp. 8132–8140, Aug. 2010, doi: 10.1128/JVI.02699-09.

[52] O.-W. Ng, C.-T. Keng, C. S.-W. Leung, J. S. M. Peiris, L. L. M. Poon, and Y.-J. Tan, "Substitution at Aspartic Acid 1128 in the SARS Coronavirus Spike Glycoprotein Mediates Escape from a S2 Domain-Targeting Neutralizing Monoclonal Antibody," *PLOS ONE*, vol. 9, no. 7, p. e102415, Jul. 2014, doi: 10.1371/journal.pone.0102415.

[53] F.-X. Lescure *et al.*, "Clinical and virological data of the first cases of COVID-19 in Europe: a case series," *Lancet Infect. Dis.*, vol. 20, no. 6, pp. 697–706, Jun. 2020, doi: 10.1016/S1473-3099(20)30200-0.

[54] D. N. Mastronarde, "Automated electron microscope tomography using robust prediction of specimen movements," *J. Struct. Biol.*, vol. 152, no. 1, pp. 36–51, Oct. 2005, doi: 10.1016/j.jsb.2005.07.007.

[55] S. Q. Zheng, E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng, and D. A. Agard, "MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy," *Nat. Methods*, vol. 14, no. 4, pp. 331–332, Apr. 2017, doi: 10.1038/nmeth.4193.

[56] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, "cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination," *Nat. Methods*, vol. 14, no. 3, Art. no. 3, Mar. 2017, doi: 10.1038/nmeth.4169.

[57] S. H. W. Scheres, "A Bayesian View on Cryo-EM Structure Determination," *J. Mol. Biol.*, vol. 415, no. 2, pp. 406–418, Jan. 2012, doi: 10.1016/j.jmb.2011.11.010.

[58] B. Webb and A. Sali, "Comparative Protein Structure Modeling Using MODELLER," *Curr. Protoc. Bioinforma.*, vol. 54, no. 1, p. 5.6.1-5.6.37, 2016, doi: 10.1002/cpbi.3.

[59] E. F. Pettersen *et al.*, "UCSF Chimera--a visualization system for exploratory research and analysis," *J. Comput. Chem.*, vol. 25, no. 13, pp. 1605–1612, Oct. 2004, doi: 10.1002/jcc.20084.

[60] T. D. Goddard *et al.*, "UCSF ChimeraX: Meeting modern challenges in visualization and analysis," *Protein Sci. Publ. Protein Soc.*, vol. 27, no. 1, pp. 14–25, 2018, doi: 10.1002/pro.3235.

[61] T. I. Croll, "ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps," *Acta Crystallogr. Sect. Struct. Biol.*, vol. 74, no. 6, Art. no. 6, Jun. 2018, doi: 10.1107/S2059798318002425.

[62] R. Y.-R. Wang, Y. Song, B. A. Barad, Y. Cheng, J. S. Fraser, and F. DiMaio, "Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta," *eLife*, vol. 5, p. e17219, Sep. 2016, doi: 10.7554/eLife.17219.

Chapter 3

Dispatched uses Na^+ flux to power lipid-modified Hedgehog release

Contributing Authors

Qianqian Wang[1†], Daniel E. Asarnow[2†], Ke Ding[1], Randall Mann[1], Jason Hatakeyama[1], Yunxiao Zhang[1‡], Yong Ma[3§], Yifan Cheng[2,4,*], Philip A. Beachy[1,5,*]

[1]Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, California 94305, USA

[2]Department of Biochemistry and Biophysics, University of California, San Francisco, California 94158, USA

[3]Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, CA 21205

[4]Howard Hughes Medical Institute, University of California, San Francisco, California 94158, USA

[5]Departments of Urology, and Developmental Biology, Stanford University School of Medicine, Stanford, California 94305, USA

§Current address: Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, 88 Keling Road, SND, Suzhou, Jiangsu 215163, China

‡Current address: Howard Hughes Medical Institute, Neuroscience Department, The Scripps Research Institute, La Jolla, CA 92037, USA.

†These authors contributed equally to this work.

Abstract

The DISP1 protein, related to the NPC1 and PTCH1 cholesterol transporters[1], [2] and to H^+-driven transporters of the RND family[3], [4], enables tissue patterning activity of the lipid-modified Hedgehog protein by releasing it from tightly-localized sites of embryonic expression[5]–[10]. A 2.5 Å resolution cryo-EM structure of mouse DISP1 revealed three Na^+ ions coordinated within a channel that traverses its transmembrane domain; the rate of Hedgehog export furthermore is dependent on the plasma membrane Na^+ gradient. The transmembrane channel and Na^+-binding are disrupted in DISP1-NNN (3.2 Å resolution), a variant with asparagine substitutions for three intramembrane aspartates that each coordinate and neutralize charge for one of the three Na^+ ions. DISP1-NNN and variants that disrupt single Na^+ sites retain binding to but are impaired in export of the lipid-modified Hedgehog protein to SCUBE2 acceptor. Interaction of Sonic hedgehog protein (ShhN) with DISP1 (2.7 Å complex structure) comprises extensive buried surface area and contacts with an extended furin-cleaved DISP1 arm. Variability analysis reveals ShhN binding restricted to one extreme of a continuous series of DISP1 conformations. The bound and unbound DISP1 conformations display distinct Na^+ site occupancies, suggesting a mechanism by which transmembrane Na^+ flux may power extraction of the lipid-linked Hedgehog signal from the membrane. Na^+-coordinating residues in DISP1 are conserved in PTCH1 and other metazoan RND family members, suggesting utilization of Na^+ flux to power their conformationally-driven activities.

Introduction

During embryogenesis the Hedgehog protein signal projects its tissue patterning influence many cells beyond its sites of expression in “organizers” such as the vertebrate notochord, the floor plate of the neural tube, and the zone of polarizing activity in developing limb[11]–[14]. The

mature Hedgehog protein signal (ShhNp) is covalently lipid-modified[15] by autoprocessing-mediated cholesterol attachment at its carboxy-terminus[16] and acyl transferase-mediated attachment of palmitate at its amino-terminus[17], [18]. The resulting dually lipid-modified protein is tightly membrane associated and thus requires dedicated machinery for its packaging and release in long-range patterning. This machinery includes the DISP1 protein[5]–[10], required in Hedgehog-producing cells, and SCUBE, a secreted protein that can be supplied by other cells that do not produce or respond to the Hedgehog signal[19]–[21]. Curiously, although the DISP1 protein and the related Hedgehog receptor PTCH1 are similar in sequence and transmembrane topology and both interact with the lipid-modified Hh signal, DISP1-mediated Hh release promotes pathway activity whereas PTCH1 attenuates activity by inhibiting the essential response component *Smoothened* (*Smo*) (**Figure 3.5A-B**; see supplemental discussion).

Results

DISP1 structure determination

We determined the structure of a stable, functional N-terminal truncation of DISP1, hereafter referred to as DISP1-A (Extended Data Figs. 1c-e, 2). Monomeric DISP1-A purified into a mixture of LMNG/GDN/CHS was efficiently cleaved at a recently reported essential furin site[22] (**Figure 3.5F**), and 3D classification of a single-particle cryo-EM dataset revealed high-resolution structures of two major classes, R (relaxed, 2.5 Å resolution) and T (tense, 2.5 Å resolution), each major class comprising at least two subclasses (**Figure 3.7** and **Figure 3.9**; **Table 2.1**). We further applied 3D variability analysis (3DVA, see Methods) [23], to systematically model the conformational landscape within our dataset as a series of orthonormal principal components representing specific modes of motion. We found conformations resembling R and T

at opposite ends of the second principal component (PC2) of this dataset (**Figure 3.9**; **Movie 3.1**), and we present here a model of DISP1-A built from the density of R.

DISP1-A structure resembles that of PTCH1, with twelve transmembrane helices (TM1-TM12) derived from apparent duplication of a six-transmembrane helix unit (TM1-6, TM7-12) arranged in pseudodyad symmetry and preceded by amphipathic transverse helices at the cytoplasmic membrane surface (pre-TM1 and pre-TM7). Each unit contains a large extracellular loop derived from corresponding positions between TM1-TM2 and TM7-TM8 (**Figure 3.1A-B**). Unlike the extracellular domains (ECDs) of PTCH1, which are juxtaposed to form a conduit for sterol transport[1], the mouse DISP1 ECDs are splayed apart in a Y-shaped manner (**Figure 3.1A**), as also recently reported for *Drosophila* and human DISP [24], [25]. DISP1 also resembles PTCH1 in the presence of ferredoxin-like α+β open-faced sandwich folds in the membrane-proximal parts of ECD1 and ECD2 (**Figure 3.5G**).

Each ECD accepts secondary structure elements from the other loop, enclosing a bowl-shaped space between the two domains. The linker within the TM1-TM2 loop is the site of furin cleavage (RERR, 276-279) [22], opening the bowl at a location we designate as "front". The ECDs contain 7 disulfide bonds, 5 in ECD1 and 2 in ECD2 (**Figure 3.1B**, **Figure 3.6**, and **Figure 3.10**), and 5 N-linked glycosylation sites (**Figure 3.1B, Figure 3.5G**, and **Figure 3.6**). Except for pre-TM1 and pre-TM7, most intracellular sequence in DISP1-A is not resolved, including four amino-terminal residues (1- 2, 172-3), 14 residues preceding pre-TM7 (667-680), and 378 residues at the carboxy-terminus (1144-1521). In the ECDs, 19 residues including the furin cleavage site (264-282) and residues 399-409 are not resolved.

We modeled 28 lipid or detergent molecules, mostly at annular positions around the transmembrane domain. Twenty-six were CHS, seventeen corresponding to the inner leaflet and

nine to the outer leaflet of the membrane, presenting the appearance of a sterol bilayer (**Figure 3.1C**; **Movie 3.2**). Two of the nine outer leaflet sterols are raised, and the front side lifted sterol may correspond to an intermediate state in ShhNp export (see below); a nearby lipidic group, modeled as LMNG, levers the lifted sterol upward (**Figure 3.1C**; **Movie 3.2**). We also modeled a non-proteogenic density within a hydrophobic cavity at the distal tip of ECD2 as an LMNG molecule (**Figure 3.1C**; **Movie 3.2**).

Na^+ ions in a transmembrane tunnel

The transmembrane domain contained three striking non-proteogenic densities (sites I, II and III), each associated with one of the three acidic residues in TM4 and TM10 that are conserved in PTCH1 (**Figure 3.2A**, **Figure 3.11A-C**). These densities appear to be metal cations, well-coordinated by oxygens from side-chain hydroxyls and main-chain carbonyls, and at least one oxygen from water (site III) (**Figure 3.2A**). Site I and II cations lie on the axis of pseudodyad symmetry between TM1-6 and TM7-12, each receiving coordination bonds from symmetry-related oxygens in both opposing six-transmembrane helix units. The site III cation is asymmetrically positioned, its ligands deriving entirely from TM4 and TM5.

The coordination number is five for sites I and II and six for site III, the center-to-center distances between ions and their associated oxygens averaging 2.50 Å, 2.51 Å, and 2.47 Å, respectively. Each density in these three sites is associated with a single carboxylate and with oxygens from hydroxyls and main chain carbonyls, ruling out alkaline earth metal ions such as Ca^{++} or Mg^{++}, which require additional charge neutralization, display a more rigid geometry and a tighter coordination sphere, and are rarely coordinated by main-chain carbonyls. The average bond distances match those characteristic of Na^+ (2.4 – 2.5 Å) rather than K^+ (2.7-3.2 Å) [26], and we

conclude that the densities at sites I, II, and III derive from Na^+, as further validated by the *CheckMyMetal web server* [26].

To assess the role of Na^+ in DISP1 activity, we modified a quantitative assay for release of ShhNp [6] by: (*i*) replacing luciferase between amino acid residues 131 and 132 of SHH with nanoluciferase, (**Figure 3.12A**); (*ii*) use of DISP1-A-mEGFP (**Figure 3.5C**); and (*iii*) use of purified SCUBE2 protein in defined medium as the acceptor (**Figure 3.12B-C**). This assay provides a rapid, quantitative measure of the rate of nanoLuc-ShhNp release into the medium over a period of minutes (**Figure 3.12D**). Under reduced Na^+, with the organic cation NMDG used to maintain osmolarity [27], we found that release by DISP1-A activity ranged from near-zero at 10 mM extracellular Na^+ to partial and full restoration of release at extracellular Na^+ concentrations of 40 and 155 mM (**Figure 2B**, **Figure 3.12D**); a similar effect of Na^+ was recently reported in ref. [28].

Simultaneous alteration to asparagine of D571, D572, and D1049, the charge-neutralizing residues within TM4 and TM10 in DISP1-A (DISP1-A-NNN) resulted in disruption of ShhNp release (**Figure 3.2C**) *in vitro*, as previously established *in vivo* [6]. Disruption of sites I, II, and III individually also disrupted release of ShhNp (**Figure 3.2C**, **Figure 3.12E**). Given that all of these variant proteins are stable and retain the ability to bind ShhNp (**Figure 3.12F**), the results indicate that Na^+ coordination is critical.

Utilization of a Na^+ gradient to power DISP1 activity implies Na^+ flux across the membrane, and we indeed found a path extending from the extracellular to the intracellular surfaces of the DISP1 transmembrane domain [29]; most water molecules within the transmembrane domain are clustered near this channel (**Figure 3.3A**; **Movie 3.3**). The structure of DISP1-A-NNN determined at 3.2 Å resolution (**Figure 3.7**; **Table 2.1**), revealed that the Na^+ ion

densities at sites I, II, and III are indeed absent, and the side chains of the substituted aspartates and some of the remaining liganding side chains are rearranged (**Figure 3.3B**). We compared water-excluded internal cavities within the transmembrane domains of these proteins (**Figure 3.11D**, see supplemental discussion), and found that a large cavity within DISP1-A containing the three Na^+ ions is strikingly disrupted in DISP1-A-NNN by a series of side-chain rotamer swaps, including interactions across the pseudodyad axis between the altered residues D571N and D1049N, with concomitant straightening and movement of TM4 and TM10 toward each other (**Figure 3.11D**, bottom panels; **Figure 3.3C**; supplemental discussion). These motions, associated with Na^+ loss in DISP1-A-NNN, bring into closer proximity the pseudodyad halves of the transmembrane domain (TM1-6 and TM7-12), suggesting possible modes of allosteric communication, from the nexus of Na^+ coordination outward toward ECD1, ECD2, and other Hh-handling portions of DISP1.

Structure of a ShhN:DISP1-A complex

We determined the structure of DISP1-A in complex with a truncated Shh protein containing residues 26-189 and lacking lipid modifications (ShhN) (2.7 Å overall resolution; **Figures 2.8**, **Figure 3.10**, and **Figure 3.13**; **Table 2.1**). The largest of the two major density classes contained substantial additional density corresponding to ShhN perched near the top front side of DISP1-A, between ECD1 and ECD2 (**Figure 3.4A**). In the model of this ShhN:DISP1-A complex (**Figure 3.4B**), the ShhN portion of the map resolved residues 39-189, as well as the previously described Zn^{++} ion[30] and the di-atomic Ca^{++} cluster [31], along with all major secondary structure elements, and most side chains (**Figure 3.4A-B**, **Figure 3.10**, and **Figure 3.13**).

In the complex, ECD1 and ECD2 of DISP1-A bind ShhN in a pincer-like grasp (**Figure 3.4A-B**), including hydrophobic contacts and possible ionic interactions with total buried surface

area of 989.3 Å^2 (-8.6 kcal/mole; see Methods). The position of ShhN relative to DISP1-A in our complex structure is distinct from the position recently proposed by Cannac et al. [24] for *Drosophila* HhN (**Figure 3.13C**). Strikingly, the furin-cleaved front-side linker of DISP1-A, which is considerably sharper than in the DISP1-A apostructure, extends from ECD2 upward and around ShhN in a one-armed molecular embrace (**Figure 3.4A-B**). We extended the model of this linker by an additional 13 residues preceding the furin cleavage site (**Figure 3.6**). Improved ECD resolution in the ShhN:DISP1-A map also allowed us to extend the DISP1-A model by an additional 11 residues (399-409) in distal ECD1 (**Figure 3.4B**; **Figure 3.6**), revealing a hydrophobic cavity that could interact with an extended hydrophobic moiety (**Figure 3.10F**), and reminiscent of the hydrophobic cavity at distal ECD2 (**Figure 3.1C**; **Figure 3.10G**).

