Atlas Host-Pathogen Interaction Report

Logan Hallee[1,2] **and Jason P. Gleghorn**[1,2,3]

[1]Synthyra, PBLLC; [2]Center for Bioinformatics and Computational Biology, University of Delaware; [3]Department of Biomedical Engineering, University of Delaware

Synthyra Technical Report · May 9, 2026

Summary

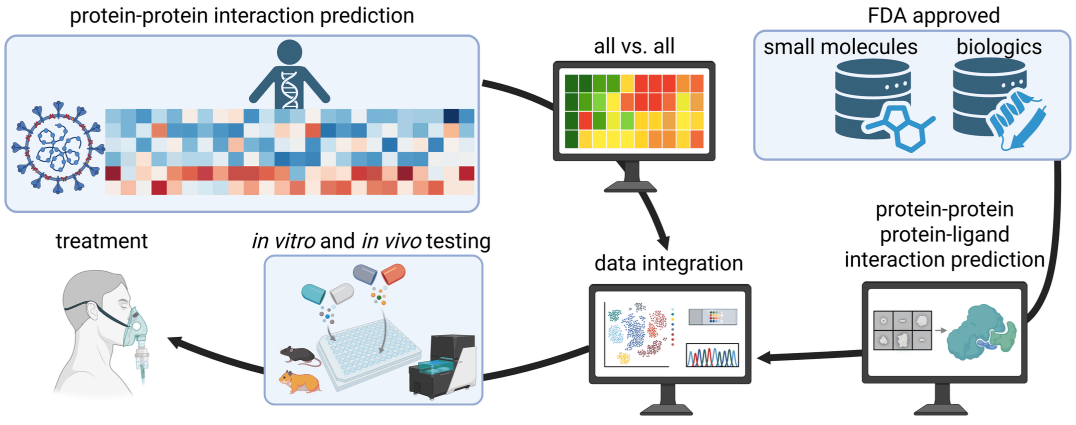
Host-pathogen protein-protein interactions are the molecular events through which viruses and bacteria commandeer human cells, evade immunity, and produce disease. Every major symptom of infection traces back to a pathogen protein interacting with a human one; identifying those contact points reveals both the mechanistic basis of pathology and the most tractable targets for intervention. A 2026 *Nature Machine Intelligence* study found that every tested computational method trained on human protein interaction data scores no better than random chance when asked to predict human-SARS-CoV-2 interactions. This is a gap that has limited both mechanistic discovery and therapeutic development for emerging pathogens. Atlas is a sequence-based protein interaction predictor designed to overcome this gap. Without modification for any specific pathogen, Atlas identifies host-pathogen interaction networks at outbreak speed, enabling both mechanistic discovery and rapid drug repurposing. Here, we demonstrate this dual capability on three pathogens with high clinical relevance: SARS-CoV-2, Andes hantavirus, and Bundibugyo ebolavirus.

Without any coronavirus in its training data, Atlas identified previously uncharacterized SARS-CoV-2 accessory proteins predicted to engage human olfactory receptors at 47-fold above background and mitochondrial machinery in olfactory support cells. These results are consistent with the published mechanism of COVID-19 anosmia and provide a molecular hypothesis for experimental validation.

For Andes hantavirus, Atlas screened the entire FDA-approved drug database against ANDV structural proteins, identifying two candidates with decades of clinical safety and structural validation. This includes one already administered by inhalation from a nebulizer directly to the lung, the primary organ of hantavirus disease. This demonstration shows what rapid outbreak response looks like: from pathogen sequence to prioritized therapeutic candidates in minutes, enabling immediate experimental follow-up.

For Bundibugyo ebolavirus, the agent of an ongoing outbreak with no approved treatment, the same screen was run as a blind recall test. Four FDA-approved monoclonal antibodies bind ebolavirus glycoprotein and nothing else in the viral proteome. Atlas assigned all four to glycoprotein out of nine viral proteins in one minute, without target annotation, and with no score overlap against any non-target protein. The screen also shows where the model fails, missing the best-characterized VP24 interaction outright. Folding the same complexes ranks three certain non-interactions above the correct one, so the two approaches fail on different inputs and their agreement, not either alone, is what should gate an experiment.

Atlas shows that sequence-only protein interaction prediction can generalize across species boundaries without retraining, and that high-throughput screening at outbreak speed is now practical. Applied to pathogens with no approved treatment and active outbreaks, it prioritized FDA-approved drug candidates for immediate experimental validation and recovered a known therapeutic target class blind, providing the concrete output needed by a translational team confronting an emerging infectious disease.



ABOUT SYNTHYRA

Synthyra Public Benefit LLC is dedicated to empowering researchers across medicine, agriculture, and environmental science with access to accessible molecular modeling tools. We aim to accelerate innovations that benefit both science and society through biochemical computation.

Protein-protein interaction | Protein language model | Host-pathogen | Interactome | Drug repurposing | Hantavirus | Ebolavirus | SARS-CoV-2 | COVID-19
Correspondence: ***logan@synthyra.com*** **|** ***gleghorn@udel.edu***

Introduction

Host-pathogen protein-protein interactions (PPIs) are the molecular substrate of infectious disease.[1] Every successful viral or bacterial infection consists of a sequence of physical contacts between pathogen proteins and host proteins: receptor-mediated cell entry, hijacking of cytoskeletal and trafficking machinery, evasion of innate immune sensors, repurposing of host translation and metabolism, and disruption of cell-cell junctions during dissemination. The clinical phenotypes that physicians observe at the bedside, including fever, hypoxia, vascular leak, organ-specific tissue damage, anosmia, and ageusia, are downstream readouts of these interactions playing out at scale across infected tissues. Mapping which pathogen proteins bind which host proteins is therefore simultaneously a mechanistic question, since it explains how a particular pathogen produces a particular phenotype, and a translational one, because the contact surfaces it identifies are the most direct candidates for small-molecule inhibition, antibody neutralisation, and host-directed therapy.

The experimental approaches that established this picture, including affinity-purification mass spectrometry, yeast two-hybrid, BioID, and structural studies of individual complexes, remain the gold standard for any individual interaction, and the SARS-CoV-2 affinity-purification map of Gordon et al. (1) demonstrated that a single coordinated effort can reshape the therapeutic landscape for a novel virus within months of an outbreak. Recent structure-based campaigns have extended this further: the eukaryotic-complex pipeline of Humphreys et al. (2), the human proteome-wide screen of Zhang et al. (3), and the human-pathogen pipeline of Humphreys et al. (4) have used coevolutionary signal and AlphaFold-style folding (5) to enumerate plausible interactions at proteome scale. These methods are accurate where the underlying multiple-sequence alignments are deep, but they are computationally expensive at hours to days per candidate pair on GPU clusters, they require either an MSA or a representative homologous structure, and they degrade on viral accessory proteins, intrinsically disordered regions, and orphan sequences with no detectable homologs. Sequence-only methods that score arbitrary protein pairs in milliseconds, with no MSA, no structure, and no per-species adaptation, offer a practical route to truly proteome-scale screens against novel pathogens, especially under the time pressure of an emerging outbreak.

Sequence-only PPI prediction has nonetheless had a difficult decade. A long line of work, including DeepFE (6), PIPR (7), D-SCRIPT (8), Topsy-Turvy (9), RAPPPID (10), INTREPPPID (11), MINT (12), PLM-interact (13), TUnA (14), C3PI (15), and our own earlier Synteract series (16), has reported steady benchmark gains, but the apparent progress has been repeatedly punctured by methodological audits. Park and Marcotte (17) showed more than a decade ago that pair-input prediction tasks decompose into three nested generalization regimes, namely test pairs that share both proteins with training, test pairs that share exactly one, and test pairs that share neither, and that performance on the easiest regime is almost uninformative about performance on the hardest. Bernett, Blumenthal, and List (18) demonstrated that the standard public PPI benchmarks contain a homology shortcut: when negatives are sampled uniformly from non-positive pairs, models can reach high apparent accuracy by recognising sequence-similarity neighborhoods rather than learning anything about the binding interface. Reim et al. (19) extended this finding, showing that on rigorously homology-controlled splits, current sequence-based deep models plateau at roughly 0.70 ROC AUC (65% accuracy). A 2026 *Nature Machine Intelligence* audit by Szymborski and Emad (20) closed the loop by showing that every tested method, when scored on a cross-species human-SARS-CoV-2 setup, performed indistinguishably from random chance.

These findings carry an unforgiving implication for the original translational ambition of the field. A model that scores well on intraspecies splits with weak homology controls but at chance on cross-species novel-pathogen splits cannot help a clinician reasoning about a newly described accessory protein from an emerging outbreak. It will return ranked partner lists, but acting on them is no better than guessing. The benchmarks the community has been climbing have therefore been measuring something narrower than what was advertised, and the most important translational use case, predicting interactions for pathogens with little or no available training data, has remained essentially unaddressed.

Atlas is a sequence-only PPI prediction model designed against this gap. Four training variants appear in this report. The production checkpoint, referred to throughout as Atlas, is trained on a multi-species PPI corpus assembled from STRING (21) and BIOGRID (22) with cluster-level homology controls and no pathogen-specific holdout; this is the model used for the per-pair benchmarks, the intra-actome comparison against ProteomeLM (23), and the Andes hantavirus and Bundibugyo ebolavirus drug screens. Atlas-Bernett is trained against the Bernett gold-standard split (18) and used only for that head-to-head benchmark. Atlas-Human is trained on a human-only PPI corpus matched to the supervision profile of the comparator panel in the Szymborski and Emad audit (20), and supplies the methodologically matched cross-species comparison on the human-SARS-CoV-2 surface. Atlas-Sars is trained on the same multi-species corpus as the production model, but with all SARS-CoV-2 sequences held out under a strict CD-HIT 90% homology cut, so that no SARS-CoV-2 protein in the test set shares more than 90% sequence identity with any training protein; this is the model used for the COVID-19 anosmia case study. The intuition with the 90% cutoff is to simulate an emerging pathogen that has known homologs in nature, which is almost always true. To the best of our knowledge, Atlas is the first sequence-only method to clear the random-chance floor on a strict homology-controlled human-SARS-CoV-2 interspecies PPI benchmark.

This report is organized in five parts that follow this introduction. The first part establishes that Atlas is competitive on the per-pair benchmarks the field has converged on: Atlas-Bernett reaches the highest reported MCC on the Bernett gold-standard split (18) while running substantially faster than published comparators, and the production multi-species checkpoint exceeds ProteomeLM's published intra-actome ROC AUC on *H. sapiens* by 0.123 absolute on the strict zero-shot subset and matches its mean across 19 bacterial pathogens

[1]Throughout this report we use *intra-actome* to refer to all-vs-all PPI prediction within a single proteome (the upper-triangular pair set of a proteome scored against itself), and *inter-actome* to refer to PPI prediction between two distinct proteomes (the full rectangular block, e.g. host vs. pathogen).

drawn from the Humphreys host-pathogen panel (4). The second part establishes that this competitiveness is not a homology artifact: on a related homology-controlled human-SARS-CoV-2 interaction surface, the human-only Atlas-Human variant reaches ROC AUC of 0.645 and MCC of 0.210 across balanced subsamples, while the Szymborski and Emad comparator panel sits at chance on Gordon AP-MS labels, and the production-style multi-species checkpoint improves on this further. The third part (Case Study I) applies Atlas-Sars to the COVID-19 anosmia and ageusia syndrome by screening the 17-protein SARS-CoV-2 reviewed proteome against the human reviewed proteome, focusing on chemosensory and mitochondrial pathways relevant to the Zazhytska et al. (24) non-cell-autonomous mechanism. The fourth part (Case Study II) applies the production checkpoint to Andes hantavirus, screening the three reviewed ANDV proteins against the FDA-approved drug database in seconds, with the top candidates structurally investigated with Protenix v2 (25) and scored by ipTM and ipSAE. The fifth part (Case Study III) applies the same production checkpoint to Bundibugyo ebolavirus, the agent of the 2026 outbreak in the Democratic Republic of the Congo and Uganda (26), chosen because it is the one pathogen in this report with approved therapeutics of known viral target: four monoclonal antibodies that bind glycoprotein and nothing else, against which a blind screen is scored for recall rather than read for plausibility.

Every finding in this report is computational. Atlas is most useful for hypothesis generation: the case studies are presented as testable hypotheses, with explicit caveats about model-side limitations, including family co-saturation, partner-count inflation, and the meaning of model scores at decision-relevant thresholds. The value of the predictions is in the candidate space they narrow, not in the binary calls they make, and any clinical translation will require independent experimental validation.

Results

Atlas on multi-species PPI

Table 1. Per-pair validation and test metrics for Atlas on the multi-species C3 split. Top-K positive accuracy reflects the production checkpoint's high-confidence shelf.

Split	ROC AUC	PR AUC	MCC	F1	Top-100 pos	Top-1000 pos
Validation	0.918	0.929	0.692	0.844	1.00	1.00
Test	0.918	0.928	0.688	0.843	1.00	1.00

Table 1 reports per-pair validation and test metrics on the C3 multi-species split that defines the production training corpus. Top-K positive accuracy reaches 1.0 at $K = 10$, 100, and 1,000, so the highest-confidence predictions are highly enriched for positives on this distribution. We use this checkpoint for every downstream proteome-scale screen.

Atlas-Bernett on the Bernett gold-standard benchmark

Atlas-Bernett reaches ROC AUC of 0.73, PR AUC of 0.73, F1 of 0.67, and MCC of 0.34 on the Bernett test set (Figure 1). The MCC is the highest reported on this split, with the next-best published per-pair method at MCC of 0.30 (14). The matrix-mode forward produces a full $b \times b$ score grid in a single pass, so throughput after preembedding is several orders of magnitude above per-pair attention methods (TUnA reports 7.30 minutes inference time on the Bernett test set (14), equivalent to roughly 118 pair-scores per second after preembedding). The C3 protocol used by Bernett is the relevant control, since it removes the homology shortcut that other PPI benchmarks tolerate. ProteomeLM-PPI, the supervised variant of ProteomeLM (23), also reports approximately matching ROC AUC on the same split, but its inputs (per-protein ProteomeLM embeddings together with the 48 inter-protein attention coefficients read from a proteome-context forward pass) differ in kind from sequence-pair scoring; we therefore compare with ProteomeLM at proteome scale below rather than per-pair.

Atlas intra-actome screening compared to ProteomeLM

Figure 2 compares Atlas to ProteomeLM-S on the 1+19 species panel that ProteomeLM published (23), with the bacterial pathogen list sourced from (4). On *H. sapiens* at threshold 990, Atlas reaches ROC AUC of 0.968 on the full evaluation and 0.953 on the strict zero-shot subset (43.5 of 184.8 million pairs whose proteins both fall outside any Atlas training clusters), compared to ProteomeLM at 0.83. On the 19-pathogen panel, the mean ROC AUC at threshold 900 is 0.897 for Atlas against 0.898 for ProteomeLM. Atlas leads on 10 of 19 species (the largest gaps are *C. trachomatis* +0.047, *S. pneumoniae* +0.046), trails by more than 0.03 on three species (*L. pneumophila* −0.048, *H. pylori* −0.048, *M. tuberculosis* −0.035), and is within ±0.02 on the remainder.

The held-out subset analysis indicates that the human-side gap reflects generalization rather than memorization. For 13 of 19 pathogens, no proteins appear in any Atlas training cluster; for three more (*S. pneumoniae*, *Y. pestis*, *M. tuberculosis*), only a handful of training proteins overlap and the held-out drop is below 0.001. For the three species with substantial overlap, the held-out drops are 0.024 (*P. aeruginosa*, 44% of proteins in training clusters), 0.032 (*E. coli*, 50%), and 0.007 (*S. enterica*, 0.6%). On *H. sapiens*, where 51.5% of proteins fall in Atlas training clusters, the held-out drop is only 0.015, leaving the lift over ProteomeLM at +0.123 absolute on the strict zero-shot subset.

ProteomeLM and Atlas occupy distinct points on the supervision axis. ProteomeLM-S fits a per-species logistic-regression head on every proteome it benchmarks; Atlas fits one multi-species dual tower and applies it zero-shot. Despite that, the two methods are essentially tied on the bacterial pathogen panel and Atlas leads on *H. sapiens* by 0.123 absolute ROC AUC under the strict zero-shot subset, indicating that a contrastive read-out trained at scale on multi-species PPI transfers across kingdoms without per-species adaptation. Where ProteomeLM leads by more than 0.03 (three pathogens: *L. pneumophila*, *H. pylori*, *M. tuberculosis*), the gap is 0.03 to 0.05 absolute ROC AUC, plausibly reflecting bacterial co-evolution signal that the per-species head exploits and that a frozen-encoder dual

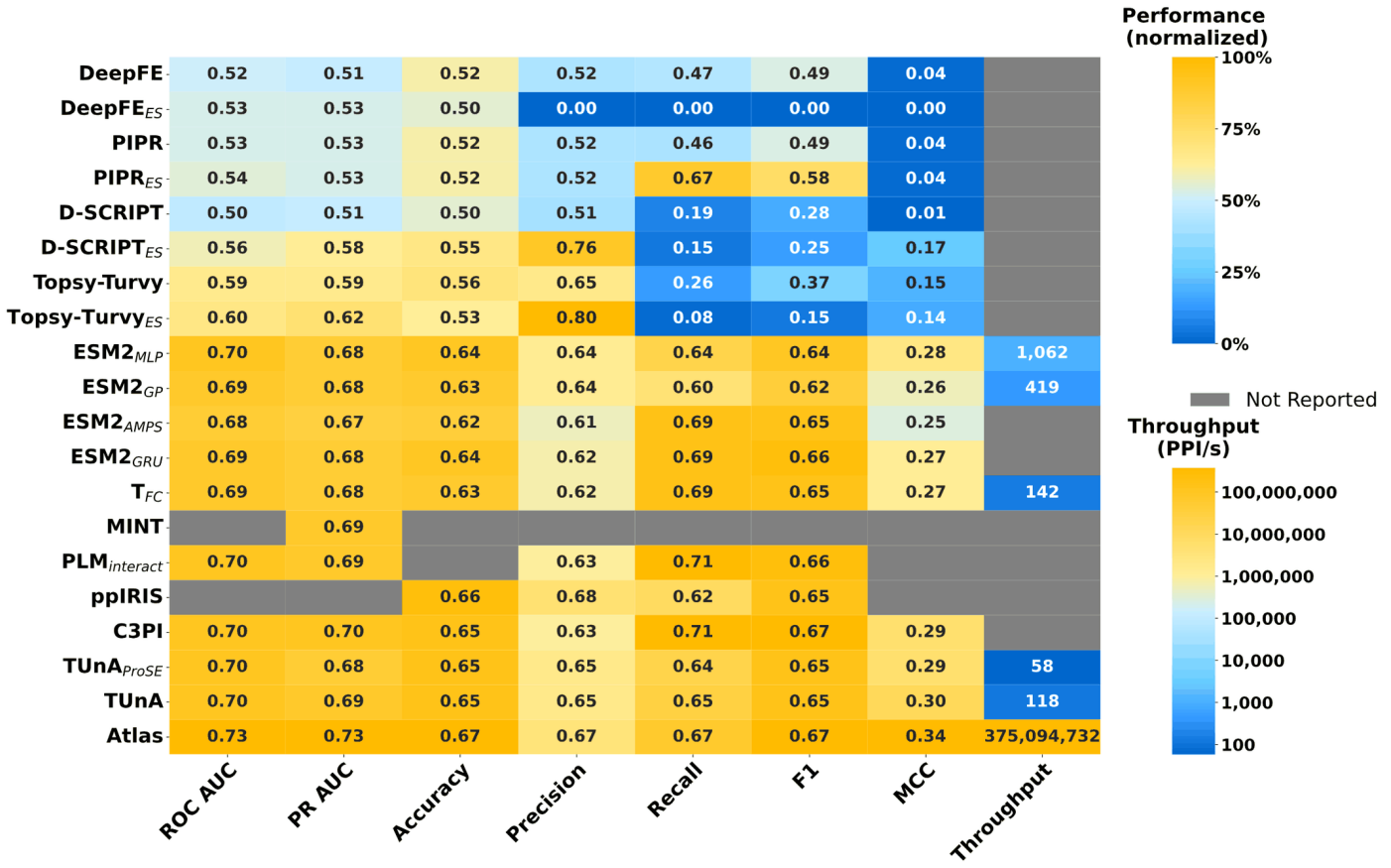


Fig. 1. Atlas-Bernett vs published comparators on the Bernett gold-standard benchmark. Heatmap of test-set ROC AUC, PR AUC, accuracy, precision, recall, F1, MCC (left colorbar, normalized), and throughput in PPI/s (right colorbar, log scale). Cells marked "Not Reported" are absent from the source publications. Bernett split protocol from (18); comparator values are from (18) for DeepFE (6), PIPR (7), D-SCRIPT (8), and Topsy-Turvy (9) (re-evaluated by Bernett et al. on the gold-standard split), and from each source publication for MINT (12), PLM-interact (13), C3PI (15), and TUnA (14).

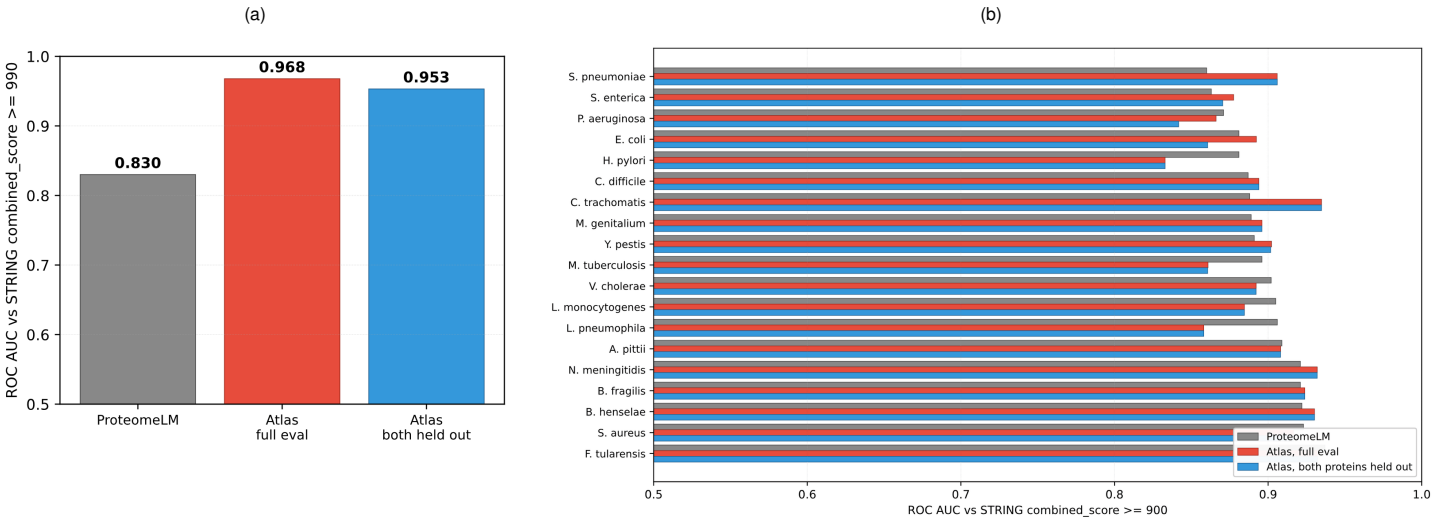


Fig. 2. Atlas vs ProteomeLM ROC AUC on the 1+19 species intra-actome panel. (a) Three-bar comparison on *H. sapiens* at STRING combined-score ≥ 990, matching ProteomeLM's Fig 3B headline view. (b) Per-species ROC AUC at STRING combined-score ≥ 900 for the 19 bacterial pathogens and ≥ 990 for *H. sapiens*. Three bars per species: ProteomeLM published values, Atlas full evaluation, and Atlas held-out (both proteins of every reported pair lie outside any 40%-identity training cluster).

tower trained on a human-heavy PPI mix does not see. Figure 3 shows the underlying score distributions and clustered heatmaps for *E. coli* K12 and *H. sapiens*, which together motivate reading the ROC AUC numbers as reflecting genuine biological structure rather than artifacts of the score distribution.