Conformational dynamics of DISP1

We applied 3DVA [23] to this dataset and found the first principal component (PC1) to be anchored at opposite ends by DISP1-A molecules either complexed with or lacking the ShhN ligand (**Figure 3.14A**, **Movie 3.4**). A similar range of DISP1-A conformations is found within PC1 of 3DVA of the DISP1-A apoprotein dataset (**Figure 3.14B**), suggesting that ShhN binds preferentially to a restricted subset of common DISP1-A conformations.

Close examination revealed that Na^+ sites are differentially occupied at opposite extremes of the PC1 reaction coordinates for datasets with or without ShhN (**Figure 3.15A**, **Movie 3.4**), with densities corresponding to all three Na^+ ions present at the end anchored by unbound DISP1-A, and only Na^+ site I occupied at the opposite end, anchored by the ShhN-bound complex. Na^+ ion occupancy thus drives or is permissive of alternative conformational states of DISP1, suggesting that chemiosmotically driven flux through the DISP1 transmembrane tunnel could induce conformational changes that cause Hedgehog release after initial binding.

Membrane extraction of Hedgehog lipid adducts

The release of mature Hedgehog signal requires extraction of its covalently linked cholesteryl and palmitoyl adducts from the membrane; the lifted sterol on the front side of DISP1-A (**Figure 3.1C**) thus may correspond to an intermediate state of the cholesteryl adduct as it is being removed from its hydrophobic environment. The hydrophobic cavities at the distal tips of ECD1 and ECD2 (**Figure 3.10F-G**) are large enough to accommodate the covalently linked Hedgehog lipid adducts, suggesting they may act as way-stations for the ShhNp lipid adducts after membrane extraction and prior to complex formation with SCUBE2.

Discussion

We propose that DISP1-A extraction and release of the lipid-modified Hedgehog signal proceeds *via* a state activated for recruitment of SCUBE2 (**Figure 3.4C**). We suggest that activation by DISP1-A binding to ShhN at a location ~25 Å above the membrane aids in extraction of the Hh-linked lipids, and the lifted sterol at the front side of DISP1-A may represent an intermediate in cholesteryl adduct extraction (**Figure 3.10F-G**). Three prominent features characterize this activated state: (*i*) a pincer-like action of the split ECDs; (*ii*) a molecular embrace of ShhN by the furin-cleaved linker arm; and (*iii*) the possible accommodation of lipid adducts in distal ECD1 and ECD2. This activated state is correlated with loss of Na^+ from Sites II and III, followed by Na^+ flux down its chemiosmotic gradient to re-occupy sites II and III and cause release of ShhNp to SCUBE2, a process incorporating opening of the ECD pincers, expulsion of the ECD1 and ECD2 lipid adducts, and withdrawal of the linker embrace, followed by resetting to permit another cycle.

We note from structure-based alignment that each of 15 protein-derived Na^+-liganding oxygens in DISP1 (10 from side-chains, five from main-chain carbonyls) is conserved in PTCH1,

as are the three charge-neutralizing intramembrane acidic residues (**Figure 3.15**). This degree of conservation strikingly exceeds that of the overall protein sequence (19% identity overall), and strongly supports the functional role of Na^+ flux in PTCH1 suppression of SMO activity, consistent with a previous demonstration of PTCH1 dependence on a transmembrane Na^+ gradient for its function [27]. All these residues are also conserved by sequence alignment in PTCH2, which also functions to suppress SMO [32]. In contrast, DISP2, which fails to rescue *disp* function in *Drosophila* [6] and probably lacks Hedgehog-releasing activity, has only 1 of three charge-neutralizing intramembrane acidic residues and 6 of 10 side-chain oxygens. Prokaryotic RND transporters harbor distinct residues at these Na^+-coordinating sites, (**Figure 3.15**), but most retain interactions between TM4 and TM10 that are conditional on cation binding, pointing to a common mechanism for harnessing chemiosmotic gradients (see Supplementary discussion).

How do DISP1 and PTCH1 produce such strikingly distinct functional effects on Hedgehog pathway activation? Our analysis suggests the use of chemiosmotic force from the metazoan transmembrane Na^+ gradient to power the release of lipids from membranes. For PTCH1, cholesterol removal *via* an enclosed sterol conduit in the PTCH1 ECD regulates SMO by depriving it of sterol that must be bound within the interior of its 7 transmembrane helical bundle for activity [33], [34]. DISP1 also appears to extract lipids from the membrane, but these lipids are covalently attached to the Hedgehog protein signal, necessitating a split of the DISP1 ECD to accommodate the protein moiety. Evolutionary acquisition of the furin cleavage site that permits this split, and of the linker arm that contacts ShhNp, apparently occurred by insertion into the *Disp1* gene of a single small exon that encodes both features (**Figure 3.6**). A fundamental mechanistic similarity is indicated by a lifted sterol in both DISP1 and PTCH1, at corresponding positions within the 12-transmembrane helix bundle [35]. The unusual positions of these sterols at

corresponding locations in both proteins suggest partial extraction from the membrane, perhaps representing a common intermediate in the lipid extraction activities of both proteins.

Supplemental Discussion

Opposing functions of Dispatched and Patched in Hedgehog signaling

DISP1 promotes pathway activation by releasing Hh from producing cells [5]–[10]: its embryonic mutant phenotype in mice is more severe than that of the *Sonic hedgehog* (*Shh*) or *Indian hedgehog* (*Ihh*) single mutants and resembles that of the *Shh;Ihh* double mutant, and of a homozygous knockout of the essential Hedgehog response component *Smoothened* (*Smo*) [6], [7], [36]. The PTCH1 protein in contrast functions to attenuate pathway activity by inhibiting SMO, and *Ptch1* homozygotes show widespread ectopic pathway activity [37]. The relationship between these opposing functions at distinct pathway levels is one of genetic epistasis, in which the *Ptch1*$^{-/-}$ phenotype, widespread pathway activation, prevails in *Ptch1*$^{-/-}$;*Disp1*$^{-/-}$ doubly homozygous mutant embryos [6], [9] (**Figure 3.1B**).

Conformational differences between DISP1-A and DISP1-A-NNN within the transmembrane domain

We examined the details of conformational differences between DISP1-A and DISP1-A-NNN in the context of the water-excluded surface within the transmembrane domain (**Figure 3.3B**). There are two extracellular branches at the top of the transmembrane domain, one narrow and admitting only a series of water molecules, and the other broader and of appropriate size to accommodate a hydrated Na^+ ion. These branches meet at Met557, which possesses two resolved rotamer conformations that can block either branch, but not both at once. Interestingly, only the conformation blocking the narrow branch is evident in DISP1-A-NNN, while a new alternate conformation of L1035 appears that would completely occlude the broader opening, and along

with M557 would seal both extracellular inlets (**Figure 3.3B**, top panels). Both extracellular openings into the transmembrane domain can thus be individually opened or closed, depending on the conformations of M557 and L1035.

The transmembrane channel is also strikingly disrupted deeper within the transmembrane domain of DISP1-A-NNN by a series of side-chain rotamer swaps and concomitant helical motions (**Figure 3.3B**, bottom panels). In DISP1-A, coordination of Na^+ excludes the main-chain carbonyls of G567 and G1045 from participating in the α-helix hydrogen bond networks of TM4 and TM10, inducing short kinks between residues 567-571 and 1045-1049, respectively (**Figure 3.3B**, lower panels; **Figure 3.3C**). In DISP1-A-NNN, however, the absence of Na^+ permits G567 and G1045 to join these hydrogen bond networks, thus straightening the TM4 and TM10 helices. Loss of these kinks in DISP1-A-NNN reduces the normal separation of TM4 and TM10, and rotates the side chain of I568 into the cavity surrounding the ion coordination sites, sterically excluding water. Simultaneously, D571N and D1049N rotate and approach each other across the transmembrane pseudodyad axis, interacting either directly or via a bridging water molecule (**Figure 3.3C**). These changes in DISP1-A-NNN isolate and dehydrate Na^+ site III, which collapses as TM5 is drawn toward TM4 by direct interactions of D572N with T610 and S611 (**Figure 3.3C**). In addition to revealing likely components of Na^+ conduction and changes in exposure of the Na^+ channel to either side of the membrane, the helical motions associated with loss of Na^+ coordination in DISP1-A-NNN also suggest plausible modes of allosteric communication outward from the nexus of ion coordination in the center of the transmembrane helical bundle to the peripheral lifted sterol site adjacent to TM1 and distally to ECD1 and ECD2.

The conservation of Na+-liganding side-chains and charge-neutralizing residues in RND family members

Na^+-coordinating residues in DISP1 are fully conserved in PTCH1 (**Figure 3.10C**, **Figure 3.14**). An intermediate level of conservation is seen for NPC1 (**Figure 3.14C**), which functions in the lysosome and thus has available to it both Na^+ and H^+ gradients [38]. Other mammalian family members, PTCHD2 and PTCHD3, display complete conservation of charge-neutralizing and liganding oxygen side-chains and thus seem likely to utilize a Na^+ gradient in their functions, whereas PTCHD1 and PTCHD4 show poor conservation of these residues and might have dispensed with transport activity or adopted a different mechanism for sustaining it (**Figure 3.14**). Most of those residues are not conserved in prokaryotic RND transporters at these Na^+-coordinating sites, consistent with their predominant utilization of proton motive force across the bacterial inner membrane to power their transport activities. Indeed, some prokaryotic transporters, such as the SecDF1 protein in halophiles (**Figure 3.13**), do require Na^+ for their activities, suggesting that coupling of Na^+ flux to transport may have evolved more than once to exploit alternate chemiosmotic gradients.

Figures

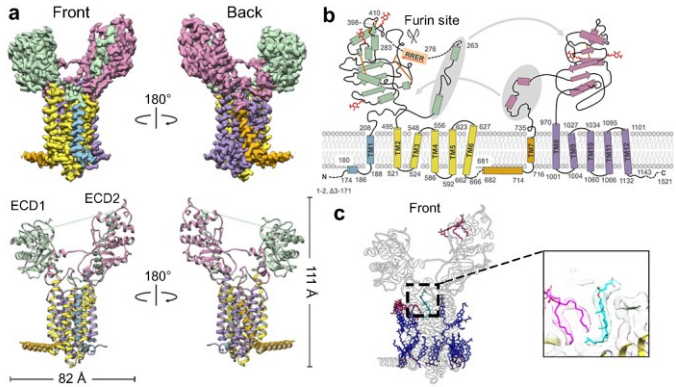


Figure 3.1 Structure and lipid binding sites of DISP1

(A) Corresponding views of the cryo-EM map (top) and atomic model (bottom) of DISP1-A (the R conformation, as seen by 3D classification and 3DVA; see text, and **Figures 2.7-2.9**). Domains are colored as follows: pre-TM1 helix and TM1, blue; TM1-TM2 loop, lime; TM2–TM7, yellow; pre-TM7 and TM7, orange; TM7-TM8 loop, pink; TM8–TM12, lavender.

(B) Topology diagram of DISP1-A showing secondary structure elements with colors as in (A), disulfide bonds (gold lines), glycosylation sites (red symbols) and unresolved regions (black dashed lines, amino acid residues 1-2, 172-3, 264-282, 399-409, 667-680, and 1144-1521; see Extended Data Fig. 2). Secondary elements exchanged between extracellular loops to form ECD1 and ECD2 are highlighted by oval shadows.

(C) Front view of DISP1-A (pale ribbon) with resolved LMNG (magenta) and CHS (blue) molecules (see also **Movie 3.2**). A sterol molecule (cyan) located at the front of DISP1 is positioned above other sterols in the outer leaflet, and is shown in the magnified inset (right) as a stick model with corresponding density and an associated, tilted LMNG molecule (magenta); this lifted sterol may correspond to an intermediate in the extraction of Hh-associated lipids from the membrane (see text).

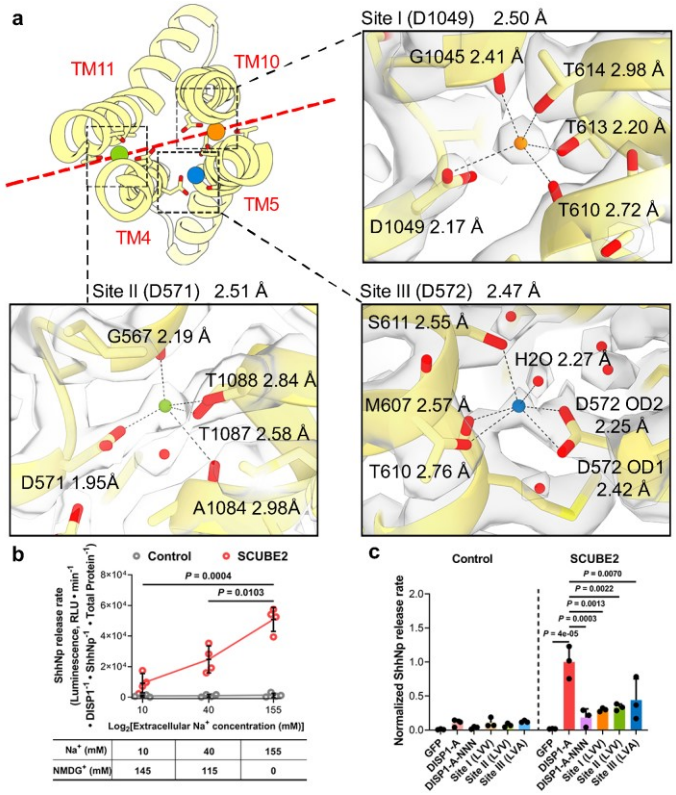


Figure 3.2 Three Na^+ densities in the transmembrane domain.

(A) Close-up views of the full set of coordinating ligands for each of the three Na^+ ions are shown with protein-derived densities superimposed upon the atomic model of DISP1-A (overview, top left, dashed red line showing the pseudodyad axis between TM1-6 and TM7-12). Dashed lines in magnified ion sites denote coordination bonds and corresponding center-to-center distances between Na^+ ions and associated oxygens. Contour levels are 6σ for sites I and II, and 5σ for site III.

(B)-(C) Na^+-dependent ShhNp release by DISP1. (B) NanoLuc-ShhNp released from *Disp*$^{-/-}$ MEFs transfected to express DISP1-A-mEGFP and NanoLuc-ShhNp, and incubated in defined medium at different Na^+ concentrations, either with or without (control) purified SCUBE2 protein (Mouse SCUBE2, lacking amino acids 30-281). $NMDG^+$ was added to maintain osmolarity for Na^+ concentrations < 155 mM (see Extended Data Fig. 8c). Slopes from linear regression applied to the linear portions of normalized NanoLuc-ShhNp release data were used to calculate release rates (see Methods; Extended Data Fig. 8d)**.** NanoLuc-ShhNp release was critically dependent on the Na^+ gradient. (C) Impaired NanoLuc-ShhNp release, as measured in (B), for DISP1-A (normalized to 1.0) and variants with alterations disrupting Na^+ binding at all three sites (DISP1-A-NNN: D571N/D572N/D1049N) or individually, at site I (LVV: D1049L/T613V/T614V), site II (LVV: D571L/T1087V/T1088V), or site III (LVA: D572L/T610V/S611A). SCUBE2 concentrations were 1 µM in (B) (30 min time course), and 0.15 µM in **c** (75 min time course). Data in (B) and (C) are represented as mean ± s.d., n = 3 (C) or 4 (B) biologically independent experiments; statistical analysis for $SCUBE2^+$ conditions (data sets normally distributed) utilized one-way ANOVA with Tukey's test in (B) or Dunnett's test in (C), with DISP1-A as comparison.

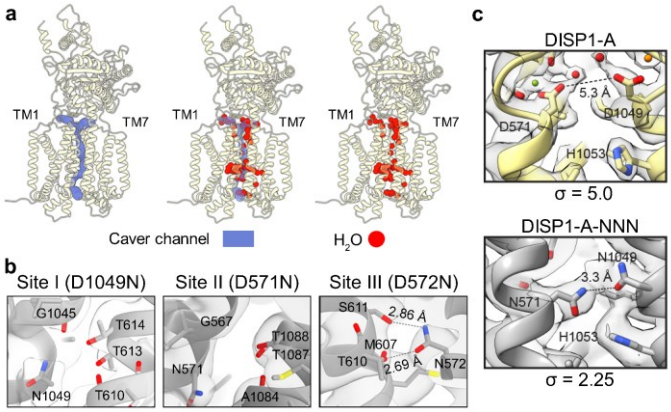


Figure 3.3 A transmembrane Na^+ permeation channel is disrupted in the structure of DISP1-A-NNN

(A) A continuous pathway through the transmembrane domain of DISP1-A, shown as a purple surface (left panel), is identified by the CAVER 3 [29] PyMOL plugin with a probe of radius 1.0 Å, appropriate for the passage of Na^+. This channel is partly occupied by resolved water molecules, shown as red spheres (right panel, superimposition with Caver channel in center panel), and provides access to the transmembrane cavity that contains the Na^+ ions and water (see text and Supplementary video 3).

(B) Close-up views of the three Na^+ ion sites in the DISP1-A-NNN structure (3.2 Å resolution) show no discernible Na^+ densities, even when contoured at 3σ (Na^+ densities in DISP1-A were clearly seen at ≥5σ for all three sites). Site III in particular has collapsed entirely, with TM4 and TM5 moving close enough to permit D572N to form hydrogen bonds with the side-chains of T610 and S611.

(C) In DISP1-A, D571 and D1049 side chains coordinate and neutralize their respective Na^+ ions, maintaining a minimum distance of 5.3 Å, whereas in DISP1-A-NNN the absence of Na^+ permits D571N and D1049N side chains to rotate toward and interact across the pseudodyad axis at a reduced distance of 3.3 Å. The motion of D571N together with an upward tilting of H1053 together disrupt the connection between the channel and the intracellular exit (see Supplementary discussion).

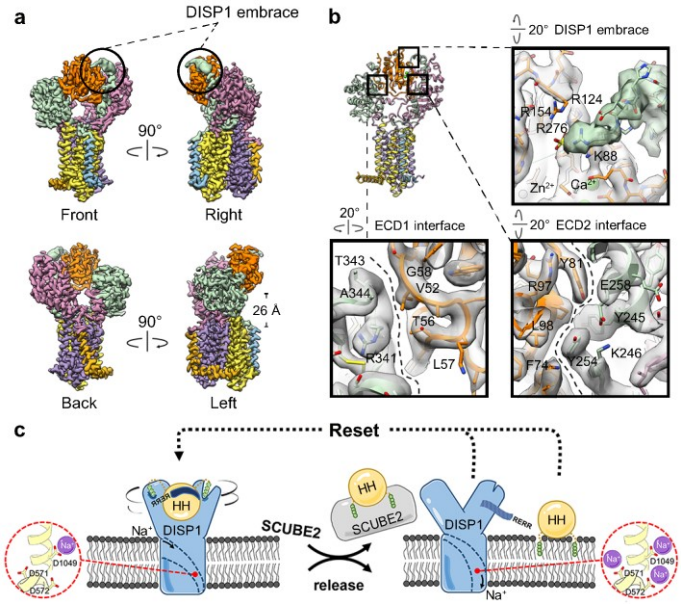


Figure 3.4 DISP1-A binds ShhN between its distal ECDs

(A) The cryo-EM density map of a ShhN:DISP1-A complex at 2.7Å resolution permitted *de novo* modeling of ShhN at a location 26 Å above the TM domain of DISP1-A. ShhN is in orange, DISP1-A colors are as in **Figure 3.1A**. The furin-cleaved linker, although not resolved in the DISP1-A apostructure, is stabilized within the complex and seen to extend from ECD2 in a one-armed molecular embrace of ShhN. The region of ShhN contacted by this linker corresponds to a polybasic region that interacts with heparin sulfate or other sulfated glycosaminoglycans (GAGs) [39], which would be displaced by the DISP1 molecular embrace.

(B) Three major interaction interfaces between ShhN and DISP1-A. The ECD1 and ECD2 interfaces feature primarily van der Waals interactions, with the extensive ECD2 interface additionally including a hydrogen bond between ShhN K75 and the backbone carbonyl of DISP1-A R241 (located just below the expanded view; not shown). Near the tip of the furin-cleaved linker arm, R276 of DISP1-A forms an ionic interaction with a sulfate ion that also interacts with basic residues in ShhN (R124, R154, and K88).

(C) A model for Na^+-driven conformational dynamics in the DISP1 activity cycle. Membrane anchored, dually lipid-modified Hedgehog protein is activated for release to SCUBE2 by interaction with DISP1 ECDs. Hedgehog activation (left side) entails (*i*) a pincer-like action of the DISP1 ECDs; (*ii*) interaction with and stabilization of the DISP1 furin-cleaved linker (dark blue); and (*iii*) accommodation of Hedgehog-linked lipids in hydrophobic distal cavities of DISP1 ECDs. From this activated state, associated with restricted Na^+ occupancy (site I only), SCUBE2 availability and the chemiosmotic force of the transmembrane Na^+ gradient together drive Hedgehog release to SCUBE2 and resetting of DISP1 for resumption of the cycle (right side). See **Movie 3.4**.

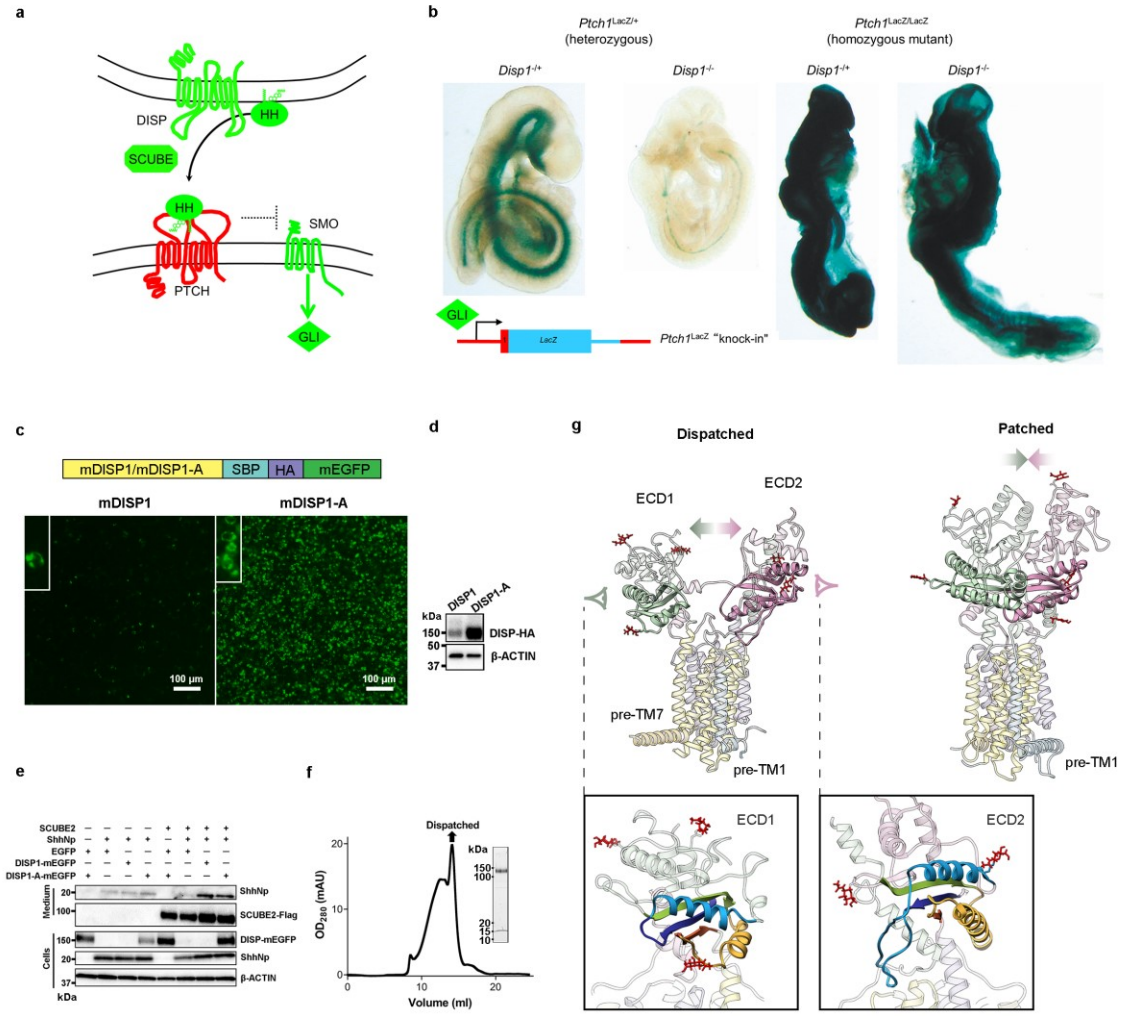


Figure 3.5 Structurally related transporters Dispatched and Patched in Hedgehog signaling
(A)-(B) Opposing functions of Dispatched and Patched in Hedgehog signaling. (A) The HH protein signal, covalently modified by cholesterol and palmitate, requires the action of DISP1 and SCUBE for release from the membrane of producing cells. HH then uses its palmitoyl adduct to clog the sterol transport conduit and block the function of its receptor PTCH1 in responding cells. The loss of PTCH1 sterol transport activity permits accumulation of cholesterol within the inner leaflet to levels that activate SMO by binding within its seven transmembrane helical bundle, resulting in activation of the GLI transcriptional effector of Hedgehog signaling. (B) As *Ptch1* is a target for GLI activation, the X-Gal staining of a *Ptch1*^{LacZ} knock-in allele[37] provides an indication of Hedgehog pathway activity (leftmost embryo). Homozygous mutation of *Disp1*[6]–[10] causes a loss of nearly all embryonic Hedgehog pathway activity (2nd embryo from left)[6], [36], whereas homozygous disruption of *Ptch1* leads to unregulated ectopic pathway activity, regardless of the functional status of *Disp1* (rightmost two embryos).
(C)-(F) Truncated DISP1 protein. As murine full-length DISP1 protein was poorly expressed in HEK293 cells, we tested a variety of constructs, and ultimately settled on an N-terminal truncation. Although its export activity was partially reduced, truncated DISP1-A protein nevertheless mediated efficient release of ShhNp (autoprocessed, lipid-modified Shh protein) into cell culture medium containing SCUBE2 (Mouse SCUBE2, lacking amino acids 30-281) upon transfection

into *Disp*$^{-/-}$ mouse embryonic fibroblasts[6], [20] (MEFs), indicating preservation of its function. (C) DISP1-mEGFP and DISP1-A-mEGFP in HEK293T cells with magnified insets showing expression on the cell membrane. (D) Western blot showing high level of DISP1-A expression relative to DISP1 (both proteins were SBP and HA-tagged at the C-terminus). (E) Functional assay (see Methods, ShhNp release endpoint assay (Western blot)) in *Disp*$^{-/-}$ mouse embryonic fibroblasts (MEFs). Culture media and cell lysates from transiently transfected *Disp*$^{-/-}$ MEFs were probed by immunoblotting for expression of DISP-mEGFP (SBP, HA and mEGFP tagged at the C-terminus, see panel (C)), ShhNp, and SCUBE2. DISP1 and DISP1-A both released ShhNp in the presence of SCUBE2. β-ACTIN, loading control. (F) Size exclusion chromatography of purified DISP1-A, together with the SDS-PAGE of the indicated fraction, corresponding to monomeric DISP1-A. Panels (C)-(F) show representative results (n = 4 biologically independent replicates).

(G) Structural comparison between DISP1-A and mouse PTCH1[40] (PDB code: 7k65). The ECDs, closely apposed in PTCH1 to form a conduit for sterol transport, in DISP1-A are splayed apart. Conserved ferredoxin-like α+β open-faced sandwich folds are highlighted (lime for ECD1 and pink for ECD2). Magnified views in spectral sequence from N to C termini of the α+β open-faced sandwich folds in DISP1-A ECD1 (bottom left) and ECD2 (bottom right). Distal structures inserted into the peripheral loops of the ferredoxin-like folds are structurally unrelated to each other or to distal PTCH1 ECD structures. Red symbols indicate the 5 N-linked glycosylation sites, 3 in ECD1 (N362, N390 and N475) and 2 in ECD2 (N834 and N915), which can be inferred from additional densities that extend from N-X-S/T sequences in the extracellular loops.

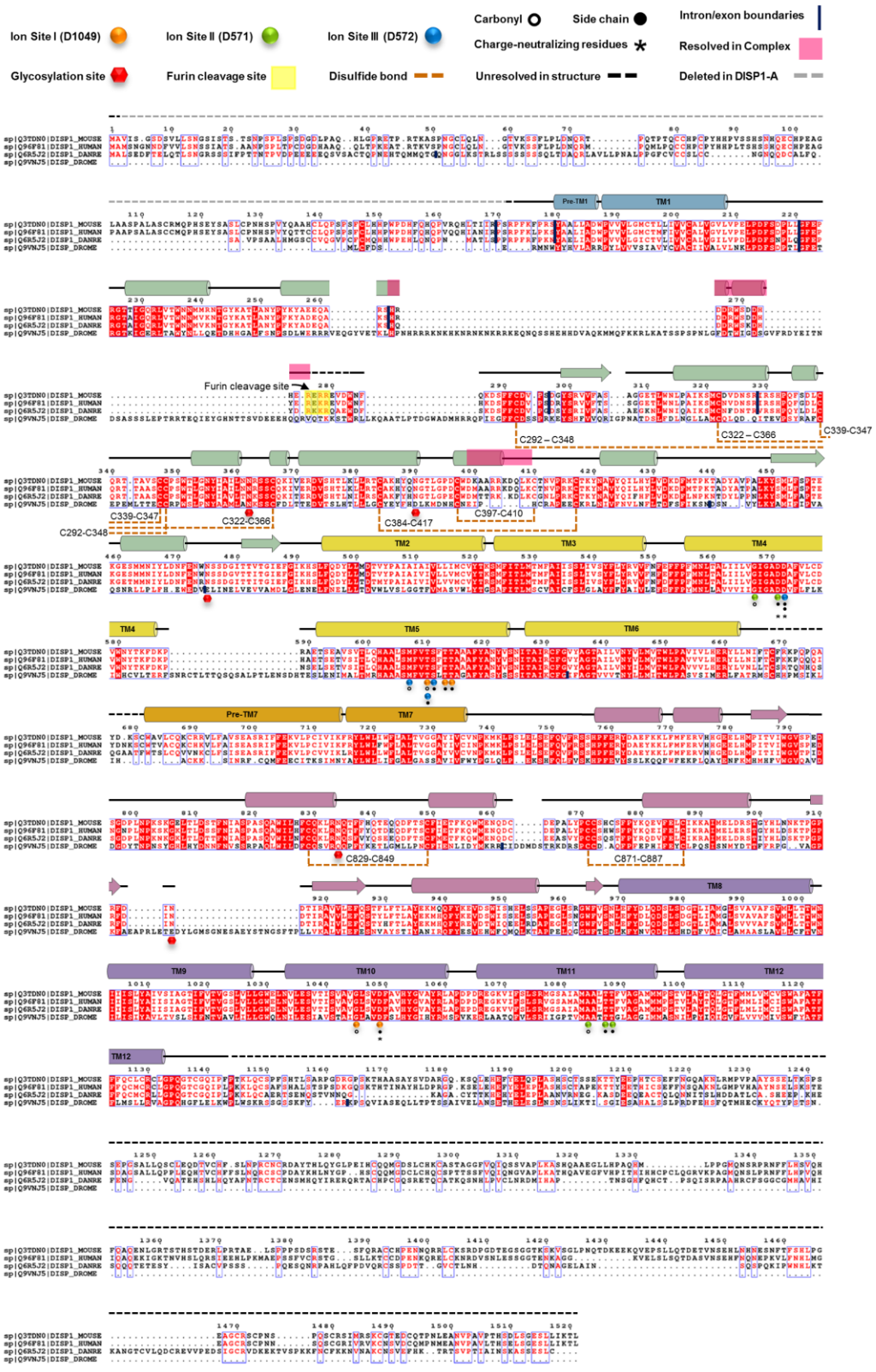


Figure 3.6 Dispatched protein sequence alignment and structural features

Sequence alignment of Dispatched proteins from *Mus musculus* (mouse), *Homo sapiens* (human), *Danio rerio* (zebrafish) and *Drosophila melanogaster* (fruit fly). Secondary structure elements and other features are indicated above the sequence, with subdomains colored according to Fig. 1a. Protein sequences were aligned using Clustal Omega[41] and the ESPript server[42]. (TM, transmembrane helix.)