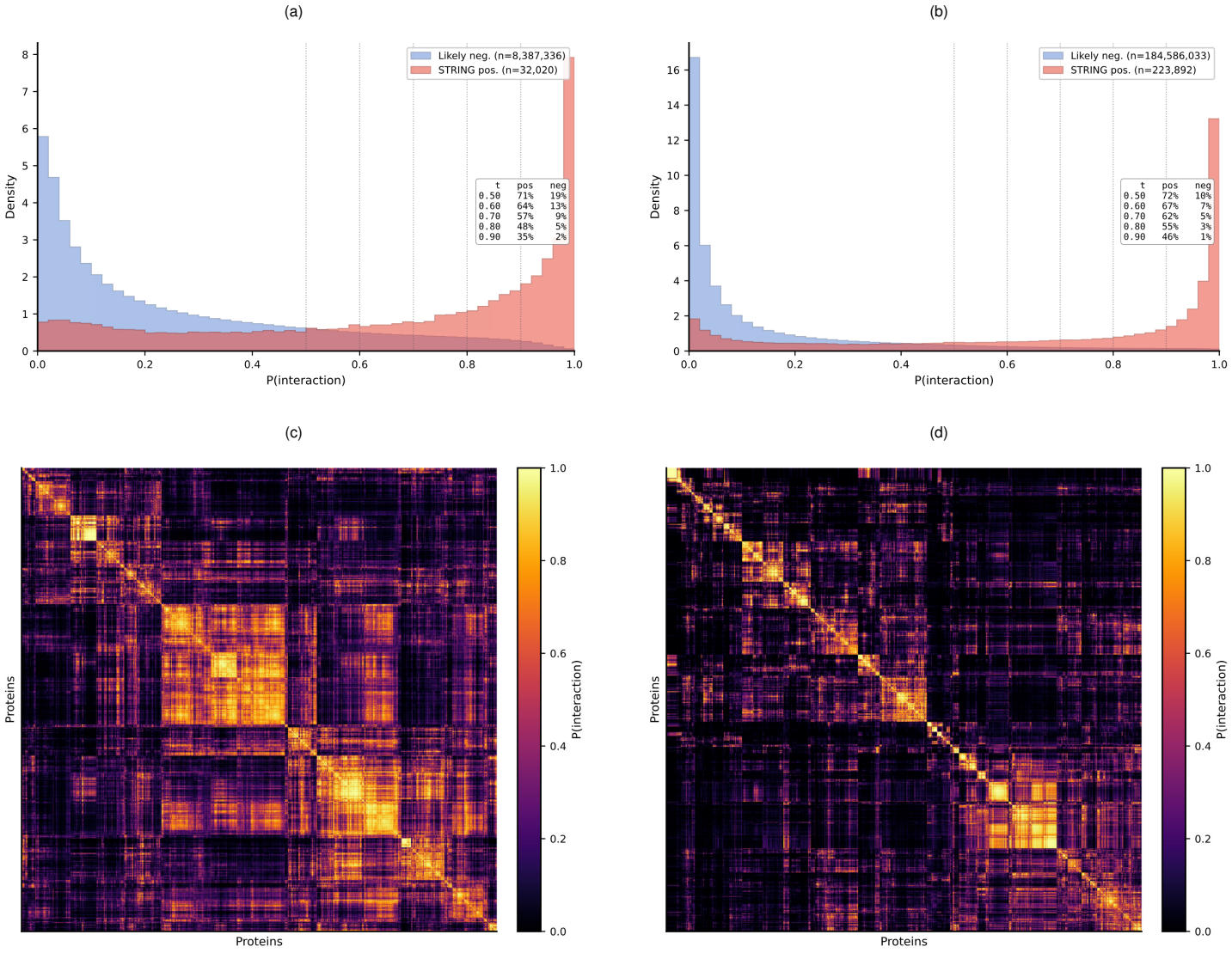


Fig. 3. Intra-actome score distributions and clustered heatmaps under Atlas for *E. coli* K12 (left column) and *H. sapiens* (right column). (a, b) Score distributions over all upper-triangular pairs, showing a bimodal split between the STRING-positive shelf and the likely-negative bulk. (c, d) Hierarchically clustered intra-actome heatmaps; the visible block structure on the diagonal reflects coherent functional modules. The *H. sapiens* run covers 19,226 × 19,226 pairs; the *E. coli* K12 run 4,104 × 4,104.

Atlas produces above-random predictions on a homology-controlled human-SARS-CoV-2 test set

Table 2. Combined human-SARS-CoV-2 per-pair comparison. Atlas rows score the homology-controlled human-SARS-CoV-2 test set used in this report ($6{,}488$ positives, $239{,}975$ negatives), with ROC AUC, Balanced Accuracy, and MCC computed across 10 balanced subsamples. Comparator rows are reproduced verbatim from (20) (averaged across $n = 3$ random training/testing splits at the 50% threshold on Gordon et al. AP-MS labels (1)). The rows therefore compare closely related human-only-to-human-SARS-CoV-2 transfer settings rather than identical pair lists. Atlas rows are shaded.

Method	ROC AUC ↑	Balanced Accuracy ↑	MCC ↑
ProtT5-XL	0.507	0.500	8.34×10^{-5}
ESM2-650M	0.482	0.478	−0.0113
ProtBERT	0.511	0.495	−0.00250
SqueezeProt-SP (Strict)	0.468	0.479	−0.0107
SqueezeProt-SP (Non-strict)	0.515	0.530	0.0157
ProteinBERT	0.444	0.462	−0.0189
ProSE	0.488	0.495	−0.00279
RAPPPID	0.483	0.474	−0.0133
Atlas-Human	0.645	0.598	0.210
Atlas-Sars	0.602	0.541	0.116
Atlas (production)	0.790	0.718	0.435

Table 2 places the Szymborski and Emad (20) per-pair comparison panel next to the three Atlas variants on the related homology-

controlled human-SARS-CoV-2 test set used here. The comparator rows and Atlas rows are not scored on identical pair lists; both settings ask whether human-only PPI supervision transfers to human-SARS-CoV-2 interactions, which is the methodological question this table is meant to summarize. The closest comparison against the Szymborski and Emad baselines is Atlas-Human, which is supervised on the same human-only PPI corpus profile (BIOGRID-MV human together with STRING human ≥ 900, no SARS-CoV-2 proteins in training). Atlas-Human reaches ROC AUC of 0.645 and MCC of 0.210, clearing the random-chance floor where all eight comparator rows sit at chance. The homology-controlled multi-species variant Atlas-Sars remains above random under a CD-HIT 90% virus-side homology cut. The production multi-species checkpoint reaches ROC AUC 0.790 and MCC 0.435 on the same surface, but its training distribution was not subjected to a SARS-CoV-2-specific homology cut and included some human-SARS-CoV-2 interspecies examples, so this jump in performance was expected. The clean methodological claim is therefore the Atlas-Human row: a sequence-only contrastive dual tower trained on a comparable human-only distribution clears random on a related human-SARS-CoV-2 interaction surface, where the previously published human-only methods do not on Gordon AP-MS labels.

Case Study I: SARS-CoV-2 accessory proteins predict interactions with human chemosensory receptors

COVID-19 frequently causes anosmia (loss of smell) and dysgeusia (loss of taste). The dominant proposed mechanism is non-cell-autonomous: supporting cells of the olfactory epithelium express ACE2/TMPRSS2 and are productively infected, while olfactory sensory neurons (OSNs) lack these receptors and are not directly infected. Yet Zazhytska et al. (24) demonstrated that SARS-CoV-2 infection disrupts the "Greek Islands" enhancer hubs that maintain OSN olfactory receptor gene expression, producing a transcriptional collapse without direct OSN infection; the paracrine signal responsible has not been identified. Viral protein persists in nasal mucus for 81-102 days in patients with persistent olfactory dysfunction (27), suggesting sustained local exposure of OSN cilia to secreted viral factors.

We used Atlas-Sars to screen the full SARS-CoV-2 reviewed proteome (17 proteins, UniProt UP000464024) against the complete human reviewed proteome (19,933 proteins, UP000005640) at score > 0.9, with all coronavirus sequences held out at CD-HIT 90% during training so no prediction reflects memorization. Partner lists were run with a local implementation of Enrichr (28) across six pathway libraries; adjusted p-values are within-library Benjamini–Hochberg corrected hypergeometric tests. The screen tests 15 proteins across 6 libraries and hundreds of terms per library; a global Bonferroni threshold at family-wise $\alpha = 0.05$ corresponds to approximately adj$p < 5 \times 10^{-9}$.

As a calibration check, the model cleanly recovers the canonical mechanism for two of the nine accessory proteins it scored: N protein (RNA binding / stress-granule formation; GO:0003723 RNA binding, adj$p = 2.8 \times 10^{-24}$, overlap 26 of 31 partners) and ORF8 (secreted virokine character; 207 of 408 partners in the extracellular region, adj$p = 1.7 \times 10^{-99}$; Reactome defensins, adj$p = 1.3 \times 10^{-51}$). Critically, Spike shows *no* enrichment for olfactory receptor activity (it recovers cell-surface adhesion and integrin interactions instead), ruling out a trivial artifact in which all SARS-CoV-2 proteins look like olfactory-receptor binders to the model.

The six SARS-CoV-2 proteins carrying novel chemosensory predictions are designated Protein A through Protein F throughout this report. Their UniProt accessions, sequence lengths, and topology are withheld pending patent prosecution and ongoing experimental work. Proteins whose screen results only reproduce already-published mechanisms are named normally, since those claims are not novel.

The headline chemosensory leads are Protein A and Protein B (Supplementary Figure S11; per-protein screen details in the Additional Material). Protein A predicts 553 human partners above threshold. The same partner set covers 409 of 433 human olfactory receptors (GO:0004984 olfactory receptor activity) and 37 of 41 mitochondrial Complex I subunits. Two near-saturated overlaps on unrelated protein families co-occurring on one partner set is more parsimoniously explained by a shared model feature than by Protein A carrying two distinct functional binding surfaces, and the 409 of 433 olfactory receptor overlap carries the information content of roughly one family-level hit rather than 409 independent calls, because the human OR family is internally homologous (Class A rhodopsin-like). If Protein A–OR binding is real, it is consistent with the Zazhytska mechanism, since broad perturbation of OR function in OSN cilia would be predicted to collapse the OR-driven PERK/ATF5/Adcy3 feedback loop that maintains OSN identity (29–31).

Protein B predicts 6 of 23 bitter taste receptors (TAS2R4, TAS2R7, TAS2R8, TAS2R9, TAS2R10, TAS2R38; GO:0033038 bitter taste receptor activity) and 3 vomeronasal receptors. This signal does not appear in the prior SARS-CoV-2 taste literature, which has focused on Spike–ACE2 entry into Type II taste cells and one *in silico* Spike–TAS2R46 docking study (32). Protein B's canonical function is RAE1/NUP98 nucleocytoplasmic transport disruption, which the model does not recover, placing the TAS2R prediction in the "novel lead requiring validation" category.

Calibration against published mechanisms

Atlas-Sars was trained without seeing the SARS-CoV-2 proteome at $\geq 90\%$ sequence identity. The CD-HIT 90%/$-n$ 5 holdout allows the model to inherit "small protein may interact with X" priors from vaguely similar PPI training data; "recovery" below means "the model returned the expected partner-family signal without seeing this exact protein," not "blind discovery from no prior."

Net calibration: 2 of 9 (Nucleoprotein, ORF8) are clean blind recoveries of canonical mechanisms. Five of 9 (Spike, ORF3a, Envelope, Membrane, ORF7a) are weak/post-hoc/contested. Two of 9 (Protein B, ORF9b) are outright misses of canonical large-effect partners that the field has multiply replicated. A model that misses canonical large-effect partners on 2 of 9 accessory proteins, while predicting novel partners on the same proteins (Protein B $\rightarrow$ TAS2R, ORF9b $\rightarrow$ TFIID), should not be treated as having earned strong priors on those novel calls from calibration alone. A primary-source audit of the established mechanisms, the extracellular-accessibility argument for Protein A and Protein B, the OR-driven feedback precedent, and the screen-specific caveats are in the Additional Material, under Supplementary detail for the SARS-CoV-2 anosmia screen.

Table 3. Calibration of Atlas-Sars against published mechanisms. Support ratings: clean = canonical mechanism specifically reproduced; weak = generic enrichment anticipated from UniProt FUNCTION; post-hoc = different molecular partner from canonical, salvaged into same broad axis; contested = model recovers a published mechanism subsequently questioned; miss = canonical large-effect partner entirely absent from predictions.

Protein	Published mechanism	What the model recovers	Support
Nucleoprotein (P0DTC9)	RNA binding, stress granules, LLPS	RNA binding, cytoplasmic stress granules, mRNA processing	Clean
ORF8 (P0DTC8)	Secreted virokine, IL-17 mimic	Extracellular region (207/408 partners), defensins, hormones	Clean
Spike (P0DTC2)	ACE2/NRP1, integrin fusogenicity	Cell-surface, integrin-mediated adhesion, ECM-receptor interaction	Weak
ORF3a (P0DTC3)	"Viroporin"/K^+ efflux/NLRP3	Voltage-gated Ca^{2+}, Na^+ regulation, ion homeostasis interactome	Contested
Envelope (P0DTC4)	Viroporin (UniProt FUNCTION)	K^+ channel regulator, cardiac action potential	Weak
Membrane (P0DTC5)	Plasma-membrane/virion assembly	Plasma membrane (193 partners), Na^+ transport, gap junctions	Weak
ORF7a (P0DTC7)	BST2/tetherin antagonism, MHC-I downregulation	TLR/MyD88 pathway	Post-hoc
Protein B	Disrupts RAE1/NUP98 transport; IFN antagonist	TAS2R/chemosensory GPCRs (canonical mechanism missed)	Miss
ORF9b (P0DTD2)	Binds TOMM70, disrupts IFN activation	TFIID/Pol II preinitiation (canonical mechanism missed)	Miss

Case Study II: Atlas-guided drug repurposing identifies FDA-approved candidates for Andes hantavirus

Andes hantavirus (ANDV) is the primary causative agent of hantavirus cardiopulmonary syndrome (HCPS) in the Americas, with case fatality rates up to 40%. HCPS presents as acute respiratory distress: the virus replicates in pulmonary endothelial cells and drives massive vascular leak into the lung. There are no FDA-approved treatments; clinical management is supportive, with ribavirin used off-label for the related hemorrhagic fever with renal syndrome despite limited efficacy for HCPS. A recent hantavirus minigenome study from the CDC (33) supported the pharmacological accessibility of hantavirus replication machinery, reporting *in vitro* activity for remdesivir and baloxavir marboxil in Hantaan and Seoul virus minigenome systems. Renewed case clusters in the United States in 2025 have increased the urgency of identifying candidates for experimental follow-up.

We screened the three reviewed ANDV proteins (N protein, nucleocapsid O36307 428 aa; L protein, RdRp/cap-snatching endonuclease Q9E005 2153 aa; and GP glycoprotein Q9E006 1138 aa) against the full FDA-approved drug database: 2,638 small molecules (protein-ligand interaction, PLI screen) and 626 FDA-approved biologics (protein-protein interaction, PPI screen) using the Atlas production checkpoint. The screen, including loading model weights and preembedding, completed in 33 seconds. Top-ranked predictions were structurally investigated with Protenix v2 (25), scored by ipTM (interface predicted TM-score; $0 =$ low confidence, $1 =$ high confidence) and ipSAE (interaction prediction score from aligned errors; most drug-protein pairs score near zero; values > 0.1 likely indicate high interface confidence, although the metric was designed for protein-protein interfaces). Four known antivirals with documented or *in vitro* hantavirus activity (remdesivir, ribavirin, baloxavir marboxil, and favipiravir) were folded through the same pipeline as calibration controls.

Table 4 reports structural validation scores. Two novel candidates emerged with ipTM and ipSAE exceeding all benchmarked antivirals on their respective targets.

Mupirocin on the N protein. Mupirocin (Bactroban) is an FDA-approved topical antibiotic that inhibits bacterial isoleucyl-tRNA synthetase; it has no prior antiviral use or connection to hantavirus. Atlas ranked it seventh by raw Atlas-PLI score among 2,638 FDA-approved drugs against the ANDV N protein (PLI probability 0.980), behind lutein as the raw rank-1 small molecule, and Protenix v2 prioritized mupirocin structurally with a high-confidence complex at the RNA-binding groove: ipTM $= 0.867$ and ipSAE $= 0.144$, the highest structural validation scores in the entire PLI dataset, above remdesivir (0.853/0.095), baloxavir (0.815/0.077), favipiravir (0.689/0.019), and ribavirin (0.662/0.012). The N protein as a druggable target is supported by Xing et al. (34), who demonstrated direct small-molecule (morin) binding to the hantavirus nucleocapsid RNA-binding groove. Mupirocin represents a structurally distinct, FDA-approved lead against the same target region. The predicted complex is shown in Figure 4a.

Dornase alfa on the L protein. Dornase alfa (Pulmozyme) is recombinant human DNase I, FDA-approved in 1993 for cystic fibrosis and delivered as an inhaled nebulizer to the lung, the primary organ of HCPS pathology. Atlas ranked it below the top raw PPI prediction against the ANDV L protein (PPI probability 0.625, $\approx 2.8\times$ higher than the highest-scoring standard antiviral on the same protein,

Table 4. Structural investigation of Atlas-predicted drug candidates against ANDV proteins via Protenix v2. ipTM: interface predicted TM-score (0–1, overall complex confidence). ipSAE: interaction prediction score from aligned errors (0–1, contact interface confidence; most drug-protein pairs score near zero). For the L-dornase PPI complex, ipSAE is reported as the larger directional ipSAE from the best-ranked Protenix sample. Atlas Prob.: raw sigmoid probability from the Atlas PLI or PPI model. Novel candidates are bold and shaded. Screen: PLI = protein-ligand (small molecule), PPI = protein-protein (biologic).

Drug	Screen	Novel vs. ANDV	Target	Atlas Prob.	ipTM	ipSAE
N protein (O36307, nucleocapsid, 428 aa)						
Mupirocin (Bactroban, FDA 1987)	PLI	Novel	N protein	0.980	0.867	0.144
Remdesivir (Veklury)	PLI	Known *in vitro*	N protein	0.117	0.853	0.095
Baloxavir marboxil (Xofluza)	PLI	Known *in vitro*	N protein	0.527	0.815	0.077
Favipiravir (Avigan)	PLI	Investigated	N protein	0.009	0.689	0.019
Ribavirin (Copegus)	PLI	Clinical (HFRS)	N protein	0.082	0.662	0.012
L protein (Q9E005, RdRp / cap-snatching endonuclease, 2153 aa)						
Dornase alfa (Pulmozyme, FDA 1993)	PPI	Novel	L protein	0.625	0.780	0.472
Remdesivir (Veklury)	PLI	Known *in vitro*	L protein	0.116	0.656	0.009
Baloxavir marboxil (Xofluza)	PLI	Known *in vitro*	L protein	0.191	0.629	0.009
Favipiravir (Avigan)	PLI	Investigated	L protein	0.044	0.499	0.000
Ribavirin (Copegus)	PLI	Clinical (HFRS)	L protein	0.223	0.437	0.000

ribavirin at 0.223), but Protenix v2 predicted a complex at the cap-snatching endonuclease domain with ipTM $= 0.780$ and max directional ipSAE $= 0.472$ in the best-ranked sample; the ipSAE of 0.472 is $> 50\times$ higher than any small molecule tested on the L protein (remdesivir and baloxavir: 0.009; ribavirin and favipiravir: 0.000). The predicted interface lies on the cap-snatching endonuclease domain rather than on the polymerase active site, so an interaction of this predicted strength at that position would be expected to block transcription priming rather than elongation; both the domain assignment and that inhibitory consequence are predictions. Hantavirus minigenome work supports drug sensitivity of the broader polymerase replication machinery (33), but the dornase complex itself remains a computational hypothesis. The known RdRp-targeting drugs score near zero on ipSAE despite their established mechanisms, consistent with these drugs targeting a deep polymerase active-site pocket that is inaccessible in the surface-docked structural prediction geometry. The predicted complex is shown in Figure 4b.

The dornase alfa interface replicates across two independent folding models. Refolding with ESMFold2 (35), architecturally distinct from Protenix v2, reproduces the reviewed ANDV L-dornase interface closely (ipTM $= 0.732$, max directional ipSAE $= 0.403$, against Protenix 0.780 and 0.472). Mupirocin agrees in ordering but not magnitude: ESMFold2 scores it at ipTM $= 0.259$ on the N protein, far below the Protenix 0.867, but still above all four antiviral controls on that target (remdesivir 0.229, baloxavir 0.224, ribavirin 0.183, favipiravir 0.178).

One reported result does not replicate. Under Protenix, the near-ANDV Cruise strain L protein carries a weaker and displaced dornase interface (rank-0 Cruise ipTM $= 0.427$, ipSAE $= 0.0119$, against ANDV 0.780 and 0.472; Figures 5 and 6), which would suggest that the 13 L substitutions in the Cruise sequence change the preferred docking geometry. ESMFold2 does not reproduce this gap, scoring the Cruise complex at ipTM $= 0.713$ and ipSAE $= 0.359$, nearly identical to reviewed ANDV. We therefore treat the L-dornase interface as replicated and the Cruise interface-pose divergence as a Protenix-specific result that should not be carried forward without further evidence. Per-target Atlas rankings, the matched Cruise sequence comparison, and the full folding score matrix are in the Additional Material, under Supplementary detail for the Andes hantavirus screen.

Case Study III: Atlas recovers the approved anti-GP antibody class from the Bundibugyo ebolavirus proteome

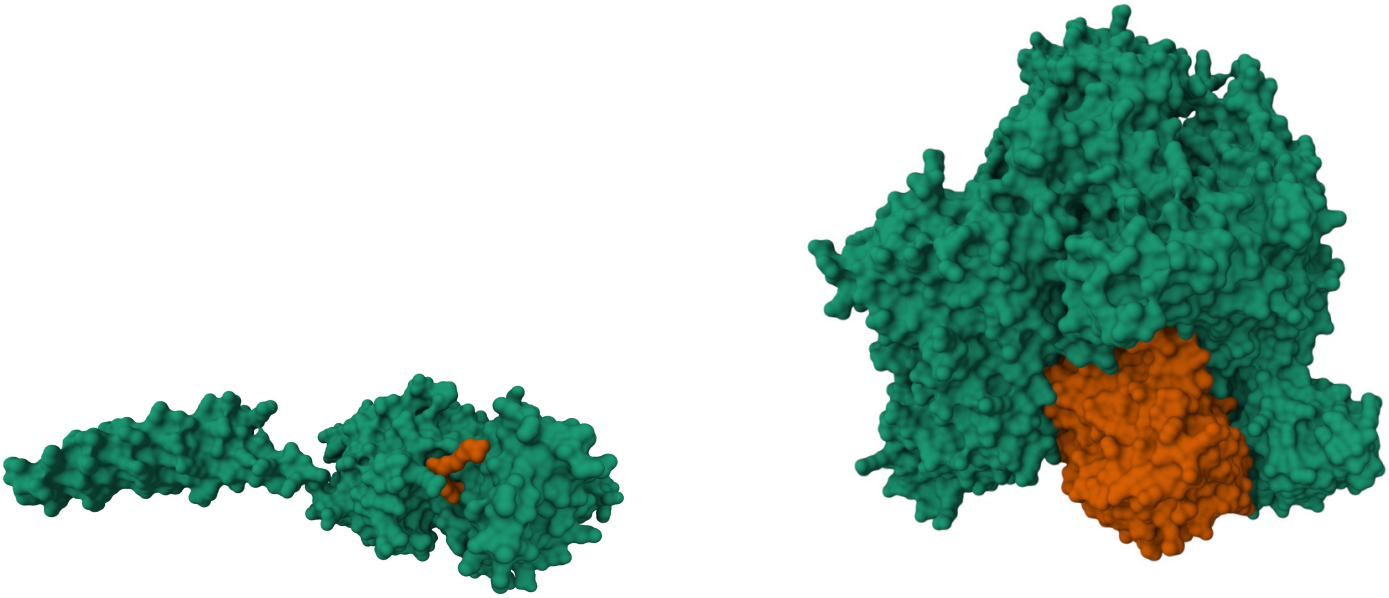
Neither preceding case study can be scored against a known answer. Bundibugyo ebolavirus (BDBV) admits a stricter test: four FDA-approved monoclonal antibodies bind ebolavirus glycoprotein and nothing else in the viral proteome, and all four already sit in the FDA biologics pool that Atlas screens. Scoring the nine-protein BDBV proteome against that pool asks whether the model assigns each antibody to GP out of nine candidates, on a virus species none of them was raised against. The clinical setting is active: BDBV caused a public health emergency of international concern declared on 17 May 2026, with 3,626 confirmed cases and 1,589 deaths and a 44% case fatality ratio in the Democratic Republic of the Congo as of 1 August 2026 (26, 36), and no approved vaccine or treatment (26). Inmazeb (atoltivimab/maftivimab/odesivimab) and Ebanga (ansuvimab) are licensed for *Zaire* ebolavirus only (37); whether a Zaire-directed biologic engages the Bundibugyo glycoprotein is an open question in the current response (38).

We screened all nine BDBV proteins (UniProt UP000143891) against 626 FDA-approved biologics, 2,638 FDA-approved small molecules, and the 19,933-protein reviewed human proteome at threshold 0.9, using the production checkpoints and pools of Case Study II. The screen completed in 60.3 seconds. Top-ranked and control complexes were then folded with ESMFold2 (35).

Antibody-class recovery

Atlas separates the correct antigen completely across all 36 antibody-protein comparisons (Table 5). Every antibody scores between 0.758 and 0.823 on the 676-aa full-length GP and no higher than 0.199 on any of the six non-GP viral proteins, a $3.8\times$ margin with zero overlap. Atoltivimab and maftivimab score highest on full-length GP (0.823 and 0.813, ranks 60 and 88 of 626 biologics) and ansuvimab lowest (0.758, rank 362). The model received no target annotation, no antibody metadata, and no ebolavirus-specific supervision.

None of the four antibodies is rank 1 on GP: the pool leaders are large multi-domain biologics that recur across most viral queries, and



(a) Mupirocin at the RNA-binding groove of the ANDV N protein (highest ipSAE in the PLI dataset). ipTM $= 0.867$, ipSAE $= 0.144$.

(b) Dornase alfa at the cap-snatching domain of the ANDV L protein (ipSAE $> 50\times$ above any small molecule on this target). ipTM $= 0.780$, max directional ipSAE $= 0.472$.

Fig. 4. Protenix v2 structural predictions for the two top-ranked novel ANDV drug candidates, rendered from the deposited `.cif` files with the Mol* (`molstar`) headless renderer used by RCSB. Both panels show molecular surfaces, with the viral receptor in teal-green and the partner highlighted in orange. (a) The ANDV N protein surface is drawn semi-transparent so the bright-orange mupirocin spacefill, lodged in the RNA-binding groove, is visible inside the pocket. (b) ANDV L protein (chain A) and Dornase alfa (chain B) are both rendered as opaque molecular surfaces in contact at the cap-snatching endonuclease domain. ipTM: interface predicted TM-score. ipSAE: interaction prediction score from aligned errors.

Table 5. Atlas-PPI recovery of the four FDA-approved anti-ebolavirus-GP monoclonal antibodies against the nine-protein BDBV proteome. Values are raw Atlas-PPI sigmoid probabilities; for multi-chain biologics the best-scoring chain is reported. The full-length GP column is shaded: the correct target of all four antibodies. Parenthesized values are the rank of that antibody among all 626 FDA-approved biologics scored against that protein. Atoltivimab, maftivimab, and odesivimab are the components of Inmazeb; ansuvimab is Ebanga. All four are licensed for *Zaire* ebolavirus, not Bundibugyo (26, 37).

Antibody	NP 739 aa	VP35 341 aa	VP40 326 aa	GP 676 aa	GP 373 aa	GP 302 aa	VP30 289 aa	VP24 251 aa	L 2,210 aa
Atoltivimab	0.194	0.090	0.092	0.823 (60)	0.184	0.193	0.022	0.046	0.019
Maftivimab	0.199	0.090	0.096	0.813 (88)	0.204	0.206	0.022	0.047	0.019
Odesivimab	0.122	0.088	0.051	0.803 (133)	0.111	0.136	0.018	0.029	0.011
Ansuvimab	0.130	0.107	0.054	0.758 (362)	0.102	0.105	0.016	0.029	0.011

the antibodies land in the top 10 to 15% (Supplementary Table S4). The result demonstrates correct antigen assignment, not discrimination between neutralizing and non-neutralizing binders of the same antigen, and the two shorter GP-gene entries score near the non-GP baseline rather than near full-length GP (Additional Material).

The small-molecule side returns no comparable signal. The four antiviral controls reach best ranks of 175 of 2,638 (brincidofovir, on NP), 397 (ribavirin, on L), 582 (remdesivir, on L), and 1,944 (favipiravir, on L). None has demonstrated efficacy against Bundibugyo virus, and remdesivir underperformed both antibody arms in PALM (37), so unremarkable ranks contradict no established ground truth.

Host inter-actome and canonical partner recovery

The BDBV-human screen returned above-threshold partners for five of nine viral proteins (95 for full-length GP, 31 for VP24, 29 for the 302-aa GP entry, 6 for NP, 3 for VP35). The GP profile is dominated by lipid and sterol transport machinery: ABC-type transporter activity (GO:0140359, adj$p = 2.6 \times 10^{-15}$, overlap 10), leukotriene transport (GO:0071716, adj$p = 3.0 \times 10^{-14}$, overlap 7), and plasma membrane (GO:0005886, adj$p = 4.2 \times 10^{-13}$, overlap 62), with top partners APOL1 (0.961), CD36 (0.958), SLC44A1 (0.949), ABCA3 (0.946), STARD4 (0.942), and SCARB1 (0.940). The canonical filovirus entry receptor NPC1 is itself a cholesterol transporter (39, 40), so the recovered signal is the correct protein family and subcellular compartment.

NPC1 itself does not clear the threshold, scoring 0.849 at rank 361 of 19,933 (98.2nd percentile), and the same holds for TIM-1/HAVCR1 (41) and for the VP35 partner DYNLL1/LC8 (42, 43) (Table 6). On a nine-protein viral proteome, the fixed 0.9 threshold inherited from the large-proteome screens discards canonical partners that a top-2% rank cutoff retains. We report both views rather than retuning the threshold post hoc.

VP24 is an unambiguous miss. Its canonical mechanism, competitive blockade of PY-STAT1 nuclear import through the NPI-1 karyopherins KPNA1, KPNA5, and KPNA6 (44, 45), is supported by a co-crystal structure, yet Atlas places all three in the bottom 60% of

Table 6. Recovery of canonical filovirus host partners in the BDBV-human inter-actome screen. Rank is out of 19,933 reviewed human proteins scored against that viral protein. Partners above the 0.9 operating threshold enter the enrichment analysis; partners below it do not, regardless of rank.

Viral protein	Canonical host partner	Score	Rank	Percentile	Outcome
GP (676 aa)	NPC1 (entry receptor) (39, 40)	0.849	361	98.2	Sub-threshold, top 2%
GP (676 aa)	TIM-1/HAVCR1 (attachment) (41)	0.752	1,019	94.9	Sub-threshold, top 5%
VP35	DYNLL1/LC8 (42, 43)	0.796	385	98.1	Sub-threshold, top 2%
VP24	KPNA1 (44)	0.123	11,216	43.7	Miss
VP24	KPNA5 (45)	0.125	11,091	44.4	Miss
VP24	KPNA6 (45)	0.098	12,939	35.1	Miss

the human proteome. It returns the adjacent compartment instead: 31 partners enriched for the AP-5 adaptor complex (GO:0044599, adj$p = 3.8 \times 10^{-10}$) and vesicle-mediated transport (GO:0016192, adj$p = 2.8 \times 10^{-5}$), including nucleoporins NUP133 and NUP210L. This is the failure mode of ORF7a and Protein B in Case Study I: the right transport axis, the wrong molecular partner class.

Structural cross-check

Folding the same complexes with ESMFold2 (35), an architecturally distinct model from the Protenix v2 pipeline of Case Study II, does not corroborate the antibody result. Ranked by best ipTM across the four antibodies, the L polymerase comes first (0.727) and full-length GP fifth (0.501), inverting the Atlas ordering (Table 7); across all 87 unsplit BDBV protein-protein complexes the two scores are weakly *anti*-correlated (Spearman $\rho = -0.326$, $p = 0.002$). This folding pass is underpowered by construction, since antibodies were folded as single chains rather than paired Fv and over-length antigens were split into windows (Additional Material), so we treat it as a failed corroboration rather than a refutation.

Table 7. Atlas score versus ESMFold2 interface confidence for the four approved anti-GP antibodies across the BDBV proteome. For each viral protein the best value across the four antibodies is shown. Atlas ranks the correct antigen first; ESMFold2 ranks it fifth of nine and places the L polymerase, which none of these antibodies binds, first. Windows gives the number of sequence windows the complex was split into for folding; multi-window systems present a partial antigen. The two highest ESMFold2 scores on certain non-targets (VP24, VP30) are both unsplit. The full-length GP row is shaded.

Viral protein	Best antibody	Atlas prob.	Atlas rank	ESMFold2 ipTM	ipTM rank	Windows
GP (676 aa)	Maftivimab	0.823	1	0.501	5	2
GP (302 aa)	Atoltivimab	0.206	2	0.582	3	1
GP (373 aa)	Atoltivimab	0.204	3	0.308	9	2
NP	Odesivimab	0.199	4	0.429	6	2
VP35	Maftivimab	0.107	5	0.319	8	2
VP40	Ansuvimab	0.096	6	0.322	7	1
VP24	Odesivimab	0.047	7	0.653	2	1
VP30	Maftivimab	0.022	8	0.561	4	1
L	Ansuvimab	0.019	9	0.727	1	4

The disagreement nonetheless bounds folding-based triage. All four antibodies are approved binders of ebolavirus glycoprotein, so their target class is certain even though their affinity for the Bundibugyo variant is the open question; none binds the L polymerase, VP24, or VP30. ESMFold2 nominates exactly those three as its top interfaces, and the two strongest, VP24 at 0.653 and VP30 at 0.561, were folded intact rather than split, so window truncation does not explain them. A structure-first screen on this proteome would have prioritized three certain non-interactions ahead of the correct one. Restricted to full-length GP the folding signal is present but too weak to act on (median ipTM 0.181 versus 0.139 for non-GP proteins, one-sided Mann-Whitney $p = 0.015$), a separation Atlas achieves with no overlap at all.

Net calibration for BDBV: one clean recovery at the level that matters translationally (the anti-GP antibody class, 4 of 4, complete separation from non-target proteins), two sub-threshold near-recoveries of canonical host receptors that a rank-based readout would surface, and one confirmed miss on the most-replicated VP24 mechanism. Screen caveats, the small-molecule artifact analysis, and the folding-run limitations are in the Additional Material.

Proteome-scale perturbation views

The screens above return ranked lists, but the object Atlas actually produces is a full score matrix over a proteome, which can be read as a perturbation map rather than a table. Figure 7 shows two such views, both rendered from precomputed Atlas scores over the SwissBioPics human cell layout (46) and both available as live interactive demos (47). Selecting a query draws its direct predicted partners together with the PPI edges that propagate outward from those partners, so a single ligand or a single pathogen protein resolves into a localized perturbation footprint across compartments.

The protein-ligand view (Figure 7a) scores FDA-relevant ligands against 20,659 localized human proteins. Cholesterol returns 25 direct targets, 17 of them assigned to the cell membrane and 4 to the endoplasmic reticulum, so the predicted footprint tracks the compartments where sterol handling occurs. This is the same lipid and sterol transport signature the model produced independently for the BDBV glycoprotein in Case Study III, reached here from a small molecule rather than from a viral protein. The cross-organism view

(Figure 7b) scores an entire bacterial proteome against the same human cell: Lpp, the Braun lipoprotein and the most abundant protein in *E. coli* by copy number, returns 6 direct human partners, 4 of them secreted, expanding to 48 first-shell and 160 second-shell PPI edges. Neither view is an experimental result, and the same caveats that govern the case-study screens govern them; their purpose is to make the compartment-level structure of a proteome-scale prediction directly inspectable, which a ranked list does not.

Discussion

Atlas is, to our knowledge, the first sequence-only PPI predictor to demonstrate above-random performance on a homology-controlled interspecies benchmark. The Atlas-Human result, trained on human PPI alone and evaluated on a human-SARS-CoV-2 surface with no SARS-CoV-2 proteins in training, establishes that sequence-only PPI supervision can learn a representation that generalizes across species boundaries without any cross-species examples in training. The eight published comparators do not achieve this on Gordon AP-MS labels, and Atlas-Human clears the random-chance floor on the related generated test set used here. The implication is that the underlying signal of host-pathogen binding is recoverable from sequence alone, given the right combination of training-corpus diversity and homology controls; the obstacle has not been the data, it has been how the data was used.

The COVID-19 anosmia case study illustrates how a validated interspecies PPI model can connect large-scale molecular hypothesis generation to physiological phenotypes. The clinical syndrome that drove the screen, persistent anosmia in a substantial fraction of COVID-19 patients, is a population-scale public-health phenomenon whose molecular cause has eluded direct experimental identification for nearly four years since the Zazhytska et al. (24) demonstration that OSN olfactory receptor downregulation is non-cell-autonomous and driven by an as-yet-unidentified paracrine signal. By scoring every SARS-CoV-2 protein against every reviewed human protein in a single proteome-scale screen, Atlas-Sars produced Protein A as a single ranked candidate whose predicted binding profile is consistent with that mechanism, providing a concrete molecular identity for the signal Zazhytska et al. explicitly left as an open question. The co-occurring Complex I signal on the same partner set means that a shared model feature cannot be ruled out without a cross-virus comparator experiment; the prediction motivates wet-lab investment, and the caveats define the controls required for that investment to be interpretable. The screen also surfaced an independent Protein B/TAS2R prediction that mirrors the rare-but-real objective bitter-taste-loss phenotype quantified in three independent psychophysical-testing series, and the same workflow generalizes to the other organ-level COVID phenotypes for which the Atlas-Sars screen produced enriched, plausible partner sets, including cardiac arrhythmia from the ORF3a/Envelope ion-channel signature and intercellular-junction disruption from the Membrane/gap-junction signal. The general pattern is the one that matters: the workflow takes a syndrome with no candidate molecular cause and converts it into a syndrome with a small, prioritized, testable list of candidates, which is the form of input a wet lab can consume.

The hantavirus application demonstrates the same physiology-from-molecules pattern in a translational rather than a mechanistic mode: proteome-scale drug repurposing in a disease with no approved treatment and active outbreak clusters. Mupirocin and dornase alfa are both FDA-approved, have well-characterized human safety profiles, and have target-level support: the N protein RNA-binding groove has direct small-molecule binding precedent from Xing et al. (34), and hantavirus minigenome work supports drug sensitivity of the polymerase replication machinery (33). For dornase alfa, the convergence of a novel computational prediction, a predicted interface on the cap-snatching endonuclease domain at which binding would plausibly inhibit transcription priming, an inhaled delivery route already targeting the primary HCPS organ, and supporting polymerase-targetability evidence makes a plausible translational case for experimental follow-up. The same workflow, scoring the FDA-approved drug database against the reviewed proteome of an emerging pathogen and structurally validating the top candidates, is applicable in days to any pathogen for which sequences are available, including pathogens that have no approved therapeutic and no published interactome data. For an outbreak response, the relevant timescale is not months to years of de novo discovery; it is hours to days to identify already-approved compounds with safety histories, mechanistic plausibility, and ready manufacturing pipelines.

The Bundibugyo screen measures what the first two case studies can only argue. Both generate hypotheses that no available experiment adjudicates, which is why they were chosen and also why they cannot establish whether Atlas is right. BDBV has an approved therapeutic class with a single known viral target, so its screen is scored rather than interpreted: Atlas assigned all four approved anti-GP monoclonals to the glycoprotein out of nine viral proteins with no overlap against any non-target protein, given no drug annotation and no ebolavirus-specific supervision. The same screen bounds the claim. It misses the VP24-karyopherin interaction outright, placing all three NPI-1 karyopherins in the bottom 60% of the human proteome despite a co-crystal structure of that complex, and it ranks NPC1, TIM-1, and DYNLL1 in the top 2 to 5% but below the operating threshold. Atlas is therefore more reliable at assigning a therapeutic to the viral protein it engages than at nominating specific host partners, and on small viral proteomes a rank cutoff recovers canonical biology that the fixed 0.9 threshold discards. A team deciding what to test next can act on both statements; a ranked list with no error model supports neither.

Folding the same complexes clarifies what a sequence-level screen contributes and what it does not. Structure prediction is the natural check on a ranked interaction list, and in Case Study II it worked: two architecturally distinct models independently support the dornase alfa interface on the ANDV L protein, which is the strongest evidence in this report for any novel candidate. That same second model also removes a claim, since the ANDV-versus-Cruise interface divergence appears under Protenix and not under ESMFold2, and we have marked it accordingly. The Bundibugyo antibodies show the reverse failure. Folding ranks the L polymerase, VP24, and VP30 above the true antigen for drugs whose target class is certain from their approval labels, while Atlas ranks those three last, seventh, and eighth. Some of that gap is attributable to a run design that folded single antibody chains and split over-length antigens into windows, and a better-configured pass might close it. The general point survives the caveat: these two families of model fail on different inputs, so agreement between them is informative and neither is a sufficient filter alone. Sequence-level scoring is cheap enough to run first over an entire proteome and an entire drug database, and its value is in choosing the small number of complexes worth the cost of folding, not in

replacing that step.

What makes these case studies practically tractable is throughput. Atlas-Bernett scores the entire Bernett gold-standard test set at approximately 375 million pairs per second after embeddings are precomputed (Figure 1), and that figure is a lower bound set by the size of the Bernett test set itself rather than by the model; in optimized settings on larger problems, sustained throughput exceeds 1 billion pairs per second on a single GPU. At this throughput, the all-vs-all comparison is no longer the constraint. The full FDA-approved drug database, or any list of molecules (ChEMBL bioactivity, Swiss-prot), against any sequenced pathogen proteome can be screened quickly enough for iterative outbreak analysis. The variable regions of every cataloged therapeutic antibody and every IMGT-annotated B-cell-receptor repertoire, scored against every surface-exposed protein of every WHO-priority emerging-pathogen list, runs comfortably overnight on a single workstation, and a saturating epitope-landscape map for one pathogen of interest is achievable in a single afternoon. Whole-proteome host-pathogen interaction maps for arbitrary species pairs, the kind of resource that has historically required years of community effort and dedicated AP-MS campaigns, become a routine output of a single screen. The bottleneck moves from compute to interpretation: deciding which of the millions of above-threshold predictions is worth a wet-lab follow-up. The caveats reported in the COVID-19 case study (family co-saturation, partner-count inflation, score-magnitude calibration) are the operational specification for that interpretation, and as the model improves they will continue to narrow.

The combination of validated above-random cross-species generalization and high-throughput all-vs-all scoring repositions sequence-only PPI prediction from a benchmark exercise into a practical outbreak-response and rare-disease-investigation tool. A clinician confronting an emerging zoonotic pathogen, a researcher trying to explain a previously inexplicable syndrome, or a translational team prioritizing candidates for a neglected disease with no approved therapeutic, can now obtain in minutes a ranked candidate set covering the entire host proteome, the entire FDA-approved drug database, or the entire chemical-bioactivity catalog, on a desk-side GPU. None of these candidate sets is a mechanism, and none replaces experimental validation. All of them collapse what was previously an open-ended search into a small, prioritized, testable list, and that collapse is the deliverable Atlas is built to provide.

Methods

Atlas training variants

Four variants of Atlas are used in this report:

- Atlas (production): trained on BIOGRID-MV ∩ STRING ≥ 900 multi-species corpus (21, 22), C3-split on a 40% global-alignment cluster map (48). Used for the hantavirus and Bundibugyo ebolavirus drug screens and intra-actome comparisons.
- Atlas-Human: trained on the human-only restriction of the same corpus, split on a human-only 40% cluster map. Used as the same-species comparator for the SARS-CoV-2 per-pair benchmark (matched supervision profile to the Szymborski and Emad comparators (20)).
- Atlas-Sars: trained on the same multi-species corpus as production, with every 90% global-alignment cluster containing any SARS-CoV-2 test protein removed from training and validation (48). This taint removes every test virus protein and any protein within 90% identity from training, so no test-side sequence or its close homolog is seen at any point. Used for the proteome-wide COVID anosmia screen (17 SARS-CoV-2 proteins × 19,933 human proteins).
- Atlas-PLI: trained on PLINDER for accurate and high-throughput protein-ligand interaction prediction (49).

The human-SARS-CoV-2 test set contains 6,488 positives and 239,975 negatives (1:37 ratio; a 1:100 target was attempted but 1:37 was the maximum achievable within 1,000 sampling attempts per negative given the 70%-identity violation clustering constraint).

SARS-CoV-2 proteome screen (Case Study I)

Atlas-Sars scored all pairwise interactions between the 17 reviewed SARS-CoV-2 proteins (UniProt proteome UP000464024) and 19,933 reviewed human proteins (UP000005640), retaining pairs above score threshold 0.9. Two polyprotein entries (P0DTC1, P0DTD1) returned zero partners at threshold and are excluded from per-protein analysis. Per-protein partner lists were submitted to Enrichr (28) across GO Biological Process, GO Molecular Function, GO Cellular Component, KEGG, Reactome, and WikiPathways; adjusted p-values are within-library Benjamini–Hochberg corrected hypergeometric tests.

Hantavirus drug repurposing screen (Case Study II)

Three reviewed ANDV proteins (O36307, Q9E005, Q9E006) were scored against 2,638 FDA-approved small molecules (PLI screen) and 626 FDA-approved biologics (PPI screen). PPI screens and host inter-actome enrichment used Atlas-PPI; PLI screens used Atlas-PLI; scores are raw sigmoid probabilities. Human inter-actome enrichment retained pathogen-human pairs above score threshold 0.9 and tested GO Biological Process, GO Molecular Function, GO Cellular Component, KEGG, Reactome, and WikiPathways with within-library Benjamini–Hochberg correction. Top-ranked predictions were structurally investigated with Protenix v2 (25). The reviewed ANDV and Cruise protein folds used MSA-enabled Protenix inference with 10 recycling cycles, 300 diffusion steps, 10 samples from seed 101, and Protenix `ranking_score` to select rank 0 as the best sample for main-text figures and tables. ipTM (interface predicted TM-score), ipSAE (interaction prediction score from aligned errors), pLDDT, and ranking score were extracted from Protenix summary outputs; for the L-dornase PPI complex, the reported ipSAE is the larger of the two directional chain-pair ipSAE values from the selected sample. The four antiviral comparators (remdesivir, ribavirin, baloxavir marboxil, favipiravir) were folded through the same pipeline as calibration controls. For the paired Cruise comparison, reviewed ANDV and Cruise proteins were rerun under identical code, checkpoint versions,

FDA pools, caches, and thresholds; the matched N proteins were required to produce identical PPI, PLI, antiviral, inter-actome, and GO enrichment rows before L and GP differences were interpreted.