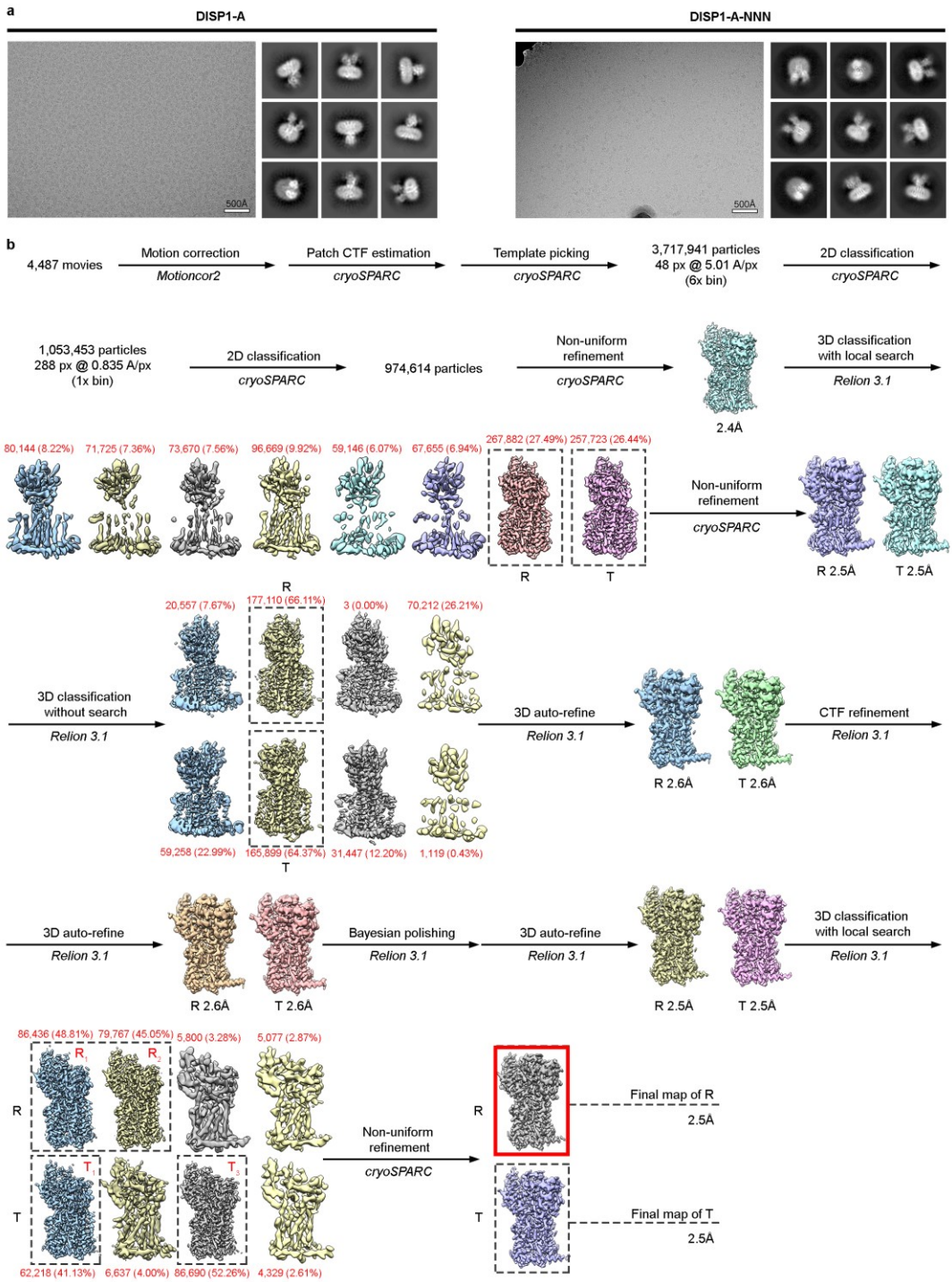


Figure 3.7 Cryo-EM data and image processing flow for DISP1-A and DISP1-A-NNN
(A) A representative cryo-EM micrograph (n = 4487 for DISP1-A, n = 4467 for DISP1-A-NNN) and several highly-populated, reference-free 2D class averages are shown for DISP1-A (left) and DISP1-A-NNN (right). The micrograph for DISP1-A-NNN has been contrast-stretched for display in order to account for the presence of a gold edge in the upper left corner of the image (the DISP1-A-NNN particle distribution on this grid necessitated targeting of the gold edge).

(B) Schematic flow-chart representing the image processing approach for DISP1-A. Thumbnail images of each 3D class or refinement are shown along with global GS-FSC resolution in black, particle counts in red, and dashed black boxes to indicate selected 3D classes. After separation of the R and T conformations in the first round of 3D classification, the identical processing flows for the two conformations are shown in parallel. Cryo-EM map (Red box) and atomic model of R conformation are used in main figures to present DISP1-A features. Subclasses of R and T conformations, R_1, R_2, T_1 and T_3, are labeled.

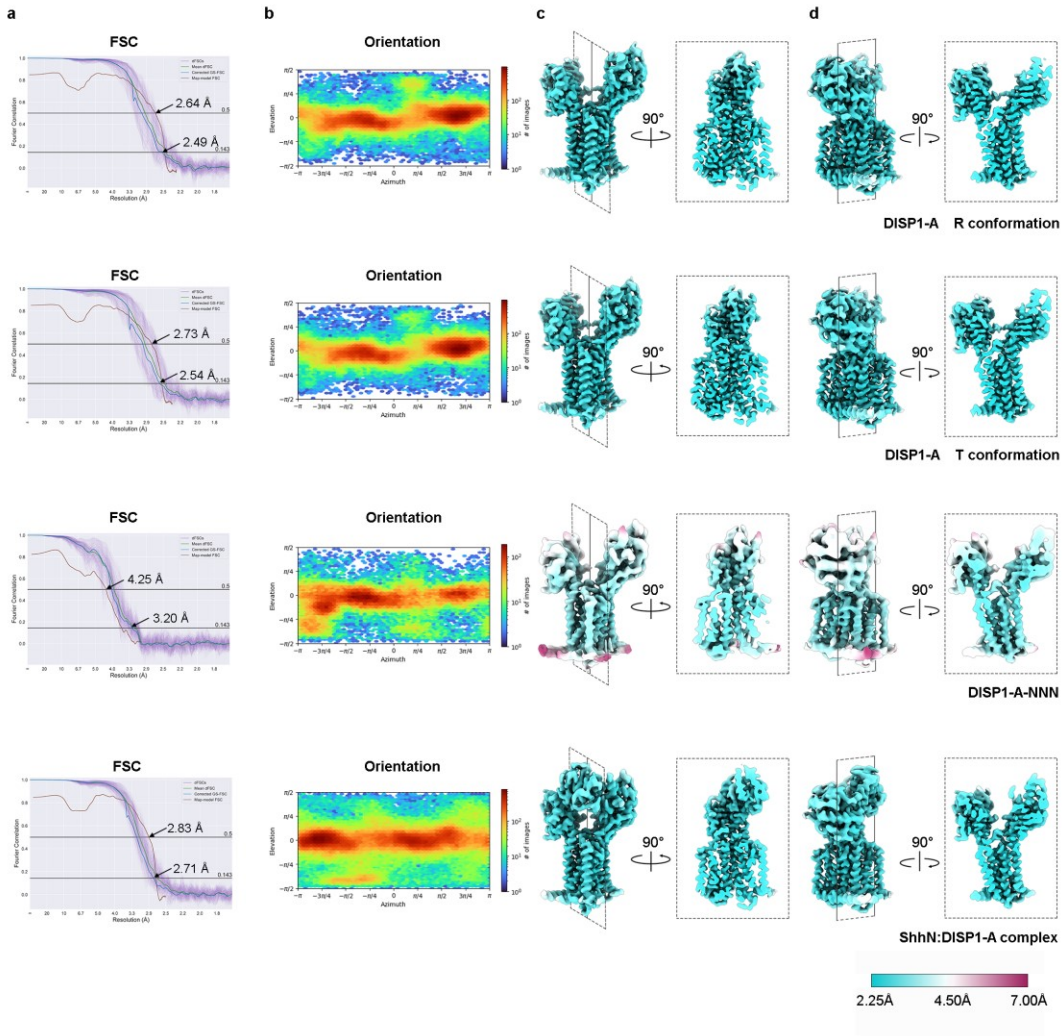


Figure 3.8 Cryo-EM density and atomic model quality

(A)-(D) Fourier shell correlation curves, particle orientation distributions, and local resolution maps are shown for R and T conformations of DISP1-A, DISP1-A-NNN, and for ShhN:DISP1-A complex. The "gold-standard" independent half-map FSC curves and orientation distributions were determined during refinement in cryoSPARC, map-to-model FSC curves were calculated in PHENIX using protein chains only, and directional FSC curves were estimated as in ref. [43]. The orientation distributions are plotted such that an elevation angle of 0˚ corresponds to a "side-view" perpendicular to the transmembrane helices; in each case the predominant views are "side-views" at a wide range of azimuthal angles. Local resolution estimates were computed using the BLOCRES algorithm as implemented in cryoSPARC.

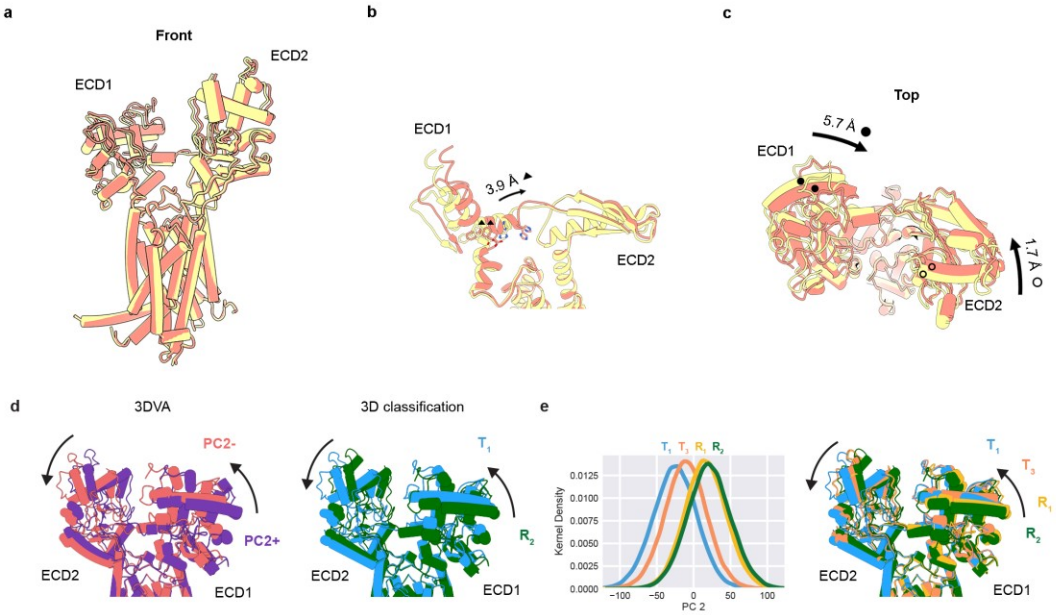


Figure 3.9 Coincident modes of DISP1 conformational flexibility revealed by 3D classification and 3D variability analysis (3DVA)

(A) Overlay comparing front view of R and T conformations (khaki and salmon, respectively) from 3D classification. Major conformational changes are localized to the extracellular domains. (B) Cut-away view showing the formation of a "kink" in the back-side linker of the T conformation, with an accompanying shift of about one helix turn that breaks a hydrogen bond between linker residue H777 and the backbone carbonyl of K767.

(C) Extracellular view of the R and T conformations, highlighting the movement of secondary structure elements in ECD1 (> 5 Å) and ECD2 (~2 Å). The shift of ECD1 and the formation of the inter-ECD linker "kink" appear intimately related. Numbers indicate distances (Å) between the Cα of F772 in R and T (marked by ▲ in (B)) and R382 and S898 (marked by ● and ○, respectively, in (C)).

(D) Comparison of the most extreme R and T conformations from 3D classification (left) and the two extreme ends of PC2, PC2+ (R-like) and PC2- (T-like), from 3DVA (right) shows that a nearly identical mode of motion is captured by both techniques. **e**, Distributions of DISP1-A particles stratified by their 3D class along the PC2 axis of 3DVA, demonstrating colinearity between the 3DVA trajectory and the 3D subclasses ordered by apparent conformation. T_1 and T_3, and R_1 and R_2, respectively, are subclasses of T and R conformations resolved by 3D classification (see **Figure 3.7**).