Bundibugyo ebolavirus screen (Case Study III)

All nine proteins of UniProt proteome UP000143891 (*Bundibugyo ebolavirus*, taxid 565995: B8XCM7 NP, B8XCM8 VP35, B8XCM9 VP40, B8XCN0 GP, B8XCN1 GP, B8XCN2 GP, B8XCN3 VP30, B8XCN4 VP24, B8XCN5 L) were scored against the same FDA-approved pools used in Case Study II (626 biologics via Atlas-PPI, 2,638 small molecules via Atlas-PLI, withdrawn drugs excluded) and against the 19,933-protein reviewed human proteome (UP000005640) via Atlas-PPI. Reported values are raw sigmoid probabilities; for multi-chain biologics each chain is scored separately and the best-scoring chain is reported per drug. Ranks are computed over the full pool for that query protein without deduplicating chains or salt forms. The complete screen, including weight loading and preembedding, took 60.3 seconds. Human inter-actome enrichment retained pathogen-human pairs above score threshold 0.9 and required at least 5 partners before testing GO Biological Process, GO Molecular Function, GO Cellular Component, KEGG, Reactome, and WikiPathways with within-library Benjamini–Hochberg correction, identical to Case Study II. Percentile ranks in Table 6 are computed over all 19,933 human proteins for the stated viral protein. The four antiviral small-molecule controls were carried over unchanged from Case Study II with brincidofovir added; the four approved anti-GP monoclonal antibodies (atoltivimab, maftivimab, odesivimab, ansuvimab) were pulled from the same FDA biologics pool by name and were not treated differently from any other pool member during scoring.

ESMFold2 structural cross-check (Case Studies II and III)

Top-ranked and control complexes from both the hantavirus and BDBV screens were folded with ESMFold2 (model `esmfold2-2026-05`) through the hosted all-atom endpoint (35), using 20 recycling loops, 100 sampling steps, an MSA depth cap of 1,024, PAE and pair-chain ipTM outputs enabled, and an ipSAE PAE cutoff of 10 Å. Multi-chain biologics were submitted as individual chains rather than as assembled multi-domain complexes, so antibody heavy and light variable domains were folded separately and no paired Fv was constructed. Inputs longer than 768 residues were split into sequence windows and each window folded independently; the reported ipTM and ipSAE for a split system are the best values across its windows, and split systems are labeled as such throughout. Interface metrics were read from the returned per-structure summaries, with ipSAE taken as the larger of the two directional chain-pair values. ipSAE is undefined for protein-ligand complexes in this pipeline and is reported only for protein-protein systems. Across the BDBV case, 1,398 of 1,556 submitted jobs returned successfully; failures were concentrated in ligands whose SMILES the endpoint rejected. Correlations between Atlas probability and ESMFold2 ipTM are Spearman rank correlations computed over unsplit complexes only, to avoid mixing partial and complete antigens.

References

1. David E. Gordon, Gwendolyn M. Jang, Mehdi Bouhaddou, Jiewei Xu, Kirsten Obernier, Kris M. White, Matthew J. O'Meara, Veronica V. Rezelj, Jeffrey Z. Guo, Danielle L. Swaney, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, 583:459–468, 2020. doi: 10.1038/s41586-020-2286-9.
2. Ian R. Humphreys, Jimin Pei, Minkyung Baek, Aditya Krishnakumar, Ivan Anishchenko, Sergey Ovchinnikov, Jing Zhang, Travis J. Ness, Sudeep Banjade, Saket R. Bagde, Viktoriya G. Stancheva, Xiao-Han Li, Kaixian Liu, Zhi Zheng, Daniel J. Barrero, Upasana Roy, Jochen Kuper, Israel S. Fernandez, Barnabas Szakal, Dana Branzei, Josep Rizo, Caroline Kisker, Eric C. Greene, Sue Biggins, Scott Keeney, Elizabeth A. Miller, J. Christopher Fromme, Tamara L. Hendrickson, Qian Cong, and David Baker. Computed structures of core eukaryotic protein complexes. *Science*, 374(6573):eabm4805, 2021. doi: 10.1126/science.abm4805.
3. Jing Zhang, Ian R. Humphreys, Jimin Pei, Jinuk Kim, Chulwon Choi, Rongqing Yuan, Jesse Durham, Siqi Liu, Hee-Jung Choi, Minkyung Baek, David Baker, and Qian Cong. Predicting protein-protein interactions in the human proteome. *Science*, 390(6771):eadt1630, 2025. doi: 10.1126/science.adt1630.
4. Ian R. Humphreys, Jing Zhang, Minkyung Baek, Yaxi Wang, Aditya Krishnakumar, Jimin Pei, Ivan Anishchenko, Catherine A. Tower, Blake A. Jackson, Thulasi Warrier, Deborah T. Hung, S. Brook Peterson, Joseph D. Mougous, Qian Cong, and David Baker. Protein interactions in human pathogens revealed through deep learning. *Nature Microbiology*, 9(10): 2642–2652, 2024. doi: 10.1038/s41564-024-01791-x.
5. Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630:493–500, 2024. doi: 10.1038/s41586-024-07487-w.
6. Yu Yao, Xiuquan Du, Yanyu Diao, and Huaixu Zhu. An integration of deep learning with feature embedding for protein-protein interaction prediction. *PeerJ*, 7:e7126, 2019. doi: 10.7717/peerj.7126.
7. Muhao Chen, Chelsea J.-T. Ju, Guangyu Zhou, Xuelu Chen, Tianran Zhang, Kai-Wei Chang, Carlo Zaniolo, and Wei Wang. Multifaceted protein-protein interaction prediction based on Siamese residual RCNN. *Bioinformatics*, 35(14):i305–i314, 2019. doi: 10.1093/bioinformatics/btz328.
8. Samuel Sledzieski, Rohit Singh, Lenore Cowen, and Bonnie Berger. D-SCRIPT translates genome to phenome with sequence-based, structure-aware, genome-scale predictions of protein-protein interactions. *Cell Systems*, 12(10):969–982, 2021. doi: 10.1016/j.cels.2021.08.010.
9. Rohit Singh, Kapil Devkota, Samuel Sledzieski, Bonnie Berger, and Lenore Cowen. Topsy-Turvy: integrating a global view into sequence-based PPI prediction. *Bioinformatics*, 38 (Supplement 1):i264–i272, 2022. doi: 10.1093/bioinformatics/btac258.
10. Joseph Szymborski and Amin Emad. RAPPPID: towards generalisable protein interaction prediction with AWD-LSTM twin networks. *Bioinformatics*, 38(16):3958–3967, 2022. doi: 10.1093/bioinformatics/btac429.
11. Joseph Szymborski and Amin Emad. INTREPPPID: an orthologue-informed quintuplet network for cross-species protein-protein interaction prediction. *Briefings in Bioinformatics*, 25 (1):bbae405, 2024. doi: 10.1093/bib/bbae405.
12. Vivian Ullanat, Bowen Jing, Samuel Sledzieski, and Bonnie Berger. Learning the language of protein-protein interactions. *Nature Communications*, 17:1199, 2026. doi: 10.1038/s41467-025-67971-3.
13. Dan Liu, Francesca Young, David L. Robertson, and Ke Yuan. PLM-interact: extending protein language models to predict protein-protein interactions. *Nature Communications*, 16: 9012, 2025. doi: 10.1038/s41467-025-64512-w.
14. Young Su Ko, Jonathan Parkinson, Cong Liu, and Wei Wang. TUnA: an uncertainty-aware transformer model for sequence-based protein-protein interaction prediction. *Briefings in Bioinformatics*, 25(5):bbae359, 2024. doi: 10.1093/bib/bbae359.
15. Seyed Morteza Hosseini, G. Brian Golding, and Lucian Ilie. Component puzzle protein-protein interaction prediction. *Briefings in Bioinformatics*, 26(6):bbaf685, 2025. doi: 10.1093/bib/bbaf685.
16. Logan Hallee and Jason P. Gleghorn. Protein-protein interaction prediction is achievable with large language models. *bioRxiv*, 2023. doi: 10.1101/2023.06.07.544109.
17. Yungki Park and Edward M. Marcotte. Flaws in evaluation schemes for pair-input computational predictions. *Nature Methods*, 9(12):1134–1136, 2012. doi: 10.1038/nmeth.2259.
18. Judith Bernett, David B. Blumenthal, and Markus List. Cracking the black box of deep sequence-based protein-protein interaction prediction. *Briefings in Bioinformatics*, 25(2):bbae076, 2024. doi: 10.1093/bib/bbae076.
19. Timo Reim, Anne Hartebrodt, David B. Blumenthal, Judith Bernett, and Markus List. Deep learning models for unbiased sequence-based PPI prediction plateau at an accuracy of 0.65. *bioRxiv*, 2025. doi: 10.1101/2025.01.23.634454.
20. Joseph Szymborski and Amin Emad. A flaw in using pretrained protein language models in protein-protein interaction inference models. *Nature Machine Intelligence*, 8:197–208, 2026. doi: 10.1038/s42256-025-01176-7.
21. Damian Szklarczyk, Rebecca Kirsch, Mikaela Koutrouli, Katerina Nastou, Farrokh Mehryary, Radja Hachilif, Annika L. Gable, Tao Fang, Nadezhda T. Doncheva, Sampo Pyysalo, Peer Bork, Lars J. Jensen, and Christian von Mering. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1):D638–D646, 2023. doi: 10.1093/nar/gkac1000.

22. Rose Oughtred, Jennifer Rust, Christie Chang, Bobby-Joe Breitkreutz, Chris Stark, Andrew Willems, Lorrie Boucher, Genie Leung, Nadine Kolas, Frederick Zhang, Sonam Dolma, Jasmin Coulombe-Huntington, Andrew Chatr-aryamontri, Kara Dolinski, and Mike Tyers. The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Science*, 30(1):187–200, 2021. doi: 10.1002/pro.3978.
23. Cyril Malbranke, Gionata Paolo Zalaffi, and Anne-Florence Bitbol. ProteomeLM: a proteome-scale language model enables accurate and rapid prediction of protein-protein interactions and gene essentiality across taxa. *bioRxiv*, 2025. doi: 10.1101/2025.08.01.668221.
24. Mariana Zazhytska, Albana Kodra, Daisy A. Hoagland, Justin Frere, John F. Fullard, Hani Shayya, Nikki G. McArthur, Rebecca Moeller, Skyler Uhl, Asif D. Omer, Jay A. Gottfried, Dhruv L. Bhatt, Stuart C. Sealfon, Fan C. Pan, Panos Roussos, Benjamin R. tenOever, and Stavros Lomvardas. Non-cell-autonomous disruption of nuclear architecture as a potential cause of COVID-19-induced anosmia. *Cell*, 185(6):1052–1064.e12, 2022. doi: 10.1016/j.cell.2022.01.024.
25. ByteDance Research. Protenix: advancing protein structure prediction with an improved AlphaFold3 implementation. *bioRxiv*, 2025. Protenix v2; available at https://github.com/bytedance/Protenix.
26. World Health Organization. Ebola disease caused by Bundibugyo virus – Democratic Republic of the Congo. https://www.who.int/emergencies/disease-outbreak-news/item/2026-DON614, 2026. Disease Outbreak News DON614, 1 August 2026; accessed 7 August 2026.
27. Gregório Costa dos Santos, Pedro Moreno Pimentel-Coelho, Lia Lima Bueno, et al. Persistence of viral proteins in nasal mucosa of patients with COVID-19 and long-term olfactory dysfunction. *Viruses*, 15, 2023. doi: 10.3390/v15040914. PMC10143909.
28. Edward Y. Chen, Christopher M. Tan, Yan Kou, Qiaonan Duan, Zichen Wang, Gabriela Vaz Meirelles, Neil R. Clark, and Avi Ma'ayan. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, 14(1):128, 2013. doi: 10.1186/1471-2105-14-128.
29. Richard P. Dalton, David B. Lyons, Sujata Bhattacharya, et al. Olfactory receptor gene choice and axonal wiring in mice with a deletion in the olfactory epithelium gene-regulatory region. *Proceedings of the National Academy of Sciences*, 2013. PMC3843319; establishes OR-driven PERK/ATF5/Adcy3 feedback loop maintaining OSN identity.
30. Hani J. Shayya, Dhruv L. Bhatt, Dhruv L. Bhatt, et al. Dynamic remodeling of sensory neuron populations in the mammalian olfactory epithelium. *Nature Neuroscience*, 2022. PMC9588687.
31. Ariel Pourmorady et al. Broad OR perturbation collapses olfactory epithelium transcriptional architecture. *Proceedings of the National Academy of Sciences*, 2024. PMC10765522.
32. Ines V. Hochheiser et al. In silico prediction of SARS-CoV-2 spike protein interaction with human taste receptor TAS2R46. *Computational and Structural Biotechnology Journal*, 2024. doi: 10.1016/j.csbj.2023.12.012. PMID 38189335.
33. Shilpi Jain, Punya Shrivastava-Ranjan, Joel M. Montgomery, Christina F. Spiropoulou, and César G. Albariño. Characterization of minigenome systems for Hantaan and Seoul hantaviruses. *Virus Research*, 363:199678, 2026. doi: 10.1016/j.virusres.2025.199678. Epub 2025 Dec 17; PMID: 41419169; PMCID: PMC12775967.
34. Jian Xing et al. Direct small-molecule binding to the hantavirus nucleocapsid protein. *Biochemical and Biophysical Research Communications*, 2025. Demonstrates direct small-molecule (morin) binding to the hantavirus N protein RNA-binding groove.
35. Chan Zuckerberg Biohub. ESMFold2 all-atom structure prediction, model `esmfold2-2026-05`. https://biohub.ai, 2026. Hosted `/api/v1/fold_all_atom` endpoint; predictions generated 23 June 2026.
36. Centers for Disease Control and Prevention. Notes from the field: outbreak of Ebola disease caused by Bundibugyo virus – Democratic Republic of the Congo and Uganda, May 2026. https://www.cdc.gov/mmwr/volumes/75/wr/mm7522e3.htm, 2026. MMWR Morbidity and Mortality Weekly Report, vol. 75, no. 22; accessed 7 August 2026.
37. Sabue Mulangu, Lori E. Dodd, Richard T. Davey, Olivier Tshiani Mbaya, Michael Proschan, Daniel Mukadi, Mariano Lusakibanza Manzo, Didier Nzolo, Antoine Tshomba Oloma, Augustin Ibanda, Rosine Ali, Sinare Coulibaly, Adam C. Levine, Rebecca Grais, Janet Diaz, H. Clifford Lane, and Jean-Jacques Muyembe-Tamfum. A randomized, controlled trial of Ebola virus disease therapeutics. *New England Journal of Medicine*, 381(24):2293–2303, 2019. doi: 10.1056/NEJMoa1910993.
38. Regeneron Pharmaceuticals. Regeneron's Ebola antibody recommended by World Health Organization for investigational use in response to current Bundibugyo ebolavirus outbreak. https://investor.regeneron.com/news-releases/news-release-details/regenerons-ebola-antibody-recommended-world-health-organization, 2026. Company news release; accessed 7 August 2026.
39. Jan E. Carette, Matthijs Raaben, Anthony C. Wong, Andrew S. Herbert, Gregor Obernosterer, Nirupama Mulherkar, Ana I. Kuehne, Philip J. Kranzusch, April M. Griffin, Gordon Ruthel, Paola Dal Cin, John M. Dye, Sean P. Whelan, Kartik Chandran, and Thijn R. Brummelkamp. Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature*, 477 (7364):340–343, 2011. doi: 10.1038/nature10348.
40. Marceline Côté, John Misasi, Tao Ren, Anna Bruchez, Kyungae Lee, Claire Marie Filone, Lisa Hensley, Qi Li, Daniel Ory, Kartik Chandran, and James Cunningham. Small molecule inhibitors reveal Niemann-Pick C1 is essential for Ebola virus infection. *Nature*, 477(7364):344–348, 2011. doi: 10.1038/nature10380.
41. Andrew S. Kondratowicz, Nicholas J. Lennemann, Patrick L. Sinn, Robert A. Davey, Catherine L. Hunt, Sven Moller-Tank, David K. Meyerholz, Paul Rennert, Robert F. Mullins, Melinda Brindley, Lindsay M. Sandersfeld, Kathrina Quinn, Melodie Weller, Paul B. McCray, John Chiorini, and Wendy Maury. T-cell immunoglobulin and mucin domain 1 (TIM-1) is a receptor for Zaire Ebolavirus and Lake Victoria Marburgvirus. *Proceedings of the National Academy of Sciences*, 108(20):8426–8431, 2011. doi: 10.1073/pnas.1019030108.
42. Toru Kubota, Michihiko Matsuoka, Tsung-Hsien Chang, Mike Bray, Steven Jones, Masato Tashiro, Atsushi Kato, and Keiko Ozato. Ebolavirus VP35 interacts with the cytoplasmic dynein light chain 8. *Journal of Virology*, 83(13):6952–6956, 2009. doi: 10.1128/JVI.00480-09.
43. Priya Luthra, David S. Jordan, Daisy W. Leung, Gaya K. Amarasinghe, and Christopher F. Basler. Ebola virus VP35 interaction with dynein LC8 regulates viral RNA synthesis. *Journal of Virology*, 89(9):5148–5153, 2015. doi: 10.1128/JVI.03652-14.
44. St. Patrick Reid, Louis W. Leung, Amy L. Hartman, Osvaldo Martinez, Megan L. Shaw, Céline Carbonnelle, Viktor E. Volchkov, Stuart T. Nichol, and Christopher F. Basler. Ebola virus VP24 binds karyopherin α1 and blocks STAT1 nuclear accumulation. *Journal of Virology*, 80(11):5156–5167, 2006. doi: 10.1128/JVI.02349-05.
45. Wei Xu, Megan R. Edwards, Dominika M. Borek, Alicia R. Feagins, Anuradha Mittal, Joshua B. Alinger, Kayla N. Berry, Benjamin Yen, Jennifer Hamilton, Tom J. Brett, Rohit V. Pappu, Daisy W. Leung, Christopher F. Basler, and Gaya K. Amarasinghe. Ebola virus VP24 targets a unique NLS binding site on karyopherin alpha 5 to selectively compete with nuclear import of phosphorylated STAT1. *Cell Host & Microbe*, 16(2):187–200, 2014. doi: 10.1016/j.chom.2014.07.008.
46. Philippe Le Mercier, Jerven Bolleman, Edouard de Castro, Elisabeth Gasteiger, Parit Bansal, Andrea H. Auchincloss, Emmanuel Boutet, Lionel Breuza, Cristina Casals-Casas, Anne Estreicher, Marc Feuermann, Damien Lieberherr, Catherine Rivoire, Ivo Pedruzzi, Nicole Redaschi, and Alan Bridge. SwissBioPics: an interactive library of cell images for the visualization of subcellular location data. *Database (Oxford)*, 2022:baac026, 2022. doi: 10.1093/database/baac026.
47. Synthyra PBLLC. Synthyra Atlas interactive demos. https://synthyra.com/demos, 2026. Protein-ligand view: https://synthyra.com/demos?organism=homo-sapiens&view=pli; cross-organism inter-actome view: https://synthyra.com/demos?organism=escherichia-coli-k12&view=interactome; accessed 7 August 2026.
48. Limin Fu, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23):3150–3152, 2012. doi: 10.1093/bioinformatics/bts565.
49. Janani Durairaj, Yusuf Adeshina, Zhonglin Cao, Xuejin Zhang, Vladas Oleinikovas, Thomas Duignan, Zachary McClure, Xavier Robin, Gabriel Studer, Daniel Kovtun, Emanuele Rossi, Guoqing Zhou, Srimukh Veccham, Clemens Isert, Yuxing Peng, Prabindh Sundareson, Mehmet Akdel, Gabriele Corso, Hannes Stärk, Gerardo Tauriello, Zachary Carpenter, Michael Bronstein, Emine Kucukbenli, Torsten Schwede, and Luca Naef. PLINDER: The protein-ligand interactions dataset and evaluation resource, 2024. Pages: 2024.07.17.603955 Section: New Results.
50. Mohana Basu and Rachel Fieldhouse. Deadly hantavirus outbreak on cruise ship puzzles scientists. https://www.nature.com/articles/d41586-026-01512-w, 2026. Nature News, 12 May 2026; accessed 14 May 2026.
51. Virological.org contributors. Complete sequence of Orthohantavirus andesense virus, swiss resident, 2026. https://virological.org/t/complete-sequence-of-orthohantavirus-andesense-virus-swiss-resident-2026/1023/9, 2026. Virological.org discussion thread; accessed 14 May 2026.
52. Virological.org contributors. Preliminary analysis of Orthohantavirus andesense virus sequences from a cruise-ship related cluster, may 2026. https://virological.org/t/preliminary-analysis-of-orthohantavirus-andesense-virus-sequences-from-a-cruise-ship-related-cluster-may-2026/1029, 2026. Virological.org discussion thread; accessed 14 May 2026.
53. World Health Organization. Hantavirus cluster linked to cruise ship travel, multi-country. https://www.who.int/emergencies/disease-outbreak-news/item/2026-DON600, 2026. Disease Outbreak News DON600, 8 May 2026; accessed 14 May 2026.
54. World Health Organization. Hantavirus cluster linked to cruise ship travel, multi-country. https://www.who.int/emergencies/disease-outbreak-news/item/2026-DON601, 2026. Disease Outbreak News DON601, 13 May 2026; accessed 14 May 2026.
55. European Centre for Disease Prevention and Control. Andes hantavirus outbreak in cruise ship, 14 may 2026. https://www.ecdc.europa.eu/en/infectious-disease-topics/hantavirus-infection/surveillance-and-updates/andes-hantavirus-outbreak, 2026. Daily update, 14 May 2026; accessed 14 May 2026.
56. Centers for Disease Control and Prevention. Andes virus outbreak on a cruise ship: Current situation. https://www.cdc.gov/hantavirus/situation-summary/, 2026. 12 May 2026; accessed 14 May 2026.
57. Centers for Disease Control and Prevention. About andes virus. https://www.cdc.gov/hantavirus/about/andesvirus.html, 2026. 9 May 2026; accessed 14 May 2026.
58. European Centre for Disease Prevention and Control. Rapid scientific advice on the management of passengers: In the context of the andes virus outbreak on the cruise ship MV Hondius. https://www.ecdc.europa.eu/en/publications-data/rapid-scientific-advice-management-passengers-context-andes-virus-outbreak-cruise, 2026. 9 May 2026; accessed 14 May 2026.
59. World Health Organization. Follow-up message by the WHO director-general to the people of tenerife regarding the hantavirus response. https://www.who.int/news/item/14-05-2026-follow-up-message-by-the-who-director-general-to-the-people-of-tenerife-regarding-the-hantavirus-response, 2026. 14

May 2026; accessed 14 May 2026.
60. Pathoplexus Team. Andes virus: supporting response to the current hantavirus outbreak. https://pathoplexus.org/news/2026-05-11-andes-virus, 2026. 11 May 2026; accessed 14 May 2026.
61. Pathoplexus Team. Andes virus [hantavirus] browse. https://pathoplexus.org/andv/search, 2026. Accessed 14 May 2026.
62. Duc Huy Nguyen, Lauriane Lecoq, Beat H. Meier, Anja Böckmann, and Marie-Laure Fogeron. SARS-CoV-2 ORF7b interacts with the transmembrane domains of E-cadherin and phospholamban via a leucine zipper motif. *Proceedings of the National Academy of Sciences*, 2024. doi: 10.1073/pnas.2407731121. PMC11573672.
63. Ludovico Cantuti-Castelvetri, Ravi Ojha, Liliana D. Pedro, Minou Djannatian, Jonas Franz, Suvi Kuivanen, Franziska van der Meer, Katri Kallio, Tuure Kaya, Maria Anastasina, et al. Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science*, 370:856–860, 2020. doi: 10.1126/science.abd2985. PMID 33082293.
64. James L. Daly, Boris Simonetti, Karl Klein, Kuan-Ying Chen, Maia K. Williamson, Carlos Anſones, Dhanmeet Bhatt, et al. Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science*, 370:861–865, 2020. doi: 10.1126/science.abd3072.
65. Mengfei Chen et al. Evolution of nasal and olfactory infection characteristics of SARS-CoV-2 variants. *Journal of Clinical Investigation*, 134(8):e174439, 2024. doi: 10.1172/JCI174439. PMCID: PMC11014658.
66. David H. Brann, Tatsuya Tsukahara, Caleb Weinreb, Mariana Lipovsek, Koen Van den Berge, Boying Gong, Rebecca Chance, Iain C. Macaulay, Hsin-jung Chou, Richard B. Fletcher, Diya Das, Kelly Street, Jose Ricardo Roux, Hyeonwoo Kim, John Ngai, Dhanmeet Bhatt, et al. Non-neuronal expression of SARS-CoV-2 entry genes in the olfactory system suggests mechanisms underlying COVID-19-associated anosmia. *Science Advances*, 2020. doi: 10.1126/sciadv.abc5801. PMID 32937591.
67. Andrew N. Miller, Paul R. Houlihan, Rodrigo Matberán, Dhaval Bhatt, Darshan Bhatt, et al. SARS-CoV-1 and SARS-CoV-2 Orf3a are not viroporins. *eLife*, 2023. doi: 10.7554/eLife.84477. PMC9910834.
68. Huanzhou Xu et al. SARS-CoV-2 viroporin encoded by ORF3a triggers the NLRP3 inflammatory pathway. *Virology*, 2022. doi: 10.1016/j.virol.2022.01.003. PMID: 35066302; PMCID: PMC8762580.
69. Michael E. Doyle, Allison Appleton, Qing-Rong Liu, Qin Yao, Caio H. Mazucanti, and Josephine M. Egan. Human taste cells are infected by SARS-CoV-2 and express two-pore domain potassium channel receptors. *American Journal of Pathology*, 191:1785–1800, 2021. doi: 10.1016/j.ajpath.2021.07.010. PMC8179718.
70. Luigi Angelo Vaira, Giovanni Salzano, Giovanna Deiana, and Giacomo De Riu. Objective evaluation of anosmia and ageusia in COVID-19 patients: single-center experience on 72 cases. *Head & Neck*, 42(6):1252–1258, 2020. doi: 10.1002/hed.26204. PMC7267244.
71. Mariano Petrocelli, Francesco Ruggiero, Paolo Boscolo-Rizzo, Roberta Marzolino, Giovanni Fornarini, et al. Objective evaluation of olfactory and gustatory disorders in COVID-19 patients: a prospective study. *Journal of Laryngology and Otology*, 2021. doi: 10.1017/S0022215120002637. PMC8111201.
72. Stéphane-Daniel Le Bon, Nicolas Pisarski, Jerome R. Lechien, Mihai Horoi, Daniel Rodenstein, Corrine Dekeyser, Laetitia Prunier, et al. Prevalence and recovery from olfactory and gustatory dysfunctions in COVID-19 infection. *International Forum of Allergy and Rhinology*, 2021. PMC8222876.
73. Alberto Muñoz-Prieto, Antonio González-Bulnes, Sergio Villanueva-Saz, et al. Salivary proteomics changes in COVID-19 patients with olfactory/gustatory dysfunction. *Scientific Reports*, 2022. doi: 10.1038/s41598-022-14747-0. PMC9237082.
74. Ada Aita, Ilaria Battisti, Nicole Contran, Serena Furlan, Andrea Padoan, Cinzia Franchin, et al. Salivary proteomic analysis in asymptomatic and symptomatic SARS-CoV-2 infection: Innate immunity, taste perception and FABP5 proteins make the difference. *Clinica Chimica Acta*, 2022. PMCID: PMC9549389.
75. Maarit Patrikainen et al. The role of carbonic anhydrase VI in bitter taste perception: evidence from the Car6-/- mouse model. *Journal of Biomedical Science*, 21:82, 2014. doi: 10.1186/s12929-014-0082-6. PMID: 25134447; PMCID: PMC4237775.
76. Ken Matsuoka, Nobuaki Imahashi, Moemi Ohno, Hirotaka Ode, Yoshihito Nakata, et al. SARS-CoV-2 accessory protein ORF8 is secreted extracellularly as a glycoprotein homodimer. *Journal of Biological Chemistry*, 2022. doi: 10.1016/j.jbc.2022.101724. PMC8832879.
77. Xinrui Wu, Tong Xia, Woo-Jin Shin, Kwang-Min Yu, Woohyun Jung, Alexander Herrmann, Shin Shan Foo, Weiqiang Chen, Pengcheng Zhang, Ji-Hwan Lee, et al. SARS-CoV-2 ORF8 is a viral cytokine regulating immune responses and coagulation. *mBio*, 2022. doi: 10.1128/mbio.00273-22. PMC9040823.
78. Xiaoyuan Lin, Beibei Fu, Songna Yin, Zhifeng Li, Huawen Liu, Haiwei Zhang, et al. ORF8 contributes to cytokine storm during SARS-CoV-2 infection by activating IL-17 pathway. *iScience*, 24:102293, 2021. doi: 10.1016/j.isci.2021.102293. PMID: 33723527; PMCID: PMC7942160.
79. Asmaa Hachim et al. ORF8 and ORF3b antibodies are accurate serological markers of early and late SARS-CoV-2 infection. *Nature Immunology*, 21:1293–1301, 2020. doi: 10.1038/s41590-020-0773-7. PMID: 32807944.
80. Scott R. Schaecher, Jason M. Mackenzie, and Andrew Pekosz. The ORF7b protein of severe acute respiratory syndrome coronavirus (SARS-CoV) is expressed in virus-infected cells and incorporated into SARS-CoV particles. *Journal of Virology*, 81(2):718–731, 2007. doi: 10.1128/JVI.01691-06. PMID: 17079322.
81. Goro Nishide, Keesiang Lim, Maiki Tamura, Akiko Kobayashi, Qingci Zhao, Masaharu Hazawa, et al. Nanoscopic elucidation of spontaneous self-assembly of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) open reading frame 6 (ORF6) protein. *Journal of Physical Chemistry Letters*, 2023. PMID: 37707320.
82. Dongwan Kim, Joo-Yeon Lee, Jeong-Sun Yang, Jun Won Kim, V. Narry Kim, and Hyeshik Chang. The architecture of SARS-CoV-2 transcriptome. *Cell*, 2020. doi: 10.1016/j.cell.2020.04.011. PMID 32330414.
83. Yaara Finkel, Orel Mizrahi, Aharon Nachshon, Shira Weingarten-Gabbay, David Morgenstern, Yfat Yahalom-Ronen, Hagit Tamir, Hagit Achdout, Dean Stein, Ofir Israeli, Dana Wolf, Tomer Almog, Avi Eldar, Moshe Schwartz, Ran Nir-Paz, and Noam Stern-Ginossar. The coding capacity of SARS-CoV-2. *Nature*, 589:125–130, 2020. doi: 10.1038/s41586-020-2739-1.
84. Zhenzhen Wang et al. Deletion of type 3 adenylyl cyclase perturbs the postnatal maturation of olfactory sensory neurons and olfactory cilium ultrastructure in mice. *Frontiers in Cellular Neuroscience*, 11:1, 2017. doi: 10.3389/fncel.2017.00001. PMCID: PMC5243839.

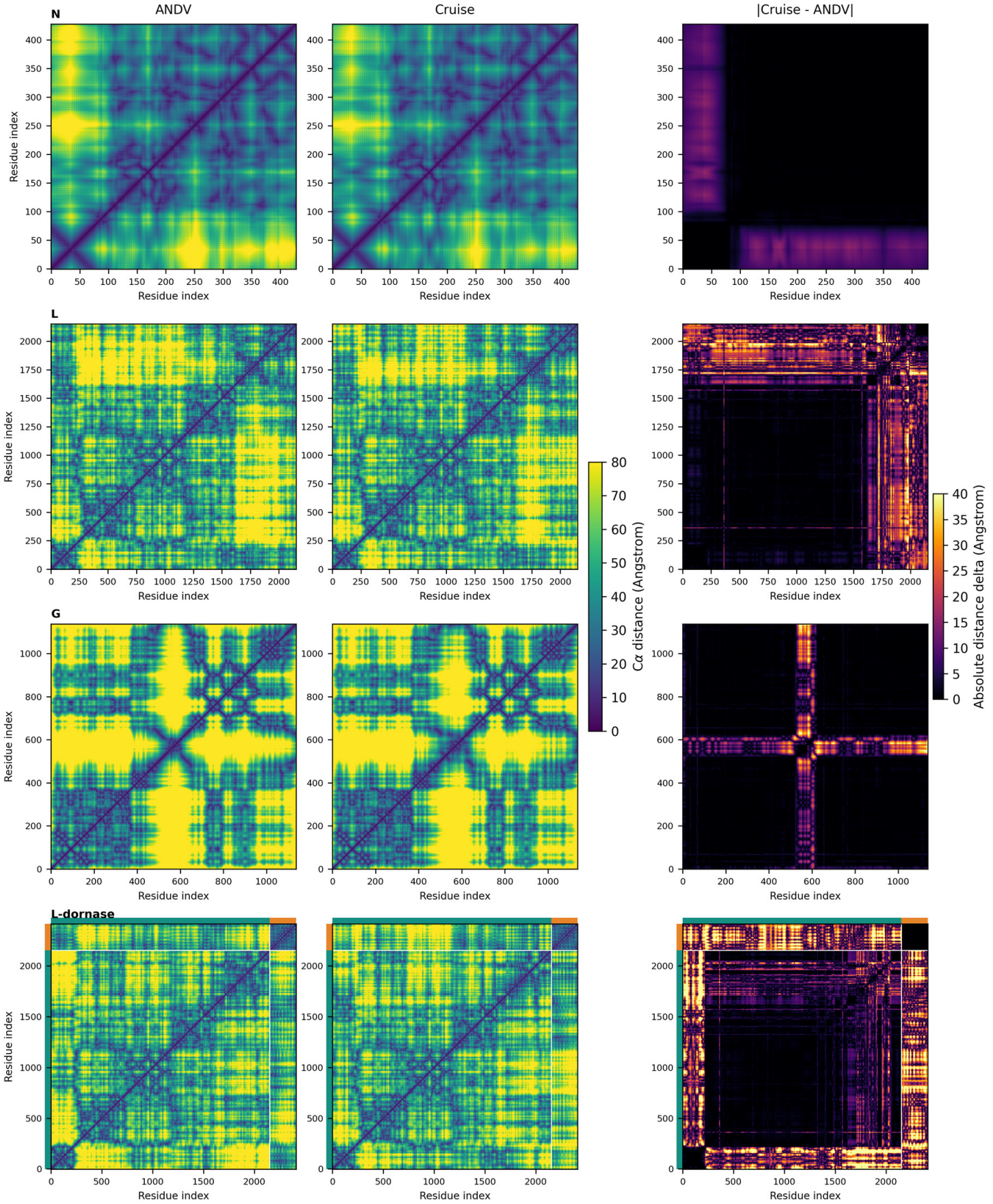


Fig. 5. Rank-0 Protenix Cα distance maps for reviewed ANDV and Cruise strain structures. Rows compare N, L, GP, and L-dornase. Columns show the reviewed ANDV map, the Cruise map, and the absolute distance-map delta in Å. For L-dornase, chain A is L and chain B is dornase alfa; teal and orange axis strips mark the two proteins and the interface quadrants. Distance maps are rigid-body invariant, so differences reflect internal geometry and relative chain placement rather than arbitrary coordinate frames.

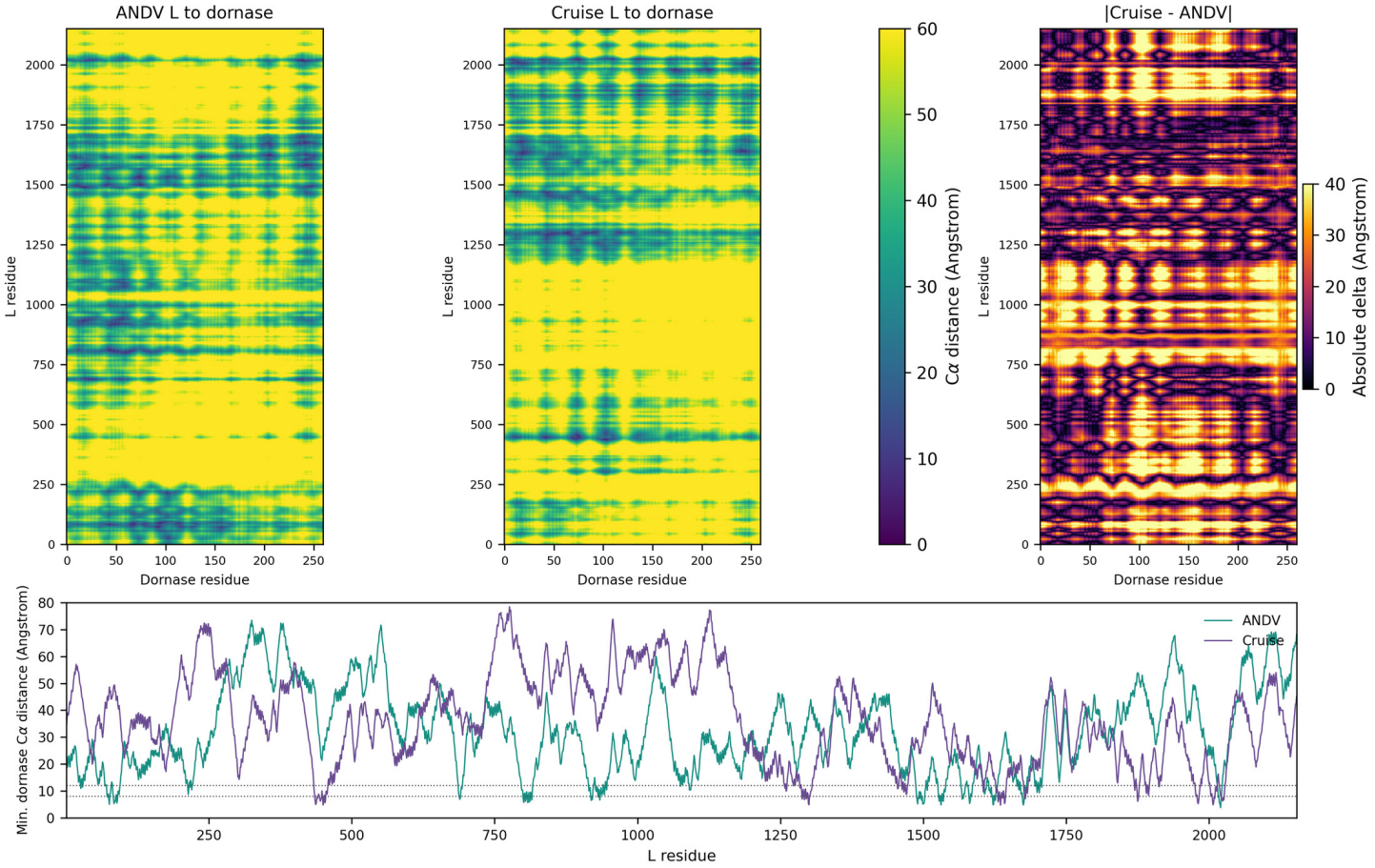
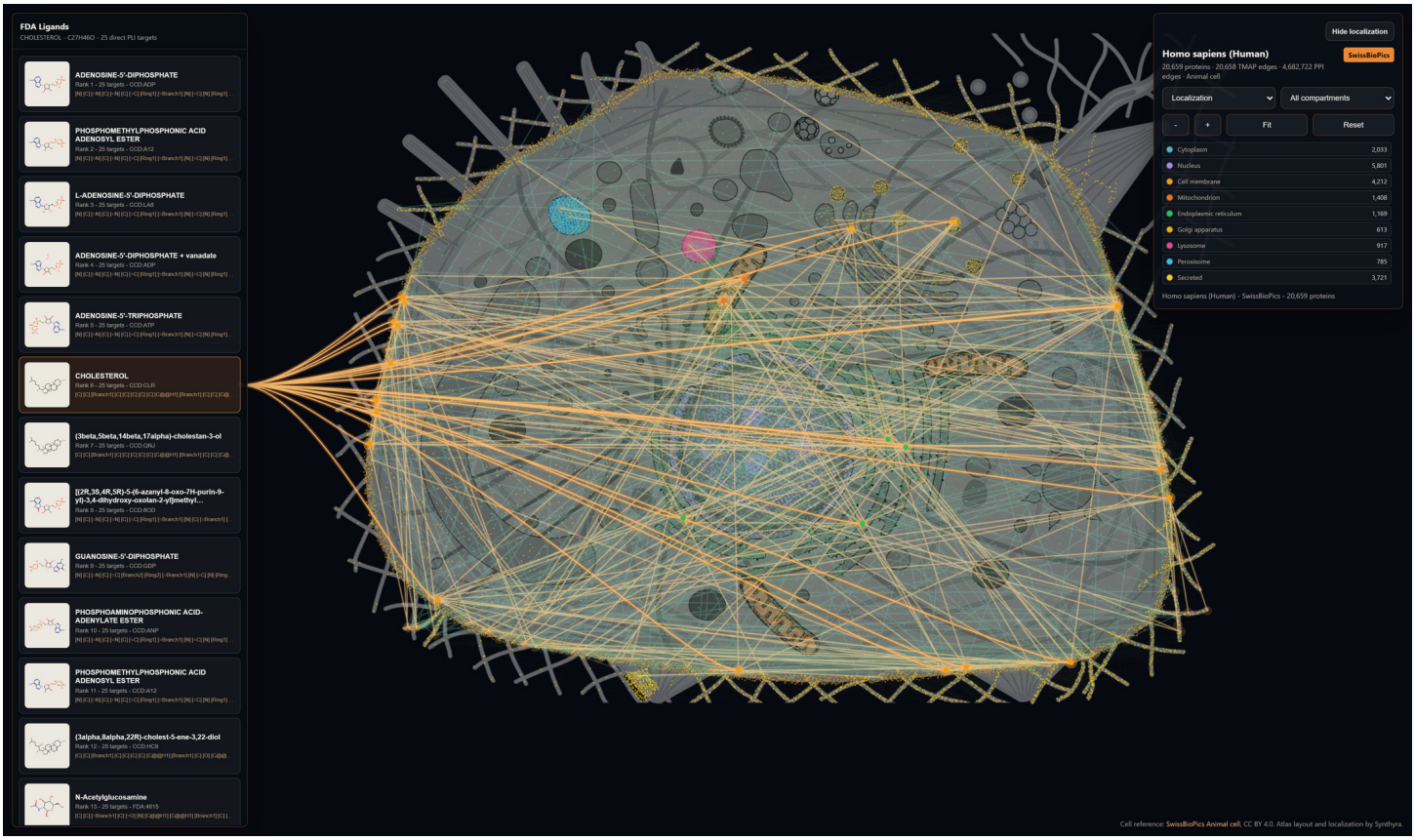


Fig. 6. Cruise L-dornase shifts away from the reviewed ANDV L-dornase interface. The Cruise complex was superimposed onto reviewed ANDV using L chain A only, then L-to-dornase Cα distances were compared. The ANDV rank-0 complex retains a denser close-contact interface and substantially higher structural confidence (ipTM $= 0.780$, max directional ipSAE $= 0.472$) than the Cruise rank-0 complex (ipTM $= 0.427$, max directional ipSAE $= 0.0119$). These values support an interface-pose difference under Protenix rather than direct experimental binding evidence, and the difference is not reproduced by ESMFold2 (main text).

(a) Protein-ligand interaction view, *H. sapiens*



(b) Cross-organism inter-actome view, *E. coli* K-12 → *H. sapiens*

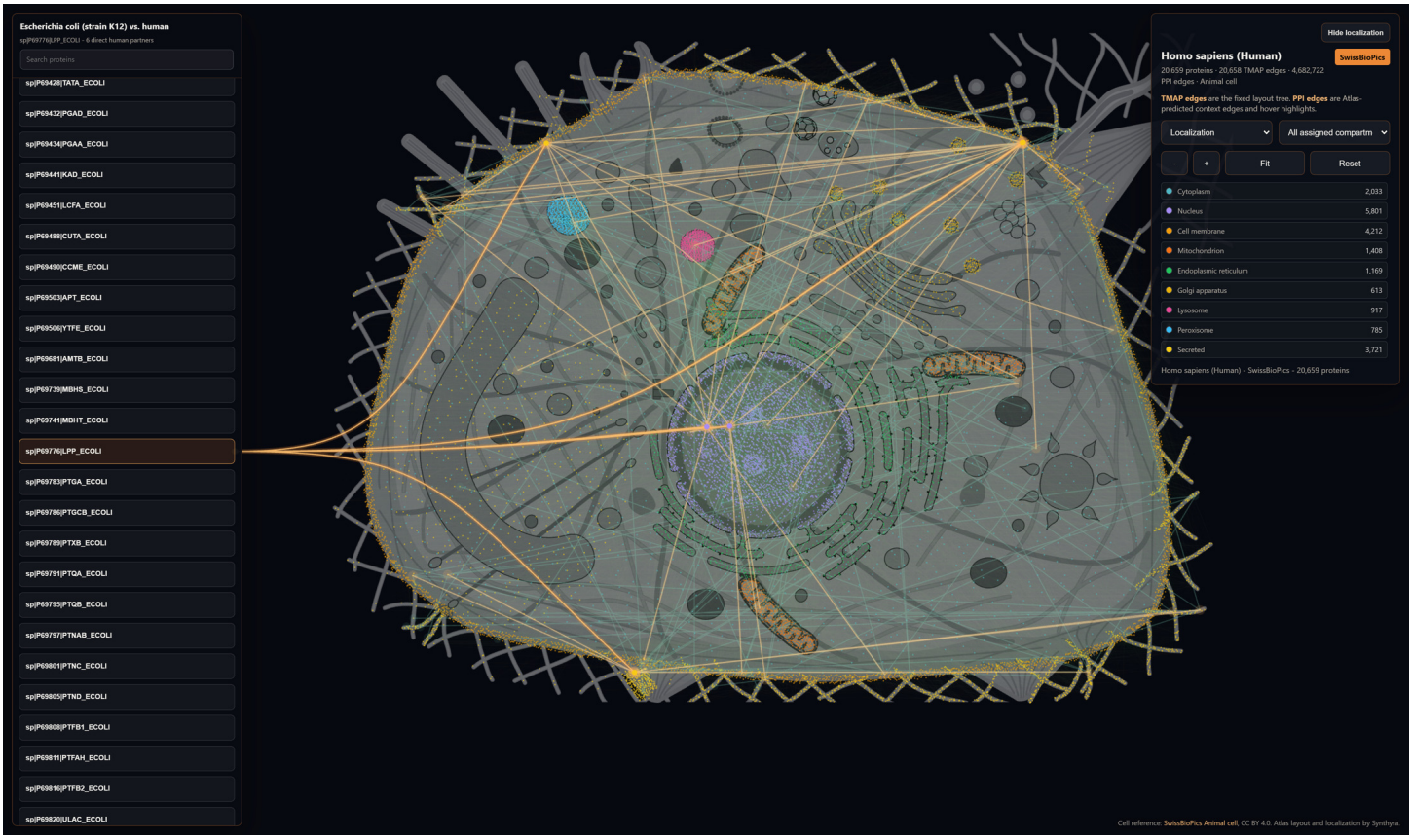


Fig. 7. Atlas perturbation views over a localized human proteome. Both panels render precomputed Atlas scores onto the SwissBioPics human cell layout (46); node color encodes subcellular compartment and the counts at right are localized proteins per compartment. (a) Protein-ligand view. Cholesterol ($C_{27}H_{46}O$) is hovered in the ligand rail, drawing its 25 direct predicted targets in orange together with the PPI edges propagating from those targets into the surrounding network; 17 targets are assigned to the cell membrane, 4 to the endoplasmic reticulum, and 2 each to mitochondrion and Golgi apparatus. (b) Cross-organism inter-actome view. Lpp (P69776), the Braun lipoprotein, is hovered in an unfiltered rail of 4,403 localized bacterial proteins, drawing 6 direct human partners (4 secreted, 2 nuclear) plus 48 first-shell and 160 second-shell propagated PPI edges. Orange edges are the selected query's direct and propagated edges; faint teal edges are intra-human context. Both views are interactive at (47). These are model predictions, not measured interactions.