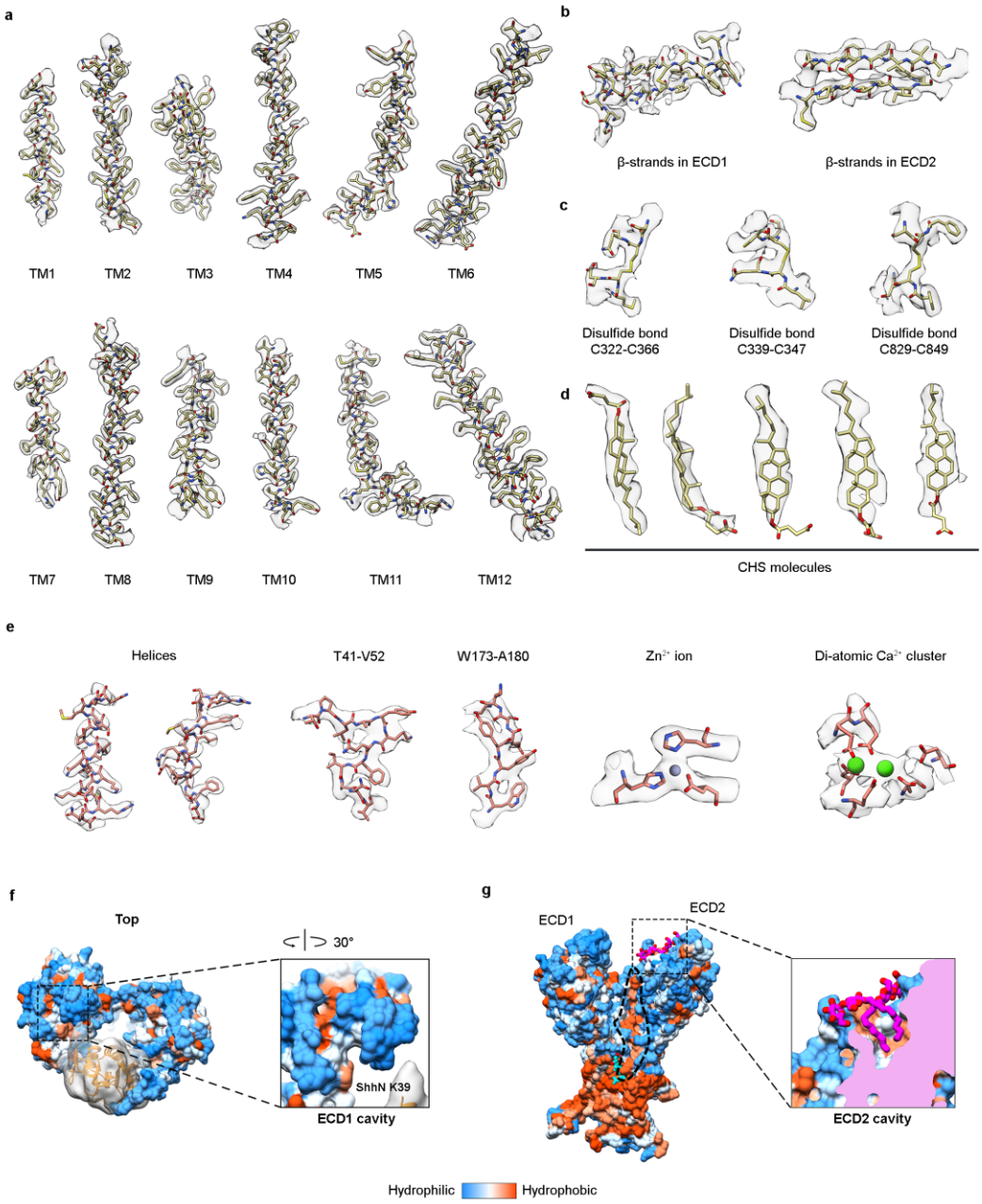


Figure 3.10 Representative cryo-EM densities from selected structural features

(A)-(D) Representative cryo-EM densities from 3D reconstruction of DISP1-A, conformation R. (A) Densities of all transmembrane helices; (B) Representative densities of a beta-sheet from ECD1 and ECD2. (C) Cryo-EM densities of three representative disulfide bonds. (D) Cryo-EM densities of five representative CHS molecules. (E) Representative cryo-EM densities in ShhN from 3D reconstruction of ShhN:DISP1-A complex.

(F)-(G) Representation of surface hydrophobicity, viewed from top ((F), left panel) or a top-front position of DISP1-A ((G), right panel). Close-up views of hydrophobic cavities in both ECD1 (left panel) and ECD2 ((G), right panel). A hydrophobic track beginning near the front-side lifted sterol (cyan) extends outward from the membrane (dotted shape). The enclosed hydrophobic conduit employed by the PTCH1 protein for sterol movement away from the membrane is formed by the

juxtaposition of ECD1 and ECD2. In DISP1, although ECD1 and ECD2 are split apart in a manner that would bisect this conduit, the ECD2 portion retains a series of hydrophobic residues that line its inner surface, which could perhaps form a partial hydrophobic conduit to the hydrophobic cavities near the distal tips of ECD1 and ECD2, analogous to the sterol conduit within the center of the conjoined ECDs of PTCH1.

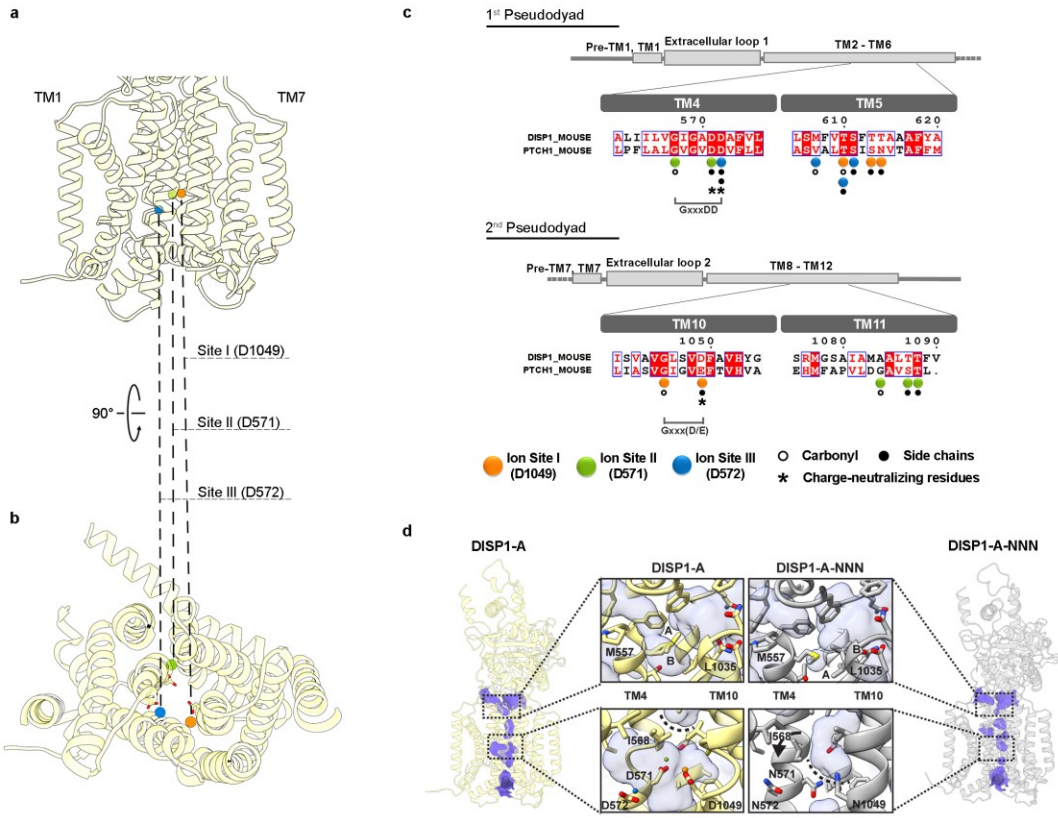


Figure 3.11 Na^+-coordinating amino acid residues in DISP1, and conformational rearrangements within the transmembrane Na^+ pathway in DISP1-A-NNN

(A)-(B) Side and extracellular views of three Na+ ion binding sites within the transmembrane domain, each labeled with its associated charge-neutralizing intra-membrane Asp.

(C) Residues in DISP1 with carbonyl oxygens or side-chain residues that coordinate Na+ and carboxylate residues that neutralize Na+ charge are conserved in PTCH1 (see also **Figure 3.15**).

d, Solvent excluded surfaces (1.4 Å probe radius) reveal transmembrane cavities within DISP1-A (left) that are altered in DISP1-A-NNN (right). Two extracellular branches (upper insets) that provide access to the central channel can be individually opened or closed depending on alternate conformations of M557 (“A” and “B” in DISP1-A) and alternate conformations of L1035 (“A” and “B” in DISP1-A-NNN). Significant rearrangements also take place around a mid-membrane water-filled cavity containing the three ion coordination sites (lower insets). In DISP1-A, short kinks in TM4 and TM10 position I568 above this cavity. In DISP1-A-NNN, however, TM4 and TM10 straighten, rotating I568 and D571N and D1049N into this cavity, dramatically reducing its volume and isolating the channel from its intracellular exit (see Supplementary discussion).

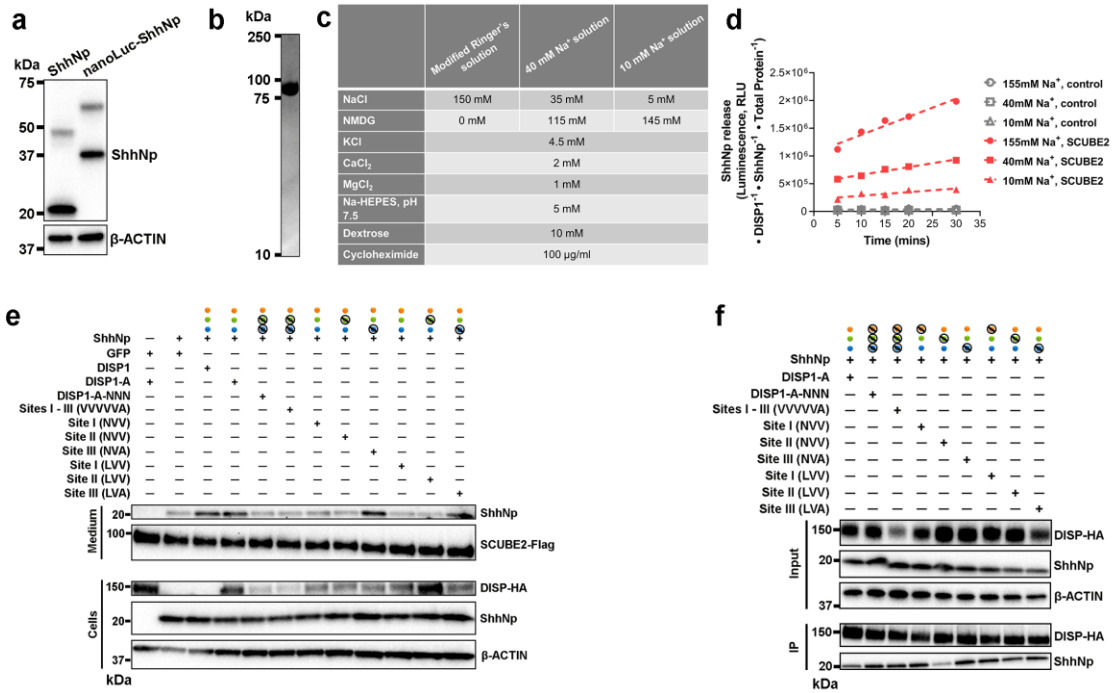


Figure 3.12 Quantitative luciferase-based assay of ShhNp release

(A) Inserting Nanoluciferase coding sequence between E131 and D132 of Shh does not affect Shh autoprocessing and associated lipid modification. Immunoblotting of ShhNp detects both precursor and processed forms of the protein. Most of the expressed nanoluciferase-inserted Shh is in the processed form.

(B) Coomassie blue staining of purified SCUBE2 (Δ30-281) protein.

(C) Composition of buffers used in this study (see **Figure 3.2B-C**).

(D) Time course of NanoLuc-ShhNp release at Na^+ concentrations in (C), with or without SCUBE2 (Δ30-281, 1µM). One representative set of normalized data (see Methods) with linear regression (dashed lines) is shown for each condition (n = 4 biologically independent experiments). The release rate determined as the slope of the linear regression line is presented in **Figure 3.2B**.

(E) Western blot-based ShhNp release assay (see Methods, ShhNp release endpoint assay (Western blot)). Culture media and cell lysates from transiently transfected Disp-/- MEFs were probed by immunoblotting for DISP-HA (SBP and HA-tagged at the C-terminus), ShhNp, and SCUBE2. β-ACTIN, loading control.

(F) DISP-ShhNp binding assay using HEK293 with stably integrated constructs for doxycycline-inducible expression of full-length Shh[20]. DISP1-A variants tagged with SBP and HA at the C-terminus (DISP-HA) were immunoprecipitated with Streptavidin resin, and ShhNp detected by Western blot. Alterations in ion site residues as follows: DISP1-A-NNN, D571N/D572N/D1049N; DISP1-A-VVVVVA, T613V/T614V/T1087V/T1088V/T610V/S611A; Site I (NVV), D1049N/T613V/T614V; Site II (NVV), D571N/T1087V/T1088V; Site III (NVA), D572N/T610V/S611A; Site I (LVV), D1049L/T613V/T614V; Site II (LVV), D571L/T1087V/T1088V; Site III (LVA), D572L/T610V/S611A. Panels (A), (B), (E)-(F), show representative results (n = 3 biologically independent replicates, see **Figure 3.5** for gel source data).

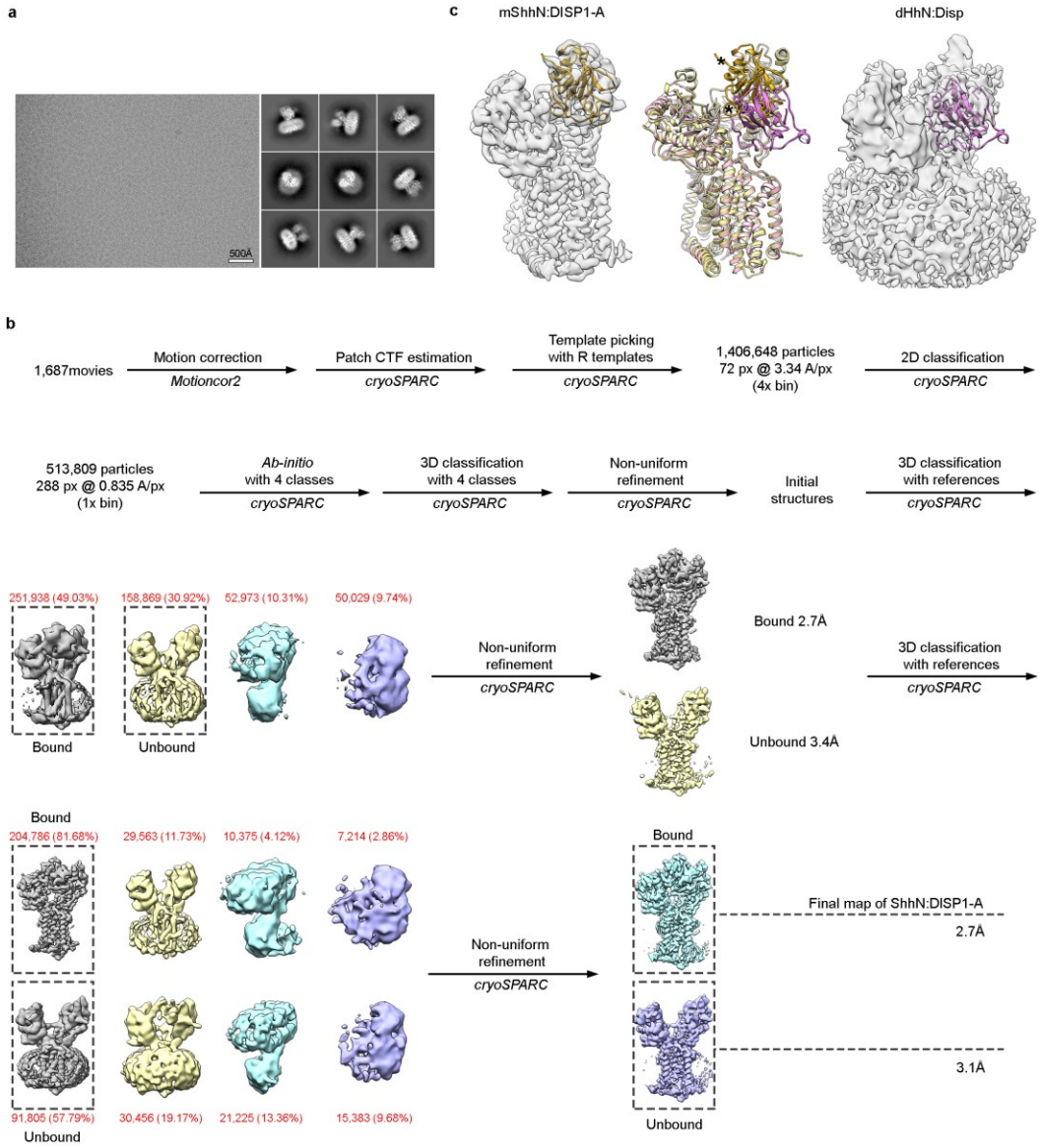


Figure 3.13 Cryo-EM data and image processing flow for ShhN:DISP1-A complex, and comparison to *Drosophila* HhN:Disp complex

(A) A representative cryo-EM micrograph (n = 1687) and highly populated reference-free 2D class averages for ShhN:DISP1-A.

(B) Schematic flow-chart illustrating the image processing used for ShhN:DISP1-A data. Thumbnail images are shown for reference-based 3D classes and high resolution refinements. Dashed black boxes indicate 3D classes selected for the next processing step, with class particle counts in red and refinement GS-FSC resolutions in black. The label, "3D classification with references," indicates that explicit "apo" and "complex" references were used to seed the classification.

(C) Left, the cryo-EM density from our ShhN:DISP1-A complex (transparent grey), overlaid on a ribbon diagram of ShhN (goldenrod). Right, cryo-EM density from the *Drosophila* HhN:Disp complex in Cannac et al.[24] (transparent grey, EMD: 10464), overlaid on the ribbon diagram of HhN (orchid, model extracted from PDB code: 6td6). Middle, superimposition of the

ShhN:DISP1-A and HhN:Disp complex models reported here and in Cannac et al.[24], with mouse DISP1-A and ShhN in khaki and goldenrod, and *Drosophila* Disp and HhN in pink and orchid. The model of *Drosophila* HhN was based on docking of a *Drosophila* HhN structure[44] within a 4.8 Å density map. Relative to *Drosophila* HhN, murine ShhN is translated upwards, away from the membrane, and rotated towards the right. Asterisks indicate corresponding positions near the amino-termini of ShhN and HhN proteins. This difference is somewhat puzzling in light of the ability of mammalian DISP1 to rescue Drosophila disp mutant function[6], and we cannot definitively account for it. One functional difference is that mammalian DISP1 cooperates with SCUBE2 for its Hedgehog-releasing activity, whereas *Drosophila* lacks a Scube orthologue. The ECD domain split and the Hedgehog interaction with the furin-cleaved linker arm together help explain the requirement for furin cleavage[22] in DISP1 function.

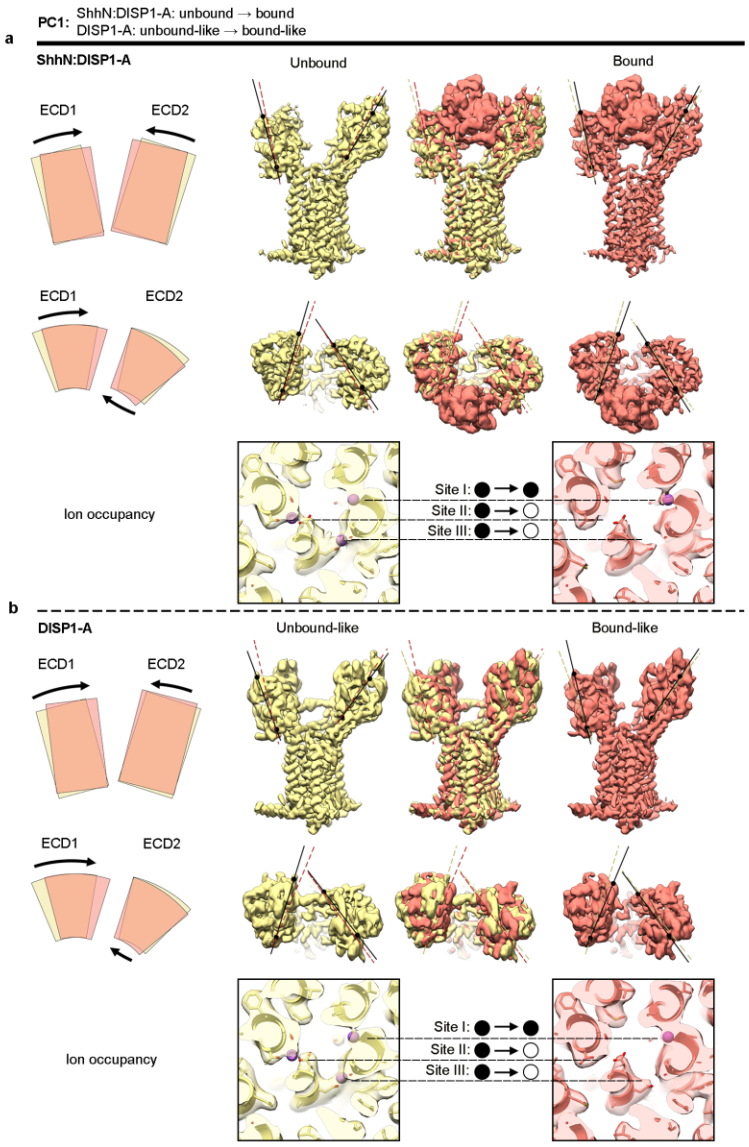


Figure 3.14 Conformational dynamics link intramembrane Na^+ site occupancy to Hedgehog release

(A) Three-dimensional variability analysis (3DVA) of the ShhN:DISP1-A complex dataset reveals a conformational series with ShhN bound or absent at opposite ends of the first principal component (PC1). Side and top views of reconstructed densities from the extremes of PC1 (unbound, khaki; bound, salmon) are shown, with a superimposed view of these extremes in the center. The overall changes in ECD position are illustrated by lines drawn atop the reconstructed densities and by schematized diagrams to the left. Reconstructed densities at the Na+ coordination sites in the transmembrane domain indicate that these sites change from fully occupied in the unbound state to site I only occupied in the bound state. See also Movie 3.4.

(B) 3DVA analysis of the apoprotein preparation shows a similar conformational series along PC1, along with similar shifts in Na+ site occupancy.

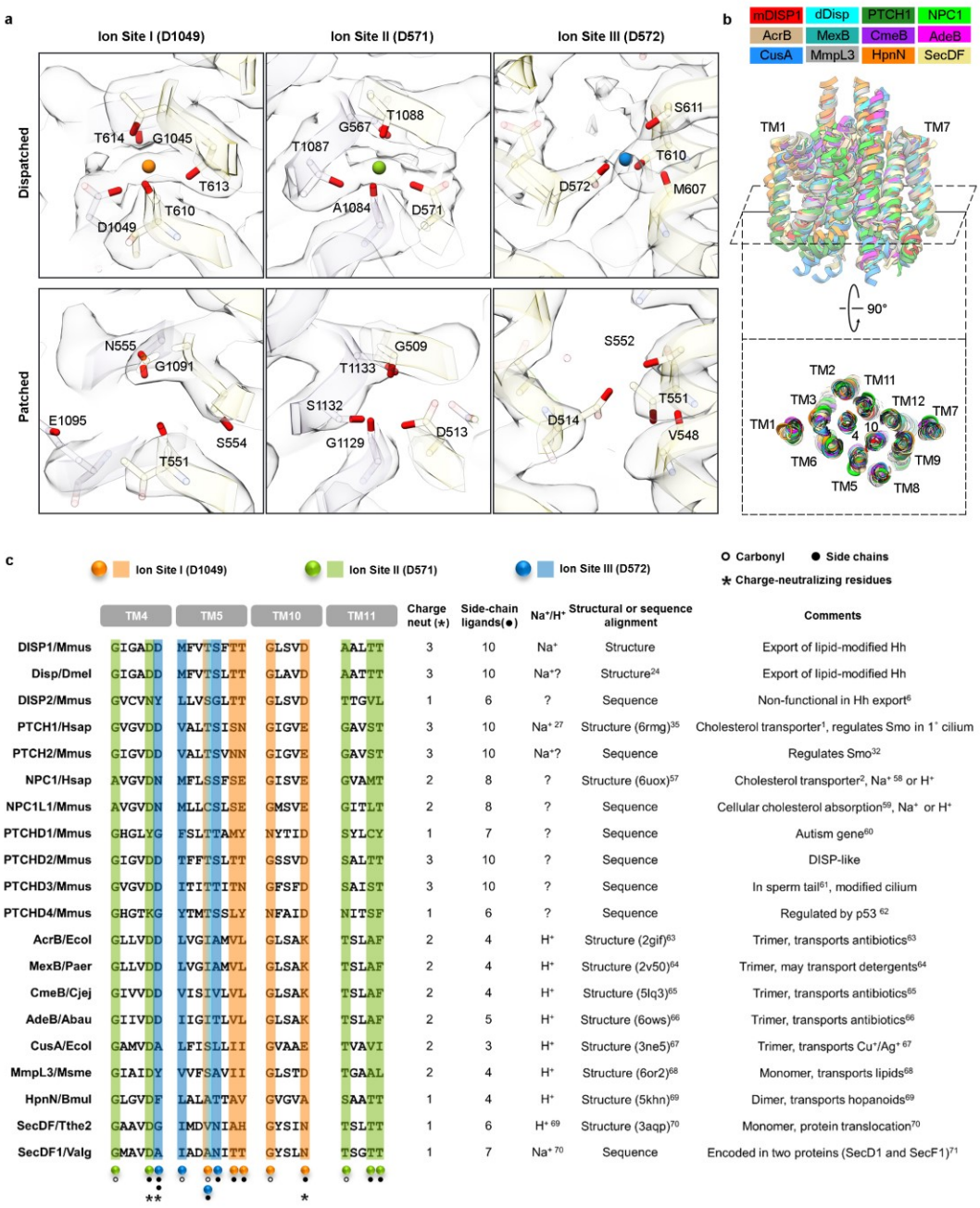


Figure 3.15 Na^+ ion utilization by DISP1, PTCH1, and other members of the RND transporter family

(A) Close-up views of Na^+ ion sites I, II, and III in DISP1, showing the locations of liganding oxygens from amino acid side chains and main-chain carbonyls, and of the corresponding locations in PTCH1[35] (PDB code: 6rmg), based on structural alignment of the two proteins. Note the presence in both proteins of a charge-neutralizing acidic residue at each site, and the conservation of side-chain oxygens as ligands.

(B) Structural alignment, using Chimera Matchmaker, of RND family members, including DISP1, PTCH1, NPC1, and several prokaryotic RND transporters.

(C) Tabulated conservation of charge-neutralizing residues (3 total) or oxygen ligands from amino acid side-chains (10 total) in the indicated proteins, aligned from structure (as in (B), see PDB

codes from the second right column) or, if possible without ambiguity, aligned from sequence. Note the close conservation of Na^+-liganding side-chains and charge-neutralizing residues in PTCH1, known to require Na^+ for its activity, and in Disp from *Drosophila melanogaster*. The prokaryotic Na^+-utilizing SecD1/SecF1 peptide translocator from *Vibrio alginolyticus* (encoded as two peptides; aligned by homology to *Thermus thermophilus* SecDF), in contrast, appears to have evolved a distinct mode of Na^+ interaction. See Supplementary discussion. Abbreviations: Mmus, *Mus musculus*; Dmel, *Drosophila melanogaster*; Hsap, *Homo sapiens*; Ecol, *Escherichia coli*; Paer, *Pseudomonas aeruginosa*; Cjej, *Campylobacter jejuni*; Abau, *Acinetobacter baumannii*; Msme, *Mycolicibacterium smegmatis*; Bmul, *Burkholderia multivorans*; Tthe2, *Thermus thermophilus*; Valg, *Vibrio alginolyticus*.

Tables

Table 3.1 Cryo-EM data collection, refinement and validation statistics

	DISP1-A "R"	DISP1-A "T"	DISP1-A-NNN	DISP1-A:ShhN Complex
Data collection and processing				
Magnification	EFTEM 105,000	EFTEM 105,000	EFTEM 105,000	EFTEM 105,000
Voltage (kV)	300	300	300	300
Electron exposure (e-/Å^2)	66.7	66.7	65.5	66.7
Defocus range (μm)	-0.5-1.5	-0.5-1.5	-0.5-1.5	-0.5-1.5
Pixel size (Å)	0.835	0.835	0.834	0.835
Symmetry imposed	C1	C1	C1	C1
Initial particle images (no.)	3,717,941	3,717,941	3,489,403	1,406,648
Final particle images (no.)	166,203	154,908	55,982	204,786
Map resolution (Å)	2.49	2.54	3.20	2.71
FSC threshold	0.143	0.143	0.143	0.143
Map resolution range (Å)				
Refinement				
Initial model used (PDB code)	*Ab initio*	*Ab initio*	*Ab initio*	*Ab initio*
Model resolution (Å)	2.64	2.73	4.25	2.83
FSC threshold	0.5	0.5	0.5	0.5
Model resolution range (Å)				
Map sharpening *B* factor (Å^2)	-69.9	-72.1	-116.7	-105.6
Model composition				
Non-hydrogen atoms	8,573	8,526	7,926	9,994
Protein residues	926	926	926	1104
Ligands	28	28	14	28
B factors (Å^2)				
Protein	54/133/76	30/140/87	30/265/180	30/147/77
Ligand	61/114/88	72/140/95	74/250/101	30/120/93
R.m.s. deviations				
Bond lengths (Å)	0.056	0.063	0.048	0.058
Bond angles (°)	2.409	2.316	2.049	2.200
Validation				
MolProbity score	1.02	0.81	1.44	0.87
Clashscore	2.22	1.05	3.26	0.60
Poor rotamers (%)	0.61	0.12	0.98	0.21
Ramachandran plot				
Favored (%)	97.93	98.04	95.42	96.99
Allowed (%)	2.07	1.96	4.47	3.01
Disallowed (%)	0	0	0.11	0

Supplemental Files

Movie 3.1 Conformational dynamics associated with R-T transition (PC2)
The structural changes associated with the transition between DISP1-A R and T subclasses are visualized using PC2 of 3DVA in cryoSPARC. The inset histogram of PC scores (X-axis) for normalized number of particles (Y-axis) assigned to R_1, R_2, or T_1, T_3 shows the quantitative relationship between this variance component and the standard 3D classification of particles presented in **Figure 3.9**.

Movie 3.2 Overview of the 28 lipids or detergent molecules associated with DISP1-A
A movie displaying DISP1-A (pale ribbon) with twenty-six CHS (blue or cyan) and two LMNG molecules (magenta) resolved in the structure. The front lifted CHS is colored cyan (see text).

Movie 3.3 Identification of a Na^+ permeation channel in DISP1-A
We searched for tunnels using the CAVER 3 PyMOL plugin with a probe of radius 1.0 Å, and identified a continuous pathway through the transmembrane domain of DISP1-A, shown as a purple surface. This channel is occupied by resolved water molecules, shown as red spheres, and provides access to the transmembrane cavity that contains the Na^+ ions and water. Finally, inspection of the exterior solvent-excluded surface of DISP1-A, computed with 1.4 Å probe radius in UCSF ChimeraX and shown in khaki, shows the extra- and intracellular openings through which water (and Na^+) may pass through the transmembrane domain.

Movie 3.4 DISP1 protein conformation and Hedgehog binding/release are linked to Na+ site occupancy
Animation of the first principal component identified by 3D variability analysis of ShhN:DISP1-A cryo-EM data, containing bound and unbound DISP1-A particles. In three synchronous views, the left shows a wide-angle view from the front of DISP1 (khaki), with bound ShhN (goldenrod) appearing in coordination with tensing of DISP1 ECDs. The center view highlights at a lower density threshold the DISP1 residues preceding the furin cleavage site, which form a one-armed embrace of ShhN upon binding (hot pink). This ECD domain split and the Hedgehog interaction with the furin-cleaved linker arm together help explain the requirement for furin cleavage in DISP1 function[24]. At right, a tilted, cut-away top view illustrating loss of density in transmembrane Na+ coordination sites II & III, again coincident with bound ShhN.

Methods

No statistical methods were used to predetermine sample size. All proteins, including DISP1 variants, ShhN, ShhNp and SCUBE2, were of mouse origin (*Mus musculus*). Cell lines were not tested for mycoplasma contamination. All cell lines grew normally.

DISP1 expression and purification

All constructs were cloned using Gibson Assembly. In DISP1-A, DISP1-A-NNN, and the ion-site mutants, amino acids 3 to 171 were deleted from the full-length (1-1521) DISP1. For protein expression, DISP1-A:ShhNp binding assays and ShhNp release assays, different DISP1 variants were cloned into the pcDNAh vector[1].