Additional Material

Supplementary outbreak review: MV Hondius Andes virus cluster and local sequence analysis

This supplement summarizes public information available on 14 May 2026. The core outbreak narrative comes from three rapidly updated source streams: the Nature News report on the quarantine and transmission uncertainty around the cruise-associated cluster (50), the Virological thread reporting the first complete Swiss resident genome (51), and the Virological preliminary analysis of the cruise-associated sequence cluster (52). Quantitative case counts and response measures are anchored to official public-health sources wherever possible because the news and message-board pages were changing during the response window (53–56).

Event chronology and public-health status

The outbreak under review was associated with the expedition cruise vessel MV Hondius, which departed Ushuaia, Argentina, on 1 April 2026. The preliminary sequence-analysis thread reports an early symptomatic passenger on 6 April, an onboard death on 11 April, a stop at Saint Helena on 24 April, and subsequent diagnosis of additional passengers after disembarkation or evacuation (52). WHO's first Disease Outbreak News item on 8 May 2026 reported eight cases linked to the voyage, including six confirmed cases, two probable cases, and three deaths (53). WHO's update on 13 May 2026 revised the event to eleven total reported cases: eight confirmed, two probable, one inconclusive, and three deaths (54). ECDC's 14 May daily update used the same case-category totals, reported no new cases or deaths since the preceding update, and assessed risk to the EU/EEA general population as very low (55). CDC's 12 May situation update stated that no outbreak-linked Andes virus disease cases had been confirmed in the United States at that time, while exposed US passengers were being monitored (56).

The public-health response reflected the unusual status of Andes virus among hantaviruses. CDC describes Andes virus as the only hantavirus known to spread person-to-person, although this appears rare and associated with close or prolonged exposure to symptomatic patients or their body fluids (57). WHO recommended a 42-day quarantine or monitoring window for high-risk contacts because symptom onset can occur weeks after exposure (54). ECDC subsequently published recommendations for self-quarantine and public-health follow-up of exposed passengers and crew (58). Nature's report captured the central uncertainty for public communication: whether the cluster reflected a single zoonotic exposure with limited onward transmission, multiple related spillovers, or more sustained person-to-person spread on the ship (50). On 14 May 2026, WHO also issued a follow-up message to Tenerife, emphasizing the local response effort and continued international coordination (59).

Table S1. Public-source chronology for the 2026 cruise-associated Andes virus cluster. Dates are event or report dates from official public-health communications and the Virological sequence-analysis threads.

Date	Finding	Source
1 Apr 2026	MV Hondius departed Ushuaia, Argentina, beginning the voyage later associated with the cluster.	Virological cluster analysis (52)
6 Apr 2026	The reported index passenger developed symptoms while onboard.	Virological cluster analysis (52)
11 Apr 2026	The reported index passenger died onboard.	Virological cluster analysis (52)
24 Apr 2026	The vessel stopped at Saint Helena, where a subset of passengers disembarked.	Virological cluster analysis (52)
2 to 3 May 2026	Early laboratory confirmations were reported for severe cases evaluated after disembarkation.	Virological cluster analysis (52)
5 May 2026	A Swiss resident was confirmed with Andes virus infection, prompting complete-genome sequencing by Swiss teams.	Swiss Virological genome thread (51)
8 May 2026	WHO reported eight linked cases, six confirmed, two probable, and three deaths.	WHO DON600 (53)
10 May 2026	Virological contributors posted preliminary multi-laboratory genomic analysis of the cruise-associated cluster.	Virological cluster analysis (52)
13 May 2026	WHO updated the event to eleven reported cases, including eight confirmed, two probable, one inconclusive, and three deaths.	WHO DON601 (54)
14 May 2026	ECDC reported the same case-category totals, no new cases or deaths since the previous update, and very low risk to the EU/EEA general population.	ECDC daily update (55)

Virological and Pathoplexus sequence evidence

The Swiss Virological thread reported complete consensus sequences for all three Andes virus genome segments from a Swiss resident confirmed on 5 May 2026. The sequencing was performed by the Geneva University Hospitals Swiss National Reference Center and the University of Zurich. The Geneva workflow used unbiased metagenomic RNA sequencing from plasma, eMAG extraction, Ribo-Zero Gold rRNA depletion, TruSeq total RNA library preparation, and Illumina MiSeq paired-end sequencing. The reported run contained 24,936,538 reads, of which 56.608% were human RNA; Andes virus mapping yielded 67,979 reads for L, 48,785 reads for M, and 17,976 reads for S, with median depths of 766, 954, and 739, respectively (51). The Zurich workflow used the medvir virome protocol, Nextera XT library preparation, and single-end 151 bp sequencing, with 3.1 million quality-filtered reads, 15,690 Andes virus reads, and segment-specific

mean depths of 154X for L, 236X for M, and 122X for S (51). The posted Swiss sequence was named ANDV/Switzerland/Hu-3337/2026, and consensus calling used a minimum coverage threshold of five reads (51).

The preliminary cruise-cluster analysis on Virological was a multi-laboratory effort involving sequence generation in South Africa, Senegal, Switzerland, and the Netherlands, with SISPA, unbiased metagenomics, TWIST capture enrichment, Illumina, Element BioSciences, and Oxford Nanopore data represented in the workflow (52). Its central genomic result was strikingly narrow variation among available cases: S- and M-segment consensus sequences were identical across the sequences analyzed, and only two L-segment single-nucleotide polymorphisms were noted relative to MN258159, C2139T in case 5 and G576A in case 7. Both L-segment substitutions were reported as synonymous (52). The same thread compared the outbreak genomes with historical Argentinian human sequences and reported approximately 98.7% nucleotide identity across L, M, and S, placing the cluster within Andes virus rather than a divergent orthohantavirus lineage (52). Segment-wise phylogenetic discussion in the Swiss thread placed the virus near human and rodent-associated Andes virus lineages from Argentina and Chile, consistent with the Oligoryzomys longicaudatus reservoir system, with no obvious cross-segment reassortment signal in the posted rapid analyses (51).

Pathoplexus added Andes virus support during the outbreak response and credited four submitting teams across the Netherlands, Switzerland, Senegal, and South Africa (60). The launch note listed current outbreak accessions PP_006W3U9, PP_006WBLH, PP_006W6RC, and PP_006WANE (60). The live Pathoplexus browse page, checked on 14 May 2026, returned an Andes virus data set containing outbreak-associated entries including PP_006W3U9.1, PP_006WBLH.1, PP_006W6RC.2, and PP_006WANE.1 (61). The most conservative interpretation is therefore the one adopted by WHO: sequence similarity supports a single zoonotic spillover or a small number of closely related spillovers followed by limited onboard transmission, but genomes alone cannot resolve every exposure link (54).

Connection to the local ANDV-like protein analysis

Our local analysis compared the Cruise protein set against Andes virus N, glycoprotein precursor, and L proteins, then summarized segment-specific and concatenated dendrograms under `case_studies/hantavirus/results`. The local protein-level result is highly concordant with an Andes virus identity assignment: N was 428 of 428 residues identical to reference Andes virus, G/GP was 1128 of 1138 residues identical, L was 2140 of 2153 residues identical, and the concatenated N+G/GP+L comparison was 3696 of 3719 residues identical, or 99.382% identity overall (Table S2). The nearest-neighbor analysis in the same result set placed Andes virus as the closest reference in N, G/GP, L, and the concatenated protein comparison, ahead of Sin Nombre virus and other orthohantavirus references.

Table S2. Cruise strain Andes virus protein comparison. Counts summarize the local Cruise protein analysis. This comparison is protein-level only and does not test nucleotide identity, synonymous substitutions, untranslated regions, segment termini, or reassortment directly.

Protein or concatenation	Aligned residues	Identical residues	Differences	Identity
N nucleocapsid	428	428	0	100.000%
G/GP glycoprotein precursor	1138	1128	10	99.121%
L polymerase	2153	2140	13	99.396%
N+G/GP+L concatenated	3719	3696	23	99.382%

This result should be interpreted as a strong local protein-level assignment to Andes virus or an Andes-virus-like sequence, not as proof that the Cruise proteins are identical to the 2026 cruise-associated outbreak genomes. The Virological outbreak analyses emphasize nucleotide-level consensus identity, L-segment synonymous variation, and segment-wise phylogenetic placement (51, 52). By contrast, our current local comparison collapses away synonymous nucleotide information and does not include direct alignment to the 2026 Pathoplexus nucleotide accessions. Establishing identity to the outbreak cluster would require direct segment-level nucleotide comparison against the posted 2026 L, M, and S genomes, including consensus ambiguities, untranslated regions where available, and segment-specific phylogenies.

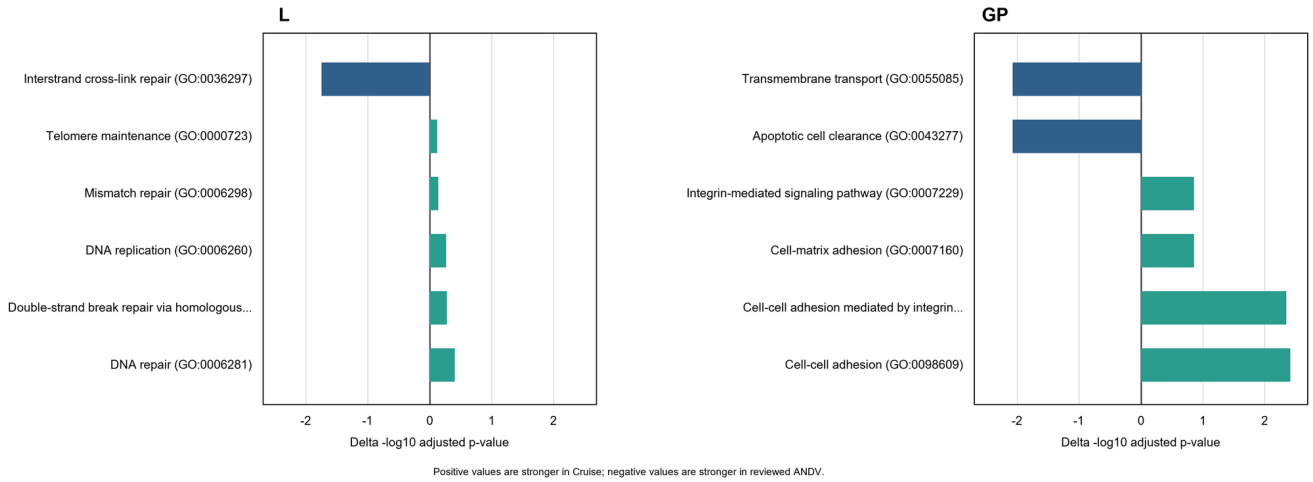


Fig. S1. Matched ANDV versus Cruise GO Biological Process enrichment shifts for L and GP. Bars show the delta in $-\log_{10}$ adjusted p-value for the top-changing terms after rerunning reviewed ANDV and Cruise proteins under the same Atlas-PPI checkpoint and score threshold 0.9. Positive values indicate stronger enrichment in Cruise; negative values indicate stronger enrichment in reviewed ANDV. N is omitted from this plot because the matched N proteins are sequence-identical and produced exactly identical enrichment output.

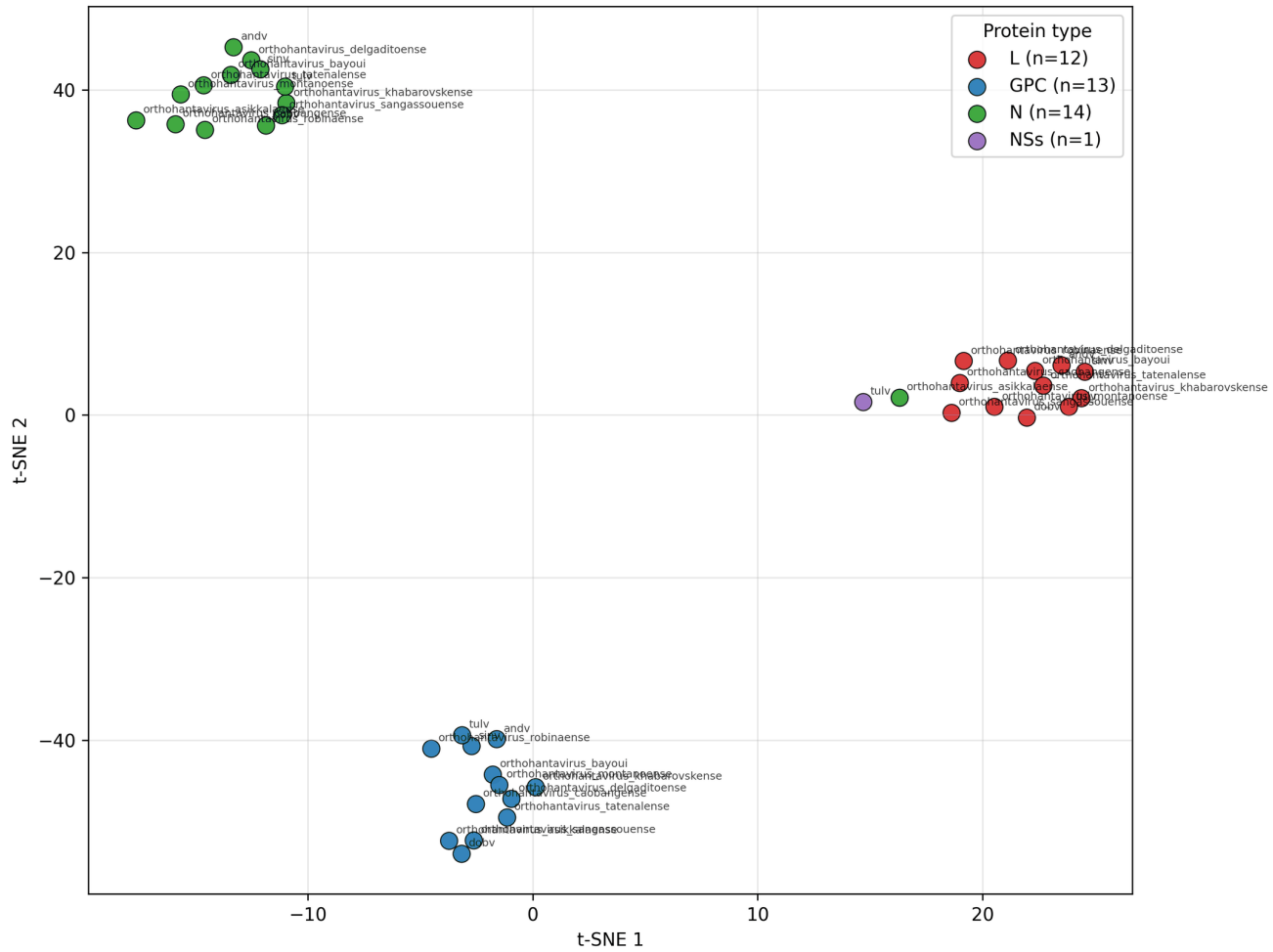


Fig. S2. Protein-level t-SNE of hantavirus proteins in human partner-probability space. Each point represents one viral protein by its Atlas score vector against 19,933 human proteins. Proximity therefore reflects model-predicted host interaction similarity, not sequence phylogenetic distance. N, L, and GPC separate primarily by predicted human partner profile, providing the protein-class context for the matched ANDV versus Cruise comparison. 40 viral proteins, t-SNE perplexity 5.0.

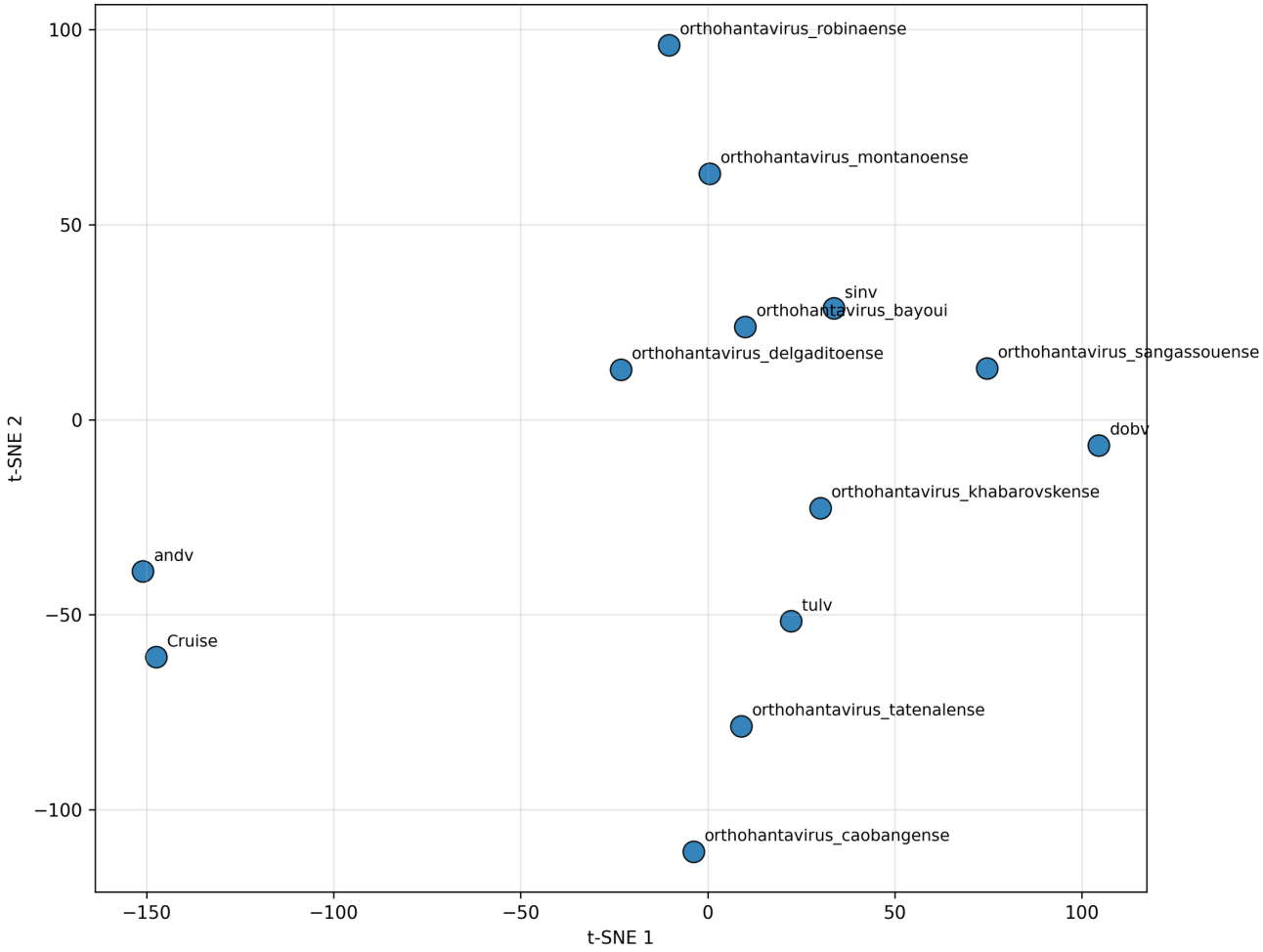


Fig. S3. Organism-level t-SNE of concatenated hantavirus host-interaction profiles. Each point concatenates L, GPC, and N Atlas score vectors against the reviewed human proteome. Proximity reflects similarity in model-predicted host partner probabilities across the three proteins, not phylogenetic distance. The Cruise profile sits close to reviewed ANDV in this representation, consistent with a near-ANDV host-interaction profile despite the small L and GP substitutions. 13 viruses, vector dimension 59,799, t-SNE perplexity 3.0.

Fig. S4. Reviewed ANDV L enrichment profile. Full human inter-actome enrichment summary for Q9E005 under the matched screen. This reviewed ANDV profile is dominated by DNA repair and replication-associated terms and is shown as the direct comparator for the Cruise L plot.

Fig. S5. Cruise L enrichment profile. Full human inter-actome enrichment summary for the Cruise-associated ANDV-like L protein. Relative to reviewed ANDV L, the Cruise profile modestly increases several DNA repair and replication enrichments while reducing interstrand cross-link repair.

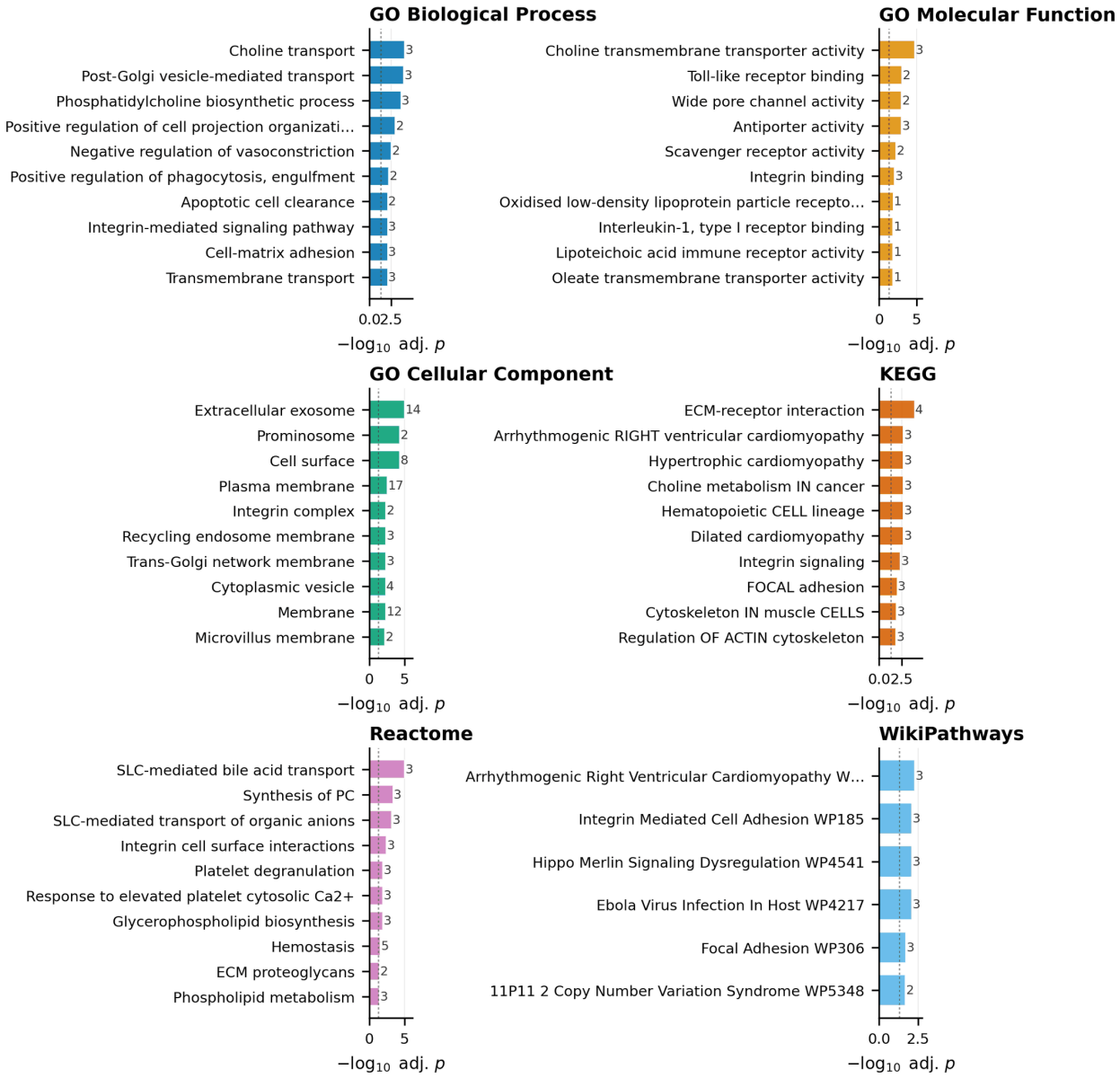


Fig. S6. Reviewed ANDV GP enrichment profile. Full human inter-actome enrichment summary for Q9E006 under the matched screen. This reviewed ANDV profile is the comparator for the Cruise GP plot and is stronger for apoptotic cell clearance and transmembrane transport terms.