For detection of protein expression levels in Extended Data Fig. 1d, HEK293T cells (ATCC, CRL-3216) were seeded into a 6-well plate. Plasmids encoding DISP1 or DISP1-A with the SBP and HA tags at the C-terminus were transiently transfected into the cells using Lipofectamine 3000 (ThermoFisher). Two days after transfection, the cells were lysed and incubated for 1 hour at 4℃ in buffer containing 20 mM HEPES pH 7.5 (Teknova), 300 mM NaCl (Teknova), 2 mg/ml iodoacetamide (Sigma), 1% DDM (Anatrace) /0.2% CHS (Anatrace) supplemented with protease inhibitors (ThermoFisher). Protein extracts were cleared by a subsequent centrifugation (100,000 xg, 30 min, 4℃). Western blotting was performed using Criterion TGX stain-Free precast gels (Bio-rad) and Immobilon-P PVDF membrane (Millipore). PVDF membranes were immunoblotted overnight at 4℃ with anti-HA rabbit polyclonal (1:1000, Cell signaling) or anti-β-ACTIN rabbit polyclonal (1:1000, Cell signaling) antibodies followed by a 1-hour incubation with the corresponding HRP-conjugated secondary antibodies (1:2000, Promega) at room temperature. The blots were developed using SuperSignal West Pico PLUS substrate (ThermoFisher).

For large-scale protein purification, C-terminally SBP-tagged DISP1-A was cloned into a BacMam expression vector, PVLAD6 vector. Baculoviruses were produced in Sf9 cells (ATCC, CRL-1711) as previously described[1], [27]. HEK293F cells (ThermoFisher, R79007) were cultured in suspension using Free-Style medium (ThermoFisher) until the density reached 1.5 – 2 $\times 10^6$ cells/ml, and cells were transduced with DISP1-A-SBP baculoviruses in the presence of 10 mM sodium butyrate (Sigma). Forty-eight hours after transfection, cells were collected and stored at -80℃. For DISP1-A-NNN, plasmids encoding DISP1-A-NNN-SBP were transfected into the HEK293F cells using linear polyethylenimines (Polysciences). Both DISP1-A and DISP1-A-NNN were purified essentially as previously described[1]. Cells were disrupted using a hypotonic buffer (20 mM HEPES pH 7.5, 10 mM MgCl2 (Teknova), 10 mM KCl (Teknova), 0.25 M sucrose (Sigma)) in the presence of benzonase (Sigma) and protease inhibitors. Following centrifugation (100,000 xg, 30 min, 4℃), the pellets were solubilized using a glass Dounce tissue homogenizer in the lysis buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 2 mg/ml iodoacetamide, 1% DDM /0.2% CHS) supplemented with protease inhibitors. The sample was incubated for 2 hours at 4℃ with gentle rotation, then centrifuged (100,000 xg, 30 min, 4℃) for collection of the supernatant. The C-terminally SBP-tagged DISP1-A or DISP1-A-NNN was captured using Streptavidin-agarose affinity resin (ThermoFisher). After a three-hour incubation period at 4℃, the resin was packed into a disposable column by gravity flow. The resin was first washed with 30 column volumes (CVs) of wash buffer (buffer W, 20 mM HEPES pH 7.5 and 300 mM NaCl) containing 0.03% DDM /0.006% CHS; then with 30 CVs of buffer W containing LMNG (Anatrace) /CHS (0.12%/0.024%) and GDN (0.025%, Anatrace) to exchange detergents; and finally with 30 CVs of buffer W containing 0.012% LMNG / 0.0024% CHS and 0.0025% GDN. Protein was eluted in buffer W containing 0.012% LMNG / 0.0024% CHS, 0.0025% GDN and 2.5 mM biotin (Sigma).

The eluted protein was concentrated by an Amicon Ultra filter (100 kDa cutoff, Millipore) and loaded onto Superose 6 Increase 10/300 GL column (GE Healthcare) in buffer W containing 0.012% LMNG / 0.0024% CHS and 0.0025% GDN plus 5 mM $CaCl_2$ (Sigma).

ShhN expression and purification

DNA fragment encoding mouse Sonic hedgehog residues 26-189 was cloned into a pHTSHP vector with N-terminal His_8 tags.[31] Protein was expressed in *E.* coli strain BL21(DE3) and purified as previously described in ref. [1]. The cell lysate was first centrifuged at 8,000 xg for 15 min (4℃), then the filtered supernatant was applied to the Ni-NTA resin. After elution from the Ni-NTA column, the N-terminal tags were cleaved with HRV3C protease. The protein was dialyzed against 20mM HEPES, 100mM NaCl, and 7mM β-mercaptoethanol, and further purified by SP Sepharose cation exchange chromatography.

SCUBE2 expression and purification

Mouse SCUBE2, lacking amino acids 30-281 (to mimic the effect of furin cleavage at an internal site), was 3xFlag-tagged at the C terminus and expressed in Flp-In T-REx 293 cells (ThermoFisher, R78007), as previously described[20]. Cultures were induced with doxycycline and transitioned to serum-free medium (DMEM (ThermoFisher) supplemented with 1x Insulin-Transferrin-Selenium (ThermoFisher), 10 mM HEPES and 1x GlutaMAX and 5 ug/mL doxycycline). After 3-4 days, SCUBE2-conditioned medium was collected and clarified by centrifugation and membrane filtration. After concentrating by ultrafiltration (Vivaflow 200/30 kDa, Sartorius), SCUBE2 was purified by heparin and anti-FLAG M2 affinity chromatography. Eluted protein was concentrated to 9.13 mg/mL by ultrafiltration (Amicon Ultra 30 kDa, Millipore). 3x Flag peptide was washed out during the final concentration step.

Cryo-EM sample preparation and data acquisition

For DISP1-A, 2.5 µl of purified and concentrated sample at a concentration of 1.56 mg/ml was applied to glow-discharged gold grids covered with holey carbon film (Quantifoil, 300 mesh 1.2/1.3) and plunge frozen using a Vitrobot Mark IV, with blotting time of 4 seconds and blotting force 0, at 8˚C and 100% humidity. DISP1-A-NNN cryo-EM grids were prepared similarly, but using gold-grids covered by a holey gold film (Quantifoil UltrAufoil, 300 mesh 1.2/1.3) and a sample concentration of 0.3 mg/ml. For ShhN:DISP1-A complex, purified DISP1-A was incubated with ShhN at a 2.7 : 1 molar ratio (27µM ShhN, 10µM DISP1-A) for 3 hours (4 ˚C) before cryo-sample preparation. DISP1-A:ShhN grids were prepared in the same way as that for DISP1-A.

Cryo-EM data for DISP1-A and ShhN:DISP1-A were collected with a Titan Krios microscope operated at 300keV and equipped with a Bio Quantum post-column energy filter with zero-loss energy selection slit set to 20eV. Using SerialEM[45], images were recorded automatically at a nominal magnification of 105,000X (true magnification 59,880X), with defocus range set to -0.5 to -1.5 µm. Images were recorded with a post-GIF Gatan K3 camera in super-resolution single electron counting mode with a calibrated pixel size of 0.4175 Å (physical pixel size of 0.835 Å). The DISP1-A-NNN samples were imaged on a different Titan Krios with identical parameters, but a slightly different calibrated pixel size of 0.417 Å. Data collection statistics are shown in Extended Data Table 1.

Image processing and map calculation

As part of an on-the-fly image processing pipeline, which processes micrographs immediately after acquisition, we used MotionCor2[46] to correct motion of each dose fractionated movie stack, sum motion corrected frames with and without dose weighting, and bin images in Fourier space to the physical pixel size (~0.835 Å) for further image processing. Motion-corrected,

dose-weighted sums were used for contrast transfer function determination and resolution estimation in cryoSPARC[47]. Particles were picked in cryoSPARC first using a Gaussian template, and then re-picked with low-resolution templates generated from an initial structure. Subsequent image processing was carried out in RELION[48], and cryoSPARC. The flow-chart of image processing is illustrated in Extended Data Fig. 3 (DISP1-A and DISP1-A-NNN) and Extended Data Fig. 9 (ShhN:DISP1-A). Conversion of data from cryoSPARC to RELION was performed using UCSF pyem[49]. Directional Fourier shell correlation curves were determined as previously described[43].

3DVA[23] of DISP1-A alone was conducted based on the early non-uniform refinement shown in Extended Data Fig. 3 (2.4 Å nominal resolution, 974,614 particles), using a five-component decomposition solved over 40 iterations. During 3DVA, the data were filtered by an eighth-order Butterworth high-pass filter with a cutoff frequency of 40 Å and a 1.5-order Butterworth low-pass filter with a cutoff frequency of 3.5 Å, as well as a soft mask containing the entire protein but excluding the detergent micelle. Trajectories representing each variability component were computed using "simple" mode (i.e. each step is calculated by weighted addition of variability component maps to the mean structure) with 20 frames and no filters. 3DVA of ShhN:DISP1-A was performed similarly, based on non-uniform refinement of all apo and complex particles combined (2.7 Å nominal resolution, 356,610 particles), but with no high-pass filter due to the need to capture the appearance and disappearance of ShhN.

Model building and refinement

The *de novo* atomic models of DISP1-A were built in Coot[50], refined in PHENIX[51] and finalized in ISOLDE[52]. Starting from a predicted model from PHYRE2[53] and a combined density of R and T conformers, we manually fitted transmembrane helices into the corresponding

density and then traced the backbone of ECDs and linkers in between those transmembrane helices in Coot. The preliminary model was then mutated based on Mouse DISP1 sequence and refined using real space refinement in PHENIX. Based on residual unmodeled density near the protein, ligands were added and refined in real space using Coot and PHENIX, respectively. Starting from this "common model" built referring to the combined density of R and T, we obtained models for each conformer (R, T and NNN) by refining this "common model" against the final density of each conformer individually, removing atoms unresolved in each density and then finalizing the models with ISOLDE (largely to improve steric clashes). The ShhN:DISP1-A complex was treated in the same fashion, but using several crystal structures of mammalian Shh as initial models (PDB codes: 1vhh, 4c4n, and 6pjv). All models were validated by wwPDB validation server[54] and no major issues were reported in our three models.

Visualizations of the atomic models (figures) were made using UCSF Chimera[55], ChimeraX[56] and PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.). The ion conduction path was mapped using Caver 3[29] and ion coordination geometries analyzed with the *CheckMyMetal* webserver[26]. DISP1-A:ShhN interaction interface area and the solvation free energy gain upon the formation of the interfaces were calculated using PISA[57]. Clustal Omega[41] and the ESPript server[42] was used in preparation of the protein sequence alignment.

Rapid, quantitative ShhNp release rate assay

Disp$^{-/-}$ mouse embryonic fibroblasts (MEFs)[6], [20], were cultured on 24-well plates in DMEM supplemented with 10 % FBS and 1% Penicillin-Streptomycin-Glutamine (ThermoFisher). DISP1-mEGFP variants were transfected into the cells along with a Shh-NanoLuc fusion in which the nanoluciferase coding sequence was inserted between E131 and

D132 in Shh (NanoLuc-ShhNp). The mass ratio of plasmids expressing DISP1 variants and NanoLuc-ShhNp was 1 to 2. An EGFP expression construct was used as a negative control for conditions lacking DISP1 variants. Two days after transfection, following a quick wash with DMEM, cells were incubated with cycloheximide (100 μg/mL, Sigma) in DMEM for 50 min. Cells were then incubated in the desired medium either with or without SCUBE2 protein (1 μM (Fig.2b) or 150 nM (Fig. 2c), Extended Data Fig. 8c) at 37℃ with gentle shake (125 rpm). Small aliquots of supernatant were removed sequentially at multiple time points, and collected supernatants were cleared by a centrifugation at 3000 xg for 30 min, 4℃. NanoLuc-Shh luminescence level was measured using Nano-Glo® Luciferase Assay System (Promega), following the manufacturer's protocol, and the measured luminescence was corrected to account for the decreased supernatant volume through time. At the end of the experiment, cells were washed with PBS and lysed with the passive lysis buffer (Promega). Released NanoLuc-Shh luminescence at each time point was normalized to the GFP level, the total protein in cell lysates (measured using Pierce™ BCA protein assay kit, ThermoFisher), and to the total luminescence of Shh (luminescence level in the cell lysates plus the total NanoLuc-ShhNp released into supernatant by the final time point). The linear portion of the release curves was used to calculate the ShhNp release rate.

ShhNp release endpoint assay (Western blot)

Disp$^{-/-}$ MEFs[6], [20], were cultured on 6-well plates in DMEM supplemented with 10 % FBS and 1% Penicillin-Streptomycin-Glutamine. All the cell transfections were performed using Lipofectamine 3000. Plasmids encoding DISP1 variants were transfected into the cells along with full-length Shh constructs. The ratio of plasmids expressing DISP1 variants and Shh was 1 to 2. For conditions that did not have DISP1 variants or Shh, GFP constructs were used as a negative control to replace the missing partner. Two days after transfection, following a quick wash with

DMEM, the cells were shifted to SCUBE2 (mouse SCUBE2, with an N-terminal truncation, Δ30-281) conditioned medium or control medium (600 µl/well). After 16 hours of incubation, the medium was collected and centrifuged at 18,000 xg for 30 min (4℃). The supernatant was probed by immunoblotting for ShhNp (anti-Shh rabbit monoclonal, 1:1000), 3xFlag-tagged SCUBE2 (anti-Flag m2 mouse monoclonal, 1:1000, sigma). Cells were lysed using a buffer containing 20 mM HEPES pH 7.5, 300 mM NaCl, 2 mg/ml iodoacetamide, 1% DDM /0.2% CHS supplemented with protease inhibitors (400 µl/well, 1 hour at 4℃). After a centrifugation of 100,000 xg for 30 min at 4℃, cell lysate was probed by immunoblotting for ShhNp (anti-Shh rabbit monoclonal, 1:1000, Cell signaling), DISP1-HA or DISP1-mEGFP (anti-HA rabbit polyclonal, 1:1000 or anti-GFP rabbit polyclonal, 1:2000, Invitrogen), and β-ACTIN (anti-β-ACTIN rabbit polyclonal, 1:1000).

SCUBE2 conditioned medium was prepared in the same way as described above in SCUBE2 expression and purification. Control medium was prepared as for SCUBE2 conditioned medium but without doxycycline induction.

DISP1-ShhNp co-IP (qualitative binding assay)

The DISP1-ShhNp binding assay was conducted using the HEK293 Flp-In T-REX ShhNp-producing cell line[20]. ShhNp-producing cells were maintained in DMEM with 10 % FBS and 1% Penicillin-Streptomycin-Glutamine. Plasmids of C-terminally HA-tagged DISP1-A variants were transiently transfected into the cells using Lipofectamine 3000. One day after transfection, full-length Shh expression was induced by 5 µg/ml Doxycycline (Sigma) in DMEM F-12 (ThermoFisher) supplemented with 1% Insulin-Transferrin-Selenium (ThermoFisher) and 1% Penicillin-Streptomycin-Glutamine. Another day after ShhNp expression was induced, cells were lysed and incubated for 1 hour at 4℃ in the same buffer as that in ShhNp release endpoint assay.

Following a subsequent centrifugation (100,000 xg, 30 min, 4℃), Streptavidin-agarose affinity resin was used to pull down the SBP-tagged DISP1-A variants, together with any bound ShhNp, from the supernatant. After a two-hour incubation period at 4℃, the resin was washed with buffer W containing 0.03% DDM /0.006% CHS. Protein was eluted in the same buffer supplemented with 2.5 mM biotin. Immunoprecipitates were analyzed by Western blot as described above using the following antibodies: anti-HA rabbit polyclonal (1:1000), anti-Shh rabbit monoclonal (1:1000) and anti-β-ACTIN rabbit polyclonal (1:1000).