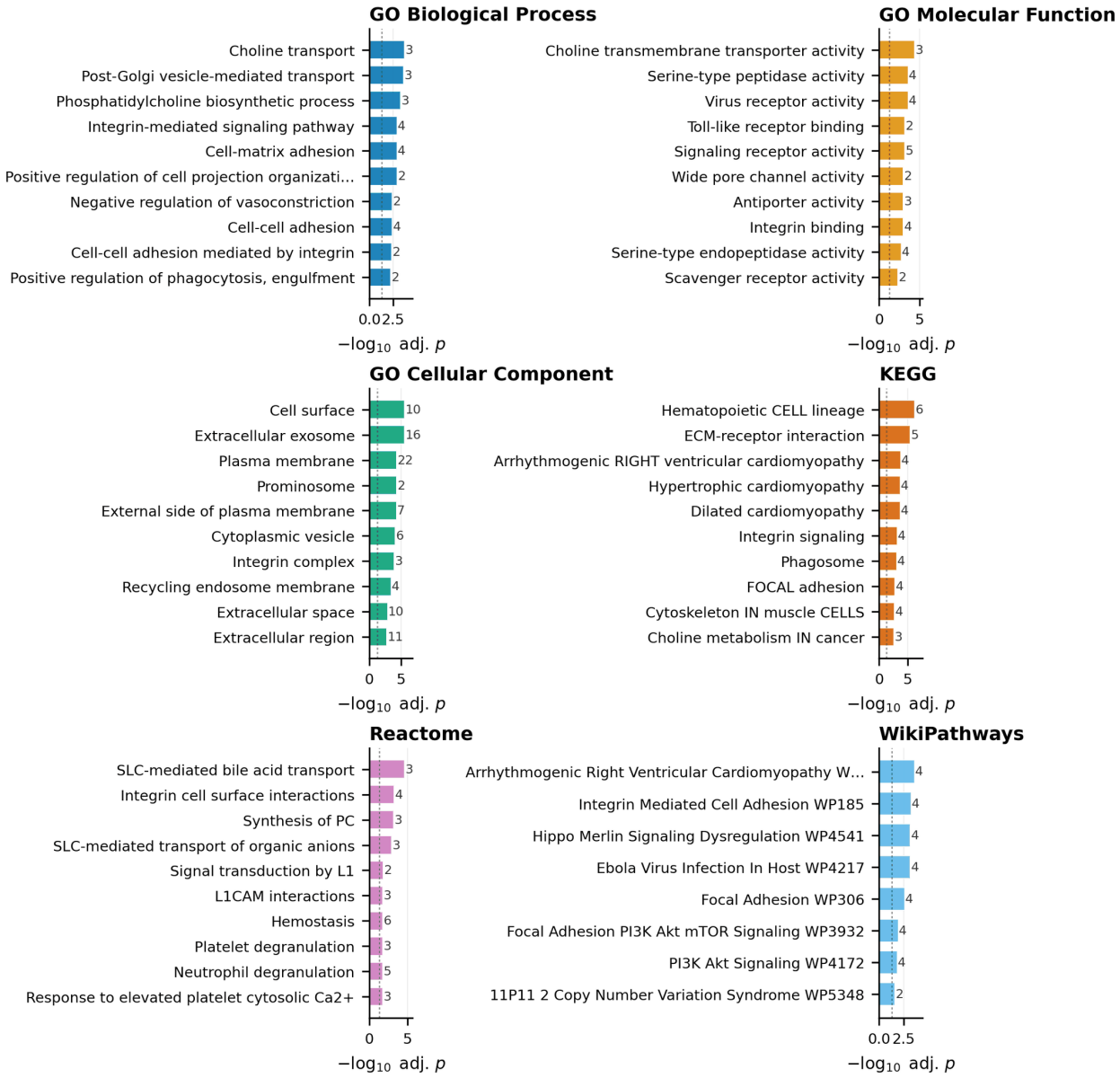


Fig. S7. Cruise GP enrichment profile. Full human inter-actome enrichment summary for the Cruise-associated ANDV-like glycoprotein. Relative to reviewed ANDV GP, this profile shifts toward cell-cell adhesion, integrin-mediated adhesion, and cell-matrix adhesion.

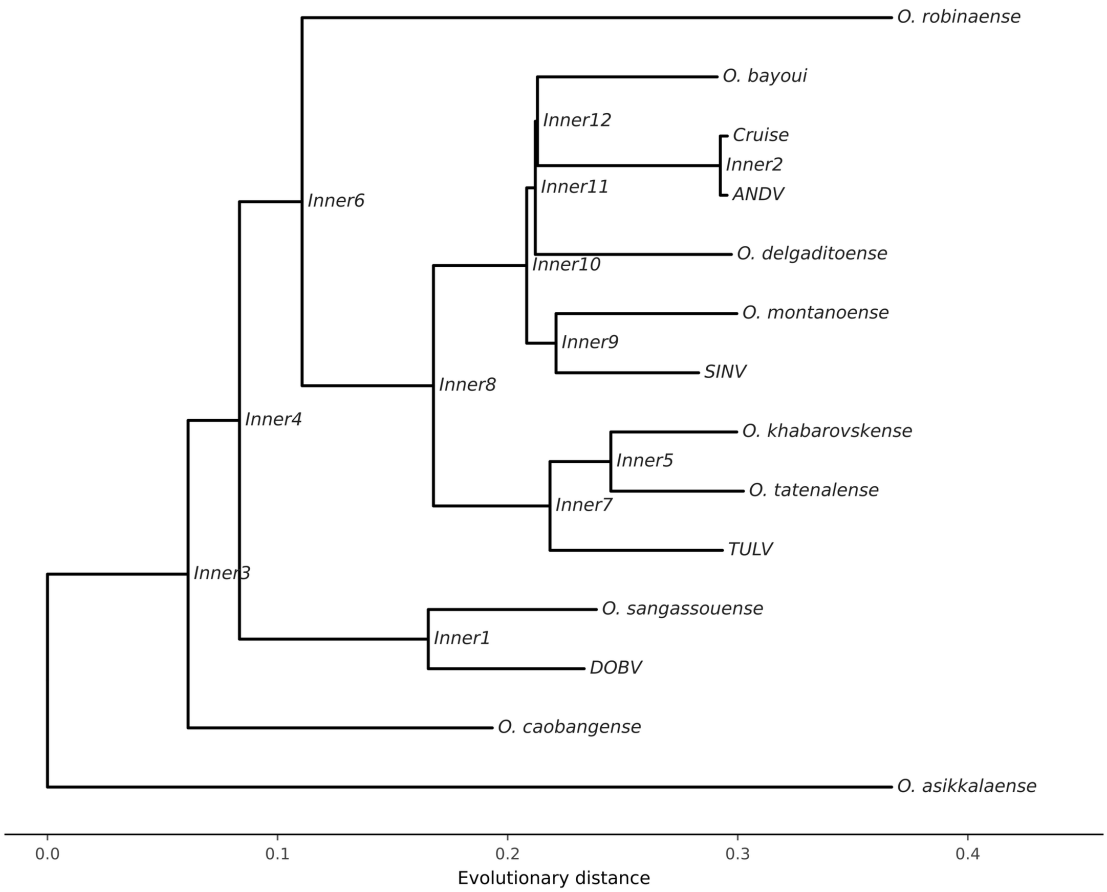
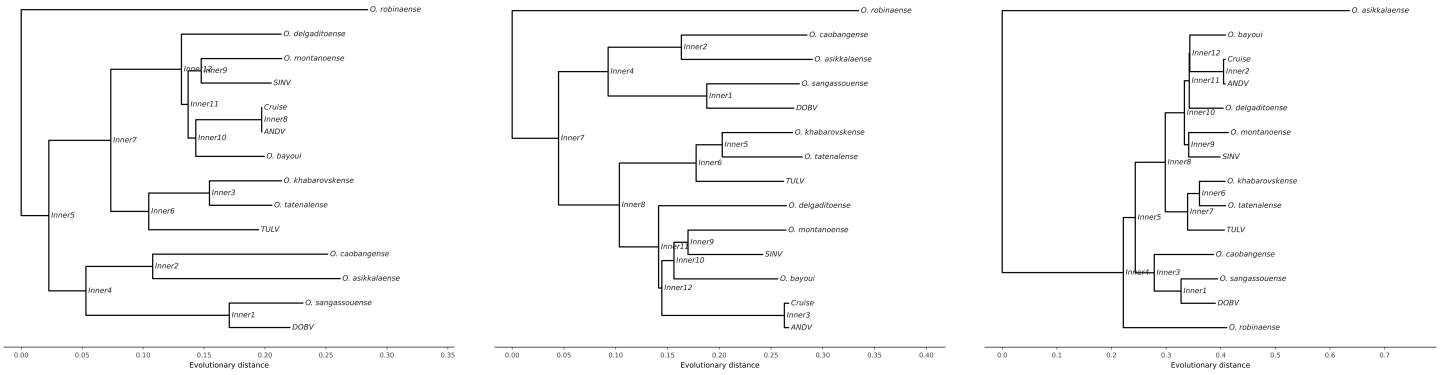


Fig. S8. Consensus dendrogram from the Cruise ANDV-like protein analysis. The concatenated N+G/GP+L protein comparison places the Cruise sequence set nearest to Andes virus, matching the 99.382% combined identity reported in Table S2.



Nucleocapsid protein (S segment) | Glycoprotein precursor (M segment) | RNA-dependent RNA polymerase (L segment)

Fig. S9. Segment-specific dendrograms from the local protein comparison. Each protein-level dendrogram supports Andes virus as the closest reference for the Cruise sequence set, while the outbreak-specific Virological analyses provide the complementary nucleotide-level evidence for the 2026 cruise-associated cluster (51, 52).

Synthesis

Taken together, the public-health and genomic evidence supports three points that are directly relevant to host-pathogen interpretation. First, the 2026 event was a severe, internationally managed Andes virus cluster with eleven reported cases and three deaths as of the 13 to 14 May public updates (54, 55). Second, rapid open genomic work on Virological and Pathoplexus showed very high similarity among available outbreak genomes, identical S and M consensus sequences in the preliminary cluster data, only two synonymous L-segment SNPs among analyzed cases, and no obvious segment reassortment signal in the rapid phylogenetic summaries (51, 52, 60). Third, our local protein-level analysis independently places the Cruise sequence set very close to Andes virus across all three coding segments, but the current local result remains upstream of the outbreak-specific nucleotide question. This makes the local analysis useful for species-level or near-species-level assignment, while the Virological and Pathoplexus analyses remain the appropriate credited sources for 2026 outbreak genomics.

Supplementary structural variability in Protenix samples

For each of the eight Protenix systems folded with seed 101 and 10 samples, we computed pairwise Cα RMSD across ranked samples (Supplementary Figure S10). Monomer matrices use whole-chain Kabsch superposition and whole-chain RMSD. L-dornase matrices use L chain A for superposition and report dornase chain B RMSD, so the values emphasize variation in the predicted dornase pose after controlling for the viral polymerase frame. Diagonal entries are zero by construction. The rank-0 figures in the main text should therefore be read as representative top-ranked Protenix predictions, while this supplemental matrix shows that the sampled structural ensemble contains substantial alternative geometries, especially for GP and the L-dornase complexes.

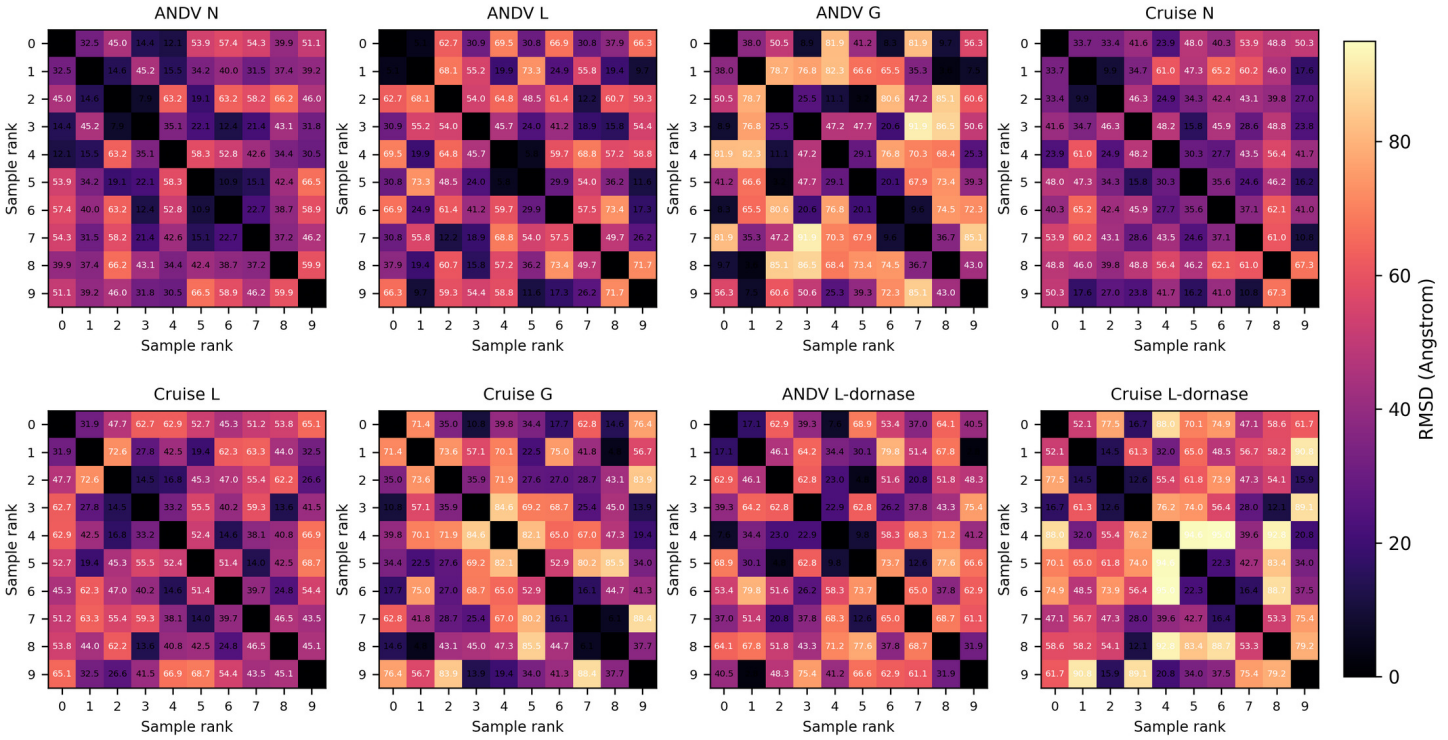


Fig. S10. Intra-sample RMSD variability across the 10 Protenix samples per system. Each panel is a 10 by 10 RMSD matrix across sample ranks from the same seed. Monomers are aligned and measured on chain A. L-dornase complexes are aligned on L chain A and measured on dornase chain B to quantify variation in the docked dornase pose. Values are in Å; diagonals are zero.

Supplementary detail for the Andes hantavirus screen

Top-ranked Atlas predictions per target

For the 428-aa N nucleocapsid protein (O36307), the raw Atlas probability leader in the PLI screen was lutein (raw_prob = 0.989), a carotenoid with no prior hantavirus connection; the PPI rank-1 was teduglutide (raw_prob = 0.880), a glucagon-like peptide 2 receptor agonist. Among all PLI candidates, mupirocin scored raw_prob = 0.980, placing it seventh by raw probability while achieving substantially higher structural validation (ipTM = 0.867, ipSAE = 0.144). The priority of mupirocin over lutein for experimental follow-up rests on its superior structural confidence, its established FDA approval and safety profile, and the precedent of the Xing et al. (34) N protein target-region study.

For the 2,153-aa L polymerase and cap-snatching endonuclease (Q9E005), the PLI rank-1 was fludarabine phosphate (raw_prob = 0.925), a ribonucleoside-diphosphate reductase inhibitor approved for chronic lymphocytic leukemia. The PPI rank-1 was corticorelin ovine triflutate (raw_prob = 0.712), a synthetic CRH diagnostic agent. Dornase alfa, ranked lower by raw probability (0.625), emerges as the priority candidate on the basis of its structural validation score (max directional ipSAE = 0.472, $> 50\times$ above any small molecule on the L protein), the position of the predicted interface on the cap-snatching endonuclease domain, where affinity of this predicted magnitude is a plausible route to transcriptional inhibition, and its inhaled delivery route directly targeting the primary HCPS organ.

For the 1,138-aa GP glycoprotein (Q9E006), the PLI rank-1 was N-acetylglucosamine (raw_prob = 0.999); the PPI rank-1 was prabotulinumtoxin A (raw_prob = 0.906). Among known antivirals benchmarked against GP, favipiravir scored highest (ipTM = 0.811, ipSAE = 0.029), followed by baloxavir (ipTM = 0.713/0.017) and remdesivir (ipTM = 0.661/0.009). No novel candidate was identified for GP that exceeded the known-antiviral structural confidence scores.

Matched ANDV-like sequence comparison

Because the Cruise strain proteins are near-identical but not fully identical to the reviewed ANDV reference outside N, we reran the reviewed ANDV and Cruise screens under the same checkpoints, FDA pools, code, and threshold before comparing them. The sequence comparison is intentionally simple and paired: N is 428/428 residues identical, L differs at 13 of 2,153 residues, and GP differs at 10 of

1,138 residues. The identical N sequences produced exactly identical PPI, PLI, antiviral, inter-actome, and GO enrichment rows, so N is treated as an internal reproducibility control rather than as a biological difference.

The L and GP substitutions changed model outputs modestly. Mean absolute human inter-actome deltas were 0.0019 for L and 0.0070 for GP, with top human-partner scores changing from 0.956 to 0.957 for L and from 0.934 to 0.938 for GP. In the FDA screen, all top PPI and PLI drug identities were retained across N, L, and GP. The priority candidates were also retained: mupirocin remained rank 7 against N with unchanged raw PLI probability (0.980), and dornase alfa improved from rank 5 to rank 4 against the Cruise L protein, increasing from 0.625 to 0.638 raw PPI probability. Thus, the matched comparison strengthens rather than weakens the L-protein follow-up argument while leaving the N-protein drug argument unchanged.

The clearest model-side differences appear in host inter-actome enrichment (Supplementary Figure S1; full per-protein enrichment plots in Supplementary Figures S4, S5, S6, and S7). L retains a dominant DNA repair and DNA replication profile, but the Cruise L profile is slightly stronger for DNA repair, double-strand break repair via homologous recombination, DNA replication, and mismatch repair, while reviewed ANDV is stronger for interstrand cross-link repair. GP shifts more qualitatively: Cruise GP gains cell-cell adhesion, integrin-mediated adhesion, and cell-matrix adhesion enrichment, while reviewed ANDV is stronger for apoptotic cell clearance and transmembrane transport. Read cautiously, these differences suggest that the Cruise protein set is not functionally novel at the scale of this model, but may modestly alter predicted host-interface emphasis in polymerase-linked nuclear repair pathways and glycoprotein-linked adhesion or transport pathways. These deltas are consistent with the small number of protein substitutions and do not change the drug-prioritization conclusions drawn from the reviewed ANDV screen.

Full folding score matrix

Table S3 reports the complete Protenix v2 ipTM/ipSAE matrix for all three ANDV proteins across novel candidates and known antivirals. Notable cross-target observations: remdesivir and baloxavir (both L protein drugs) score highest on the N protein rather than the L protein, and favipiravir (an RdRp inhibitor) scores highest on the GP protein. These off-target structural scores suggest the pipeline is capturing genuine cross-protein structural complementarity rather than simply recapitulating known mechanisms, which independently supports the significance of mupirocin's N protein score.

Table S3. Full Protenix v2 folding score matrix for ANDV proteins. ipTM: interface predicted TM-score. ipSAE: interaction prediction score from aligned errors. Novel candidates are bold and shaded. Dashes indicate the compound was not screened against that target.

Drug	N ipTM	N ipSAE	L ipTM	L ipSAE	GP ipTM	GP ipSAE
Novel candidates						
Mupirocin (Bactroban, FDA 1987)	0.867	0.144	—	—	—	—
Dornase alfa (Pulmozyme, FDA 1993)	—	—	0.780	0.472	—	—
Known antivirals (calibration)						
Remdesivir (Veklury)	0.853	0.095	0.656	0.009	0.661	0.009
Baloxavir marboxil (Xofluza)	0.815	0.077	0.629	0.009	0.713	0.017
Favipiravir (Avigan)	0.689	0.019	0.499	0.000	0.811	0.029
Ribavirin (Copegus)	0.662	0.012	0.437	0.000	0.602	0.006

Supplementary detail for the Bundibugyo ebolavirus screen

Qualifications on the antibody result

Three points bound the antibody-class recovery reported in the main text. First, none of the four antibodies is rank 1 on GP. The top biologic predictions against full-length GP are tagraxofusp (0.928), simoctocog alfa (0.911), and prabotulinumtoxin A (0.909), all large multi-domain or multi-chain proteins, and the same handful of large biologics recurs at rank 1 across most of the nine viral queries (Table S4). This is the biologic-side analogue of the partner-count inflation documented in Case Study I; the antibodies land in the top 10 to 15% of the pool rather than at the top of it. Second, discriminating GP from six structurally unrelated non-GP viral proteins is an easier task than discriminating a neutralizing antibody from a non-neutralizing one against the same antigen, and this screen demonstrates only the former. Third, the model scores the two shorter GP-gene entries (373 and 302 aa, lengths consistent with the secreted sGP and ssGP products of the edited GP locus) at 0.102 to 0.206, near the non-GP baseline rather than near full-length GP. Those products lack the GP2 subunit and much of the epitope surface these antibodies engage, which is the likely explanation.

Small-molecule leads and their artifact signature

The small-molecule results document a failure pattern rather than nominating candidates. The per-protein PLI leaders are not isolated compounds but tight pharmacophore clusters occupying consecutive ranks: vitamin D receptor agonists sweep VP30 (calcitriol 0.996, calcipotriene 0.995, calcifediol 0.995, doxercalciferol 0.995) and VP35 (calcipotriene 0.967, paricalcitol 0.966, doxercalciferol 0.959, calcitriol 0.958); SGLT2 inhibitors sweep VP40 (canagliflozin 0.984, dapagliflozin 0.975, empagliflozin 0.973); retinoic-acid receptor agonists sweep VP24 (trifarotene 0.992, palovarotene 0.989). *N*-acetylglucosamine is rank 1 on all three GP-gene entries at 0.997 to 0.999. Within-class rank coherence indicates that the model responds consistently to chemical structure, but it is not evidence of a therapeutic interaction, and the GlcNAc result is more parsimoniously read as compositional affinity for the heavily N-glycosylated, mucin-domain-bearing GP than as a binding event. One cross-pathogen observation warrants follow-up: mupirocin, the Case Study II lead against the ANDV nucleocapsid, is rank 4 against the BDBV NP at 0.949. Recurrence of the same compound at the top of two

unrelated viral nucleoproteins is either a generalizable nucleoprotein-groove affinity or a model-side nucleoprotein bias, and the two are not separable from these data.

Limitations of the ESMFold2 pass

Three properties of the folding run limit what its disagreement with Atlas can mean. Each antibody was folded as a single chain, either a heavy or a light variable domain alone and never as a paired Fv, so no complex in this set contains an intact paratope. Complexes exceeding the 768-residue input limit were split into sequence windows, so every full-length GP and L complex presents a partial antigen and the epitope need not lie in the folded window. Monomer confidence is low throughout, with median pLDDT of 30 for VP40, 34 for NP and VP24, and 50 for L, and ipSAE is undefined for all 176 ligand complexes in this pipeline. The pass is therefore underpowered by construction.

Caveats specific to this screen

(i) The antibody recovery identifies the target, not the effect. It states which viral protein the four drugs engage, not whether a Zaire-raised antibody neutralizes Bundibugyo virus; that requires neutralization assays (38).