Data availability

The accession numbers for DISP1-A R conformation, DISP1-A T conformation, DISP1-A-NNN, and ShhN:DISP1-A complex models reported in this paper are PDB: 7RPH, 7RPI, 7RPJ, and 7RPK, respectively. The maps of DISP1-A R conformation, DISP1-A T conformation, DISP1-A-NNN, and ShhN:DISP1-A complex have been deposited in EMDB under accession codes EMD-24614, EMD-24615, EMD-24616, and EMD-24617, respectively. Further information and requests for data should be directed to the corresponding authors. PDB codes of all the cited structures are provided either in the related figure legends or in the Methods section.

Acknowledgements

We thank K. Roberts and A. Kershner for technical assistance and advice, and V. Korkhov and X. Li for sharing data and information prior to publication. Y.C. is an investigator of the Howard Hughes Medical Institute. Supported by NIH grants R01GM102498, Stanford Department of Urology, and Ludwig Cancer Institute (to P.A.B.) and R35GM140847, S10OD020054 and S10OD021741 (to Y.C.).

Author Contributions

Q.W. purified and characterized DISP1-A, DISP1-A-NNN and ShhN:DISP1-A complex, SCUBE2 protein, and developed and performed ShhNp binding and release assays. D.E.A. prepared cryo-EM grids, collected and processed cryo-EM data, and performed the 3D variability analysis. R.M. devised and aided the SCUBE2 purification. J.H. performed ShhNp release assays. K.D., D.E.A., Y.Z., and Q.W. built the models. Y.M. provided mouse genetic epistasis analysis while in the laboratory of P.A.B. at Johns Hopkins University School of Medicine. All authors participated in discussion and analysis of the data. P.A.B., Q.W., D.E.A., K.D., Y.C., and Y.Z. prepared the manuscript.

References

[1] Y. Zhang *et al.*, "Structural basis for cholesterol transport-like activity of the Hedgehog receptor Patched," *Cell*, vol. 175, no. 5, pp. 1352-1364.e14, 2018, doi: 10.1016/j.cell.2018.10.026.

[2] E. D. Carstea *et al.*, "Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis," *Science*, vol. 277, no. 5323, pp. 228–31, 1997.

[3] H. Nikaido and Y. Takatsuka, "Mechanisms of RND multidrug efflux pumps," *Biochim. Biophys. Acta*, vol. 1794, pp. 769–781, 2009, doi: 10.1016/j.bbapap.2008.10.004.

[4] A. Yamaguchi, R. Nakashima, and K. Sakurai, "Structural basis of RND-type multidrug exporters," *Front. Microbiol.*, vol. 6, p. 327, 2015, doi: 10.3389/fmicb.2015.00327.

[5] R. Burke *et al.*, "Dispatched, a novel sterol-sensing domain protein dedicated to the release of cholesterol-modified Hedgehog from signaling cells," *Cell*, vol. 99, no. 7, pp. 803–815, 1999, doi: 10.1016/s0092-8674(00)81677-3.

[6] Y. Ma *et al.*, "Hedgehog-mediated patterning of the mammalian embryo requires transporter-like function of Dispatched," *Cell*, vol. 111, no. 1, pp. 63–75, 2002, doi: https://doi-org.stanford.idm.oclc.org/10.1016/S0092-8674(02)00977-7.

[7] T. Kawakami, T. N. Kawcak, Y. Li, W. Zhang, Y. Hu, and P. Chuang, "Mouse dispatched mutants fail to distribute hedgehog proteins and are defective in hedgehog signaling," *Development*, vol. 129, pp. 5753–5765, 2002, doi: 10.1242/dev.00178.

[8] E. Roessler *et al.*, "Truncating loss-of-function mutations of DISP1 contribute to holoprosencephaly-like microform features in humans," *Hum. Genet.*, vol. 125, pp. 393–400, 2009, doi: 10.1007/s00439-009-0628-7.

[9] T. Caspary *et al.*, “Mouse dispatched homolog1 is required for long-range, but not juxtacrine, Hh signaling,” *Curr. Biol.*, vol. 12, no. 18, pp. 1628–1632, 2002.

[10] Y. Nakano, H. R. Kim, A. Kawakami, S. Roy, A. F. Schier, and P. W. Ingham, “Inactivation of dispatched 1 by the chameleon mutation disrupts Hedgehog signalling in the zebrafish embryo,” *Dev. Biol.*, vol. 269, pp. 381–392, 2004, doi: 10.1016/j.ydbio.2004.01.022.

[11] R. D. Riddle, R. L. Johnson, E. Laufer, and C. Tabin, “Sonic hedgehog mediates the polarizing activity of the ZPA,” *Cell*, vol. 75, pp. 1401–1416, 1993.

[12] Y. Echelard *et al.*, “Sonic hedgehog, a member of a family of putative signaling molecules , is implicated in the regulation of CNS polarity,” *Cell*, vol. 75, pp. 1417–1430, 1993.

[13] S. Krauss, J.-P. Concordet, and P. W. Ingham, “A functionally conserved homolog of the Drosophila segment polarity gene hh is expressed in tissues with polarizing activity in Zebrafish embryos,” *Cell*, vol. 75, pp. 1431–1444, 1993.

[14] C. Chiang *et al.*, “Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function,” *Nature*, vol. 383, no. 6599, pp. 407–413, 1996.

[15] R. K. Mann and P. A. Beachy, “Novel lipid modifications of secreted protein signals,” *Annu. Rev. Biochem.*, vol. 73, pp. 891–923, 2004, doi: 10.1146/annurev.biochem.73.011303.073933.

[16] J. A. Porter, K. E. Young, and P. A. Beachy, “Cholesterol modification of Hedgehog signaling proteins in animal development,” *Science*, vol. 274, pp. 255–260, 1996, doi: 10.1126/science.274.5285.255.

[17] R. B. Pepinsky *et al.*, “Identification of a palmitic acid-modified form of human sonic Hedgehog,” *J. Biol. Chem.*, vol. 273, no. 22, pp. 14037–14045, 1998.

[18] Z. Chamoun *et al.*, “Skinny Hedgehog, an acyltransferase required for palmitoylation and activity of the Hedgehog signal,” *Science*, vol. 293, no. 5537, pp. 2080–2085, 2001.

[19] I. G. Woods and W. S. Talbot, “The you gene encodes an EGF-CUB protein essential for Hedgehog signaling in Zebrafish,” *Plos Biol.*, vol. 3, no. 3, p. e66, 2005, doi: 10.1371/journal.pbio.0030066.

[20] A. Creanga *et al.*, “Hedgehog signal in soluble form Scube/You activity mediates release of dually lipid-modified Hedgehog signal in soluble form,” *Genes Dev.*, vol. 26, pp. 1312–1325, 2012, doi: 10.1101/gad.191866.112.

[21] H. Tukachinsky, R. P. Kuzmickas, C. Y. Jao, J. Liu, and A. Salic, “Dispatched and Scube mediate the efficient secretion of the cholesterol-modified Hedgehog ligand,” *Cell Rep.*, vol. 2, no. 2, pp. 308–20, 2012, doi: 10.1016/j.celrep.2012.07.010.

[22] D. P. Stewart *et al.*, “Cleavage activates Dispatched for Sonic Hedgehog ligand release,” *eLife*, vol. 7, p. e31678, 2018.

[23] A. Punjani and D. J. Fleet, “3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM,” *J. Struct. Biol.*, vol. 213, no. 2, p. 107702, 2021, doi: 10.1016/j.jsb.2021.107702.

[24] F. Cannac, C. Qi, J. Falschlunger, G. Hausmann, K. Basler, and V. M. Korkhov, “Cryo-EM structure of the Hedgehog release protein Dispatched,” *Sci. Adv.*, vol. 6, no. 16, p. eaay7928, 2020.

[25] H. Chen, Y. Liu, and X. Li, “Structure of human Dispatched-1 provides insights into Hedgehog ligand biogenesis,” *Life Sci. Alliance*, vol. 3, no. 8, pp. 1–8, 2020, doi: 10.26508/LSA.202000776.

[26] H. Zheng, D. R. Cooper, P. J. Porebski, I. G. Shabalin, K. B. Handing, and W. Minor, "CheckMyMetal : a macromolecular metal-binding validation tool," *Acta Crystallogr Struct Biol*, vol. 73, no. Pt 3, pp. 223–233, 2017, doi: 10.1107/S2059798317001061.

[27] B. R. Myers, L. Neahring, Y. Zhang, and K. J. Roberts, "Rapid, direct activity assays for Smoothened reveal Hedgehog pathway regulation by membrane cholesterol and extracellular sodium," *PNAS*, vol. 114, no. 52, pp. E11141–E11150, 2017, doi: 10.1073/pnas.1717891115.

[28] K. Petrov, B. M. Wierbowski, J. Liu, and A. Salic, "Distinct cation gradients power cholesterol transport at different key points in the Hedgehog signaling pathway," *Dev. Cell*, vol. 55, no. 3, pp. 314-327.e7, 2020, doi: 10.1016/j.devcel.2020.08.002.

[29] E. Chovancova *et al.*, "CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures," *PLoS Comput. Biol.*, vol. 8, no. 10, p. e1002708, 2012, doi: 10.1371/journal.pcbi.1002708.

[30] T. M. T. Hall, J. A. Porter, P. A. Beachy, and D. J. Leahy, "A potential catalytic site revealed by the 1.7–Å crystal structure of the amino–terminal signalling domain of Sonic hedgehog," *Nature*, vol. 378, no. 6553, pp. 212–216, 1995, doi: 10.1038/378212a0.

[31] J. S. McLellan, X. Zheng, G. Hauk, R. Ghirlando, P. A. Beachy, and D. J. Leahy, "The mode of Hedgehog binding to Ihog homologues is not conserved across different phyla," *Nature*, vol. 455, no. 7215, pp. 979–983, 2008, doi: 10.1038/nature07358.

[32] O. Zhulyn, E. Nieuwenhuis, C. Y. Liu, S. Angers, and C. Hui, "Ptch2 shares overlapping functions with Ptch1 in Smo regulation and limb development," *Dev. Biol.*, vol. 397, no. 2, pp. 191–202, 2015, doi: 10.1016/j.ydbio.2014.10.023.

[33] I. Deshpande *et al.*, "Smoothened stimulation by membrane sterols drives Hedgehog pathway activity," *Nature*, vol. 571, no. 7764, pp. 284–288, 2019, doi: 10.1038/s41586-019-1355-4.

[34] X. Qi, H. Liu, B. Thompson, J. Mcdonald, C. Zhang, and X. Li, “Cryo-EM structure of oxysterol-bound human Smoothened coupled to a heterotrimeric Gi,” *Nature*, vol. 571, no. 7764, pp. 279–283, 2019, doi: 10.1038/s41586-019-1286-0.

[35] C. Qi, G. Di Minin, I. Vercellino, A. Wutz, and V. M. Korkhov, “Structural basis of sterol recognition by human hedgehog receptor PTCH1,” *Sci. Adv.*, vol. 5, no. 9, p. eaaw6490, 2019.

[36] X. M. Zhang, M. Ramalho-santos, and A. P. McMahon, “Smoothened mutants reveal redundant roles for Shh and Ihh signaling including regulation of L/R asymmetry by the mouse node,” *Cell*, vol. 105, no. 6, pp. 781–792, 2001, doi: https://doi.org/10.1016/S0092-8674(01)00385-3.

[37] L. V. Goodrich, L. Milenkovic, K. M. Higgins, and M. P. Scott, “Altered neural cell fates and medulloblastoma in mouse patched mutants,” *Science*, vol. 277, pp. 1109–1113, 1997.

[38] X. Wang *et al.*, “TPC proteins are phosphoinositide-activated sodium-selective ion channels in endosomes and lysosomes,” *Cell*, vol. 151, no. 2, pp. 372–383, 2012, doi: 10.1016/j.cell.2012.08.036.

[39] D. M. Whalen, T. Malinauskas, R. J. C. Gilbert, and C. Siebold, “Structural insights into proteoglycan-shaped Hedgehog signaling,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 41, pp. 16420–16425, 2013, doi: 10.1073/pnas.1310097110.

[40] Y. Zhang *et al.*, “Hedgehog pathway activation through nanobody-mediated conformational blockade of the Patched sterol conduit,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 46, pp. 28838–28846, 2020, doi: 10.1073/pnas.2011560117.

[41] F. Sievers *et al.*, “Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega,” *Mol. Syst. Biol.*, vol. 7, no. 539, 2011, doi: 10.1038/msb.2011.75.

[42] X. Robert and P. Gouet, “Deciphering key features in protein structures with the new ENDscript server,” *Nucleic Acids Res.*, vol. 42, pp. W320–W324, 2014, doi: 10.1093/nar/gku316.

[43] S. Dang *et al.*, “Cryo-EM structures of the TMEM16A calcium-activated chloride channel,” *Nature*, vol. 552, no. 7685, pp. 426–429, 2017, doi: 10.1038/nature25024.

[44] J. S. McLellan *et al.*, “Structure of a heparin-dependent complex of Hedgehog and Ihog,” *PNAS*, vol. 103, no. 46, pp. 17208–13, 2006.

[45] D. N. Mastronarde, “Automated electron microscope tomography using robust prediction of specimen movements,” *J. Struct. Biol.*, vol. 152, pp. 36–51, 2005, doi: 10.1016/j.jsb.2005.07.007.

[46] S. Q. Zheng, E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng, and D. A. Agard, “MotionCor2 : anisotropic correction of beam-induced motion for improved cryo-electron microscopy,” *Nat. Methods*, vol. 14, no. 4, pp. 331–332, 2017, doi: 10.1038/nmeth.4193.

[47] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, “cryoSPARC : algorithms for rapid unsupervised cryo-EM structure determination,” *Nat. Methods*, vol. 14, no. 3, pp. 290–296, 2017, doi: 10.1038/nmeth.4169.

[48] J. Zivanov *et al.*, “New tools for automated high-resolution cryo-EM structure determination in RELION-3,” *eLife*, vol. 7, p. e42166, 2018.

[49] D. Asarnow, E. Palovcak, and Y. Cheng, “https://doi.org/10.5281/zenodo.3576630.” 2019.

[50] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, “Features and development of Coot,” *Acta Crystallogr. Biol. Crystallogr.*, vol. D66, pp. 486–501, 2010, doi: 10.1107/S0907444910007493.

[51] P. D. Adams *et al.*, “PHENIX : a comprehensive Python-based system for macromolecular structure solution,” *Acta Crystallogr. Biol. Crystallogr.*, vol. D66, pp. 213–221, 2010, doi: 10.1107/S0907444909052925.

[52] T. I. Croll, “ISOLDE : a physically realistic environment for model building into low-resolution electron-density maps,” *Acta Crystallogr Struct Biol*, vol. D74, pp. 519–530, 2018, doi: 10.1107/S2059798318002425.

[53] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg, “The Phyre2 web portal for protein modelling, prediction and analysis,” *Nat. Protoc.*, vol. 10, no. 6, pp. 845–858, 2015.

[54] H. Berman, K. Henrick, and H. Nakamura, “Announcing the worldwide Protein Data Bank,” *Nat. Struct. Biol.*, vol. 10, no. 12, p. 980, 2003.

[55] E. F. Pettersen *et al.*, “UCSF Chimera — A visualization system for exploratory research and analysis,” *J. Comput. Chem.*, vol. 25, pp. 1605–1612, 2004, doi: 10.1002/jcc.20084.

[56] T. D. Goddard *et al.*, “UCSF ChimeraX : meeting modern challenges in visualization and analysis,” *Protein Sci.*, vol. 27, pp. 14–25, 2018, doi: 10.1002/pro.3235.

[57] E. Krissinel and K. Henrick, “Inference of macromolecular assemblies from crystalline state,” *J. Mol. Biol.*, vol. 372, no. 3, pp. 774–797, 2007, doi: 10.1016/j.jmb.2007.05.022.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Daniel Asarnow
BFA285B89A1943A...
Author Signature

9/3/2021
Date