(ii) The production checkpoint had no filovirus-specific homology holdout, so ebolavirus proteins and the approved antibodies may both appear in the training corpus. Read the antibody result as correct retrieval rather than blind discovery; the Atlas-Sars protocol is the control and has not been run here.

(iii) Five of nine viral proteins returned zero above-threshold human partners, so inter-actome coverage supports no conclusion about those proteins in either direction.

(iv) The three GP-gene entries are products of one edited locus and share substantial sequence, so their scores are not independent observations.

Per-protein top-ranked predictions

Table S4 lists the rank-1 biologic and rank-1 small molecule for each of the nine BDBV query proteins. Two patterns visible here are discussed in the main text. First, the biologic column is dominated by a small recurring set of large multi-chain proteins (prabotulinumtoxin A appears at rank 1 for three of nine queries), which is why the approved anti-GP antibodies land in the top 10 to 15% of the pool rather than at the top of it despite scoring their correct target. Second, the small-molecule column is organized by pharmacophore class rather than by antiviral plausibility, with vitamin D receptor agonists, retinoids, and SGLT2 inhibitors sweeping consecutive ranks on their respective targets.

Table S4. Rank-1 FDA-approved biologic and small molecule per BDBV query protein. Values are raw Atlas sigmoid probabilities from Atlas-PPI (biologics) and Atlas-PLI (small molecules). Withdrawn drugs excluded. None of these was structurally validated.

Accession	Gene	Length	Top biologic	Prob.	Top small molecule	Prob.
B8XCM7	NP	739	Prabotulinumtoxin A	0.884	Lutein	0.963
B8XCM8	VP35	341	Teriparatide acetate	0.830	Calcipotriene	0.967
B8XCM9	VP40	326	Prabotulinumtoxin A	0.682	Canagliflozin anhydrous	0.984
B8XCN0	GP	676	Tagraxofusp	0.928	*N*-Acetylglucosamine	0.997
B8XCN1	GP	373	Pramlintide	0.844	*N*-Acetylglucosamine	0.999
B8XCN2	GP	302	Pegunigalsidase alfa	0.929	*N*-Acetylglucosamine	0.999
B8XCN3	VP30	289	Parathyroid hormone	0.638	Calcitriol	0.996
B8XCN4	VP24	251	Prabotulinumtoxin A	0.848	Trifarotene	0.992
B8XCN5	L	2,210	Dornase alfa	0.735	Fludarabine phosphate	0.910

Table S5 reports the best rank achieved by each small-molecule antiviral control across the nine BDBV proteins. None of these compounds has demonstrated efficacy against Bundibugyo virus, and remdesivir underperformed both monoclonal antibody arms in the PALM trial (37), so these ranks are reported as a negative-control record rather than as misses.

Table S5. Small-molecule antiviral control ranks in the BDBV PLI screen. Best rank out of 2,638 FDA-approved small molecules across the nine BDBV query proteins, with the viral protein achieving that rank.

Control	Best-scoring viral protein	Raw prob.	Best rank (of 2,638)
Brincidofovir	NP	0.697	175
Ribavirin	L	0.221	397
Remdesivir	L	0.178	582
Favipiravir	L	0.043	1,944

Supplementary detail for the SARS-CoV-2 anosmia screen

Independent verification against published mechanisms

This audit is a verification exercise, separate from the chemosensory predictions. It asks whether the screen output agrees with what the primary literature already establishes for SARS-CoV-2 proteins whose mechanisms are published, in the same role that Spike plays as a within-screen specificity control in the main text. Proteins are named here because the mechanisms are public and no novelty is claimed for them, and the naming carries no mapping to the anonymized Protein A through Protein F designations used for the chemosensory leads.

Zazhytska et al. (24) established the non-cell-autonomous collapse of olfactory nuclear architecture through three experimental lines: hamster infection time-course (bulk RNA-seq + in situ Hi-C on FAC-sorted mature OSN nuclei at 1, 3, and 10 days post-infection); human autopsy (19 SARS-CoV-2 positive subjects + 3 controls); and UV-neutralised serum transfer to naive olfactory epithelium ex vivo. RNAscope detected SARS-CoV-2 S transcript in sustentacular cells but rarely in OMP+ OSNs. The Hi-C compartment scores around OR clusters persisted at 10 dpi after viral clearance. The UV-neutralised serum experiment showed that something released into the bloodstream is sufficient to alter OSN nuclear architecture in naive animals.

Key quotes: "OR downregulation likely represents a non-cell autonomous effect, since SARS-CoV-2 detection in OSNs is extremely rare both in human and hamster OEs" (24). "Genomic compartmentalization remains disrupted 10 dpi, when the virus is already cleared from the OE" (24). The senior-author commentary by Mukamel & Lomvardas is explicit about what remains unknown: "the specific molecules that are responsible are yet to be identified. Possibilities include non-viable viral particles, molecules secreted from infected cells such as cytokines" and "It is also unclear how such a molecule would elicit the profound changes in nuclear architecture in OSNs." Of the 19 human autopsy subjects, only 1 had documented anosmia; the molecular phenotype is therefore correlated with COVID death, not specifically with anosmia, in 18 of 19 patients.

The published biophysical evidence for a SARS-CoV-2 accessory protein engaging a host transmembrane domain through leucine-zipper interference is weak, which sets the standard this screen's predictions should be held to. Nguyen et al. (62) used wheat-germ cell-free protein synthesis of a full-length coronaviral accessory protein and a 33-residue E-cadherin transmembrane peptide (residues E641–R673) in ERGIC-mimicking lipid vesicles, measuring co-capture by dual-tag affinity pulldown and intermolecular solid-state NMR. The key caveats: no K_d, no SPR/BLI/ITC; full-length E-cadherin was never tested; no mammalian-cell experiment was performed. The authors' own framing is "hypothetical interferer" and "One could speculate that such a hydrophobic TM protein, carrying in addition a leucine zipper motif used also by cellular proteins, if expressed at high copy numbers in cells, has the potential to interfere with correct targeting and recycling of cellular proteins." The largest unbiased SARS-CoV-2/human interactome study (Gordon et al. (1)) recovered neither E-cadherin nor phospholamban for that protein. A hamster deletion experiment removing two adjacent accessory ORFs together showed partial anosmia rescue (2 of 8 versus 5 of 8 in wild-type) but cannot attribute the effect to either ORF individually.

Two simultaneous *Science* papers connected Spike cleavage, NRP1 engagement, and the vascular consequences of variant-dependent fusogenicity. Cantuti-Castelvetri et al. (63) showed NRP1 enhances SARS-CoV-2 entry in cell lines; a SARS-CoV-2 mutant lacking the furin cleavage site lost NRP1 dependence. Daly et al. (64) measured the Spike S1 CendR peptide / NRP1 affinity directly: "the b1 domain of NRP1 directly bound a synthetic S1 CendR peptide (679NSPRRAR685) with an affinity of 20.3 μM at pH 7.5" – a low/mid-micromolar affinity consistent with a co-receptor/facilitator role rather than a primary entry receptor. No paper has shown mature OSNs are infected via NRP1; the olfactory-bulb access claim in Cantuti-Castelvetri uses synthetic AgNP-CendR, never authentic virus.

For variant-specific olfactory tropism, Chen et al. (65) showed in hamsters: infected olfactory epithelium length was 79.2% (WA1), 70.3% (Delta), and 6.7% (Omicron), correlating with clinical anosmia rates of 52.7% (Delta period) vs 16.7% (Omicron period).

Brann et al. (66) resolved the cellular distribution of ACE2 and TMPRSS2 by analysing bulk and single-cell RNA-seq from mouse, non-human primate, and human olfactory mucosa, with immunostaining confirmation: "single cell sequencing revealed that ACE2 is expressed in support cells, stem cells, and perivascular cells, rather than in neurons" and "Immunostaining confirmed these results and revealed pervasive expression of ACE2 protein in dorsally-located olfactory epithelial sustentacular cells." The consensus that mature OSNs do not express canonical SARS-CoV-2 entry receptors is robust across modalities, though scRNA-seq dropout means an absolute lower bound on OSN ACE2 expression is dataset-dependent.

The proposed ORF3a viroporin-to-NLRP3 mechanism is weakened by Miller et al. (67), who used HEK293 whole-cell patch clamp, Xenopus oocyte two-electrode voltage clamp, ACMA flux assay, and proteoliposome patch clamp at standard 1:100 protein:lipid ratio: "Neither SARS-CoV-2 nor SARS-CoV-1 Orf3a form functional ion conducting pores" and "From our data, we conclude that SARS-CoV-1 and SARS-CoV-2 Orf3a are not viroporins." Miller et al. identified the relevant mechanism as trafficking (VPS39/HOPS/autophagosome-lysosome fusion), not ion conduction. Xu et al. (68) showed ORF3a triggers IL-1β processing, with K^+ efflux inferred from pharmacology rather than directly measured; combined with Miller 2023, the K^+ efflux narrative specifically is what is contested. No primary paper has demonstrated ORF3a-driven NLRP3 activation in taste-bud or lingual epithelial tissue.

Doyle et al. (69) connected SARS-CoV-2 exposure to Type II taste cells by detecting ACE2 in PLCB2+ Type II taste cells and reported replicative-strand ISH in Type II cells of the single acute-FP specimen that contained taste buds, with significantly reduced Ki-67+ basal cells in COVID-positive fungiform papillae ($P < 0.0001$). No infectious virus was recovered. Critically, three independent psychophysical-testing series establish that objectively measured complete ageusia occurs in approximately 1% of COVID patients: 1 of 72 (1.4%) by four-tastant gustometry (70), 1% (3 of 300) by Taste Strips at 6 months (71), and 12% objective dysgeusia with 88% normogeusic (72). The 38–88% headline ageusia rates from cohort questionnaires reflect retronasal flavour loss being conflated with literal taste loss, as Le Bon et al. explicitly note: "Most patients with COVID-19 tend to unconsciously conflate taste loss and olfactory loss, even when taste function is clearly defined beforehand." The Protein B/TAS2R prediction, if validated, would explain specifically this rare-but-real objective bitter-taste phenotype.

Two salivary proteomics studies provide a second mechanistic axis through CA6 and cystatins, reporting downregulation of carbonic

anhydrase VI (CA6) and S-type cystatins (CST1, CST2, CST4) in COVID patients. Muñoz-Prieto et al. (73) found CA6 log2FC = -1.78 ($p = 0.007$), CST1 -1.57 ($p = 0.001$), CST4 -2.54 ($p = 0.001$) in severe COVID vs healthy controls. Aita et al. (74) found CA6 direction opposite in the symptomatic/asymptomatic axis, and noted ageusia was "extremely rare" (1/35 patients) in their cohort; the proteomic CA6 signal is therefore decoupled from clinical ageusia. The CA6 knockout mouse study (75) shows altered bitter aversion without morphological taste-bud abnormality, placing CA6 as a taste-sensitivity modifier rather than a structural requirement for taste-bud existence. No proteomics study to date compares ageusic vs non-ageusic COVID patients head-to-head.

The evidence for ORF8 as a secreted virokine begins with Matsuoka et al. (76), who demonstrated that "approximately 41% of the ORF8 expressed in 293T cells were secreted extracellularly as a glycoprotein homodimer", via a conventional Golgi-mediated pathway. Wu et al. (77) provided the first quantitative serum measurement: "The median serum levels of ORF8 were 1 ng/mL in hospitalized COVID-19 patients and 414.5 ng/mL in ICU-admitted COVID-19 patients" (this 414.5 ng/mL figure is unusually high relative to typical serum cytokine levels and should be treated cautiously in downstream Kd-occupancy reasoning). Lin et al. (78) demonstrated direct ORF8/IL-17RA binding by yeast two-hybrid, co-IP, and GST pulldown with domain mapping (fnIII_D2 domain of IL17RA required); no SPR/BLI K_d was reported. The strong antibody response to ORF8 in patient sera (96.5% anti-ORF8 positive by LIPS assay; Hachim et al. (79)) is consistent with extracellular exposure during infection.

Extracellular behavior is documented unevenly across the accessory proteins, which is the relevant constraint on any mechanism requiring one of them to reach a ciliary or apical membrane. ORF8 is the only well-quantified case, above. For ORF7b, Schaecher et al. (80) showed that the SARS-CoV-1 ortholog is expressed in infected cells and incorporated into virions, which places virion-associated protein in mucus on lysis, but no primary paper reports extracellular detection of the SARS-CoV-2 protein in patient body fluids, exosomes, or extracellular vesicles; Nguyen et al. (62) describe it as ER-retained. For ORF6, Nishide et al. (81) characterized spontaneous self-assembly and reported that "Exogenous ORF6 protein released from infected cells (living or dead) could be internalized", without a serum or plasma concentration measurement. For ORF10, translation in vivo is contested outright: direct RNA sequencing and ribosome profiling place it at or below the level of the 3′ untranslated region (82, 83), so any prediction involving it describes a polypeptide whose existence is in question.

Extracellular accessibility of the chemosensory leads

For a "Protein A or Protein B binds chemosensory receptor" mechanism to operate *in vivo,* the viral protein must reach OSN cilia (which project into olfactory mucus produced by sustentacular cells) or taste-cell apical membranes. ORF8 is the established precedent: 41% secreted via conventional Golgi pathway (76), detected in patient serum at ng/mL concentrations by LC-MS/MS and ELISA (77), with strong seroresponse indicating extracellular immune exposure.

For Protein A, no published primary paper documents extracellular detection in patient body fluids, exosomes, or extracellular vesicles, and the published characterization describes an ER-retained protein. Its SARS-CoV-1 ortholog is incorporated into virions, so virion-associated protein would be released into mucus on viral lysis, placing it in the same geometric compartment as OSN cilia via apical secretion from sustentacular cells. For Protein B, the published work reports self-assembly and release from infected cells with subsequent internalization by neighboring cells, but no serum or plasma concentration measurement exists. For Protein F, translation itself is contested and no extracellular detection has been reported. Primary sources are given under named proteins in the verification subsection above.

The geometric argument for OSN accessibility is: the OSN dendrite ends in a knob from which 20–35 cilia project into the olfactory mucus produced by the same sustentacular cells the virus infects. Anything an infected sustentacular cell secretes apically enters the mucus bathing the OSN cilium meshwork within milliseconds of diffusion. Viral protein (anti-S, anti-N immunofluorescence) persists in nasal mucociliary mucus for 81–102 days post-symptom-onset in all 8 of 8 patients with persistent olfactory dysfunction (27). Local secretion from infected sustentacular cells is therefore geometrically more direct than systemic circulation for reaching OSN cilia.

OR-driven feedback and anosmia precedent

If the Protein A/OR binding is real and the protein reaches OSN cilia, would broad OR antagonism produce an anosmia phenotype? The published OSN biology provides individual molecular pieces that are consistent with this possibility.

Dalton et al. (29) showed that OR translation generates ER stress that triggers a PERK → ATF5 → Adcy3 feedback: "OR expression activates Perk in the neuronal ER, which phosphorylates the translation initiation factor eif2alpha" and "translation of the nuclear form of Atf5 induces the transcription of Adenylyl Cyclase 3 (Adcy3)." Shayya et al. (30) extended this: "the PERK arm of the unfolded protein response (UPR) regulates both the glomerular coalescence of like axons, and the specificity of their projections." Pourmorady et al. (31) showed that the OR transcript itself reinforces the dominant Greek Islands hub and inhibits competing hubs: "coding OR mRNAs possess non-coding functions that influence nuclear architecture, enhance their own transcription and inhibit transcription from their competitors." In the Adcy3 knockout mouse, MOE cell density sharply decreases by postnatal day 90 with "incomplete membranes, swollen cilia" in OSNs (84).

The end-to-end chain (broad OR antagonism → reduced OR translation flux → weakened PERK/ATF5/Adcy3 feedback → Greek Islands hub instability → OR downregulation matching the Zazhytska signature) has not been directly tested in any non-COVID system. The individual molecular pieces are precedented; the coupling is a hypothesis motivated by the screen.

Caveats specific to this screen

(i) The score (> 0.9) is a ranking output, not a calibrated affinity or probability.

(ii) Family co-saturation: Protein A returns near-complete coverage of two large, internally homologous human families on one partner set (94% of olfactory receptors, 90% of mitochondrial Complex I). A shared latent-space feature explains this more parsimoniously than two distinct binding surfaces; a cross-virus comparator is the experiment that separates them.

(iii) Partner-count inflation: the chemosensory leads return 225 to 616 partners above threshold, more than interaction-surface considerations predict at their size.

(iv) A global Bonferroni threshold at family-wise $\alpha = 0.05$ across 15 proteins $\times$ 6 libraries corresponds to approximately adj$p < 5 \times 10^{-9}$. The Protein A OR call clears it by many orders of magnitude; the Protein B taste receptor call (GO:0008527, adj$p = 7.8 \times 10^{-7}$) is borderline. Reported adj$p = 0.0$ values are floating-point underflow on the hypergeometric statistic, not posteriors.

(v) Protein F translation in vivo is contested (82, 83), so its predictions describe a polypeptide whose existence is in question.

Per-protein screen results from the SARS-CoV-2 anosmia screen.

The 419-aa nucleoprotein N (P0DTC9), which has no transmembrane helices and packages the viral genome, produced 31 partners above threshold, with leading enrichments in RNA binding (GO:0003723, adj$p = 2.8 \times 10^{-24}$, overlap 26 of 31), cytoplasmic stress granule (GO:0010494, adj$p = 1.5 \times 10^{-14}$, overlap 10), CRD-mediated mRNA stabilisation (GO:0070934, adj$p = 6.4 \times 10^{-16}$, overlap 7), Reactome Metabolism of RNA (adj$p = 1.4 \times 10^{-10}$, overlap 14). The model recovers the canonical RNA-binding / LLPS / stress-granule profile from sequence alone.

The 275-aa, three-pass transmembrane protein ORF3a (P0DTC3) produced 200 partners, with leading enrichments in regulation of sodium ion transport (GO:0002028, adj$p = 4.2 \times 10^{-6}$, overlap 6), voltage-gated calcium channel activity (GO:0005245, adj$p = 7.2 \times 10^{-5}$, overlap 6), KEGG Cardiac muscle contraction (adj$p = 4.9 \times 10^{-7}$, overlap 10), ion homeostasis (R-HSA-5578775, adj$p = 9.5 \times 10^{-5}$, overlap 7). The literature framing of ORF3a as a K+/Ca2+-permeable viroporin is contested in the main-text literature audit; the model recovers the ion-channel/cardiac-action-potential interactome that ORF3a's published interactome partners share, regardless of whether ORF3a itself is a conducting channel.

The 75-aa, single-pass Envelope protein (P0DTC4) produced 218 partners, with leading enrichments in ventricular cardiac muscle cell action potential (GO:0086005, adj$p = 1.7 \times 10^{-4}$, overlap 5), potassium channel regulator activity (GO:0015459, adj$p = 4.7 \times 10^{-5}$, overlap 7). The model places Envelope alongside cardiac action-potential/K+ regulators, parallel to ORF3a.

The 222-aa, three-pass Membrane protein (P0DTC5) produced 284 partners, with leading enrichments in sodium ion transmembrane transport (GO:0035725, adj$p = 5.2 \times 10^{-20}$, overlap 22), gap junction channel activity (GO:0005243, adj$p = 3.8 \times 10^{-13}$, overlap 11), plasma membrane (GO:0005886, adj$p = 1.7 \times 10^{-46}$, overlap 193), Reactome transport of small molecules (adj$p = 6.6 \times 10^{-32}$, overlap 65). The strong plasma membrane/Na^+/K^+ transport enrichment is expected for a 3-TM structural protein; the gap-junction-channel signal was not in the prior SARS-CoV-2 literature.

The 1,273-aa, single-pass Spike protein (P0DTC2) produced 195 partners, with leading enrichments in cell-matrix adhesion (GO:0007160, adj$p = 3.3 \times 10^{-27}$, overlap 26), cell surface (GO:0009986, adj$p = 3.4 \times 10^{-32}$, overlap 53), integrin cell surface interactions (R-HSA-216083, adj$p = 3.8 \times 10^{-26}$, overlap 23). Spike is not enriched for olfactory receptor activity; this rules out a trivial artifact in which all SARS-CoV-2 proteins look like olfactory-receptor binders to the model.

The 121-aa, single-pass ORF7a protein (P0DTC7) produced 38 partners, with leading enrichments in Toll-like receptor signaling pathway (GO:0002224, adj$p = 5.2 \times 10^{-14}$, overlap 8), MyD88 distinct input/output pathway (WP3877, adj$p = 2.8 \times 10^{-16}$, overlap 8), KEGG Toll-like receptor signaling (adj$p = 5.3 \times 10^{-8}$, overlap 7). The canonical UniProt function is BST2/tetherin antagonism, which the model does not directly recover; it surfaces the TLR/MyD88 immune-signalling axis instead.

The 121-aa ORF8 protein (P0DTC8), which has no transmembrane helix, produced 408 partners, with leading enrichments in defense response to bacterium (GO:0042742, adj$p = 1.1 \times 10^{-17}$, overlap 29), hormone activity (GO:0005179, adj$p = 1.4 \times 10^{-17}$, overlap 25), extracellular region (GO:0005576, adj$p = 1.7 \times 10^{-99}$, overlap 207), Reactome Defensins (R-HSA-1461973, adj$p = 1.3 \times 10^{-51}$, overlap 37). 207 of 408 partners are extracellular proteins, recovering ORF8's published secreted-virokine character. The defensin enrichment extends its innate-barrier-interference role.

The putative 73-aa, single-pass ORF9c protein (P0DTD3) produced 247 partners, with leading enrichments in detoxification of copper ion (GO:0010273, adj$p = 4.8 \times 10^{-18}$, overlap 12), metallothioneins bind metals (R-HSA-5661231, adj$p = 7.8 \times 10^{-14}$, overlap 9), zinc homeostasis (WP3529, adj$p = 2.3 \times 10^{-8}$, overlap 9). If real, a direct interaction between ORF9c and metallothioneins would provide a concrete molecular handle for the correlative Zn/COVID severity literature.

The 97-aa ORF9b protein (P0DTD2), which has no transmembrane helix, produced 268 partners, with leading enrichments in RNA polymerase II preinitiation complex assembly (GO:0051123, adj$p = 3.6 \times 10^{-71}$, overlap 46), TFIID complex (GO:0005669, adj$p = 4.7 \times 10^{-48}$, overlap 32), KEGG Basal transcription factors (adj$p = 4.9 \times 10^{-33}$, overlap 24). The model misses the canonical mitochondrial/TOMM70/IFN-suppression axis and predicts that ORF9b targets the RNA Pol II preinitiation complex/TFIID, consistent with a host-transcription-shutoff role.

Protein B produced 225 partners, with leading enrichments in detection of bitter taste stimulus (GO:0001580, adj$p = 3.0 \times 10^{-5}$, overlap 7), taste receptor activity (GO:0008527, adj$p = 7.8 \times 10^{-7}$, overlap 6: TAS2R4, TAS2R7, TAS2R8, TAS2R9, TAS2R10, FFAR4), bitter taste receptor activity (GO:0033038, overlap 6: TAS2R4, R7, R8, R9, R10, R38), pheromone binding (GO:0005550, overlap 3: VN1R2, VN1R5, VN1R17P). The model misses Protein B's canonical RAE1/NUP98 nucleocytoplasmic-transport function entirely.

Protein A produced 553 partners, the second-highest count among the SARS-CoV-2 proteins screened, with leading enrichments in olfactory receptor activity (GO:0004984, adj$p = 0.0$ [floating-point overflow, not a Bayesian posterior], overlap 409 of 433), KEGG Olfactory transduction (adj$p = 0.0$, overlap 376), Reactome expression and translocation of olfactory receptors (adj$p = 0.0$, overlap 387), mitochondrial Complex I Electron Transport Chain (WP111, adj$p = 5.6 \times 10^{-96}$, overlap 73), NADH dehydrogenase activity (GO:0008137, overlap 37 of 41 Complex I subunits). The same partner set also covers 406 of 685 human GPCRs. The co-occurrence of near-saturated overlaps on unrelated protein families is discussed in the calibration and caveats subsections of the main-text Case Study I.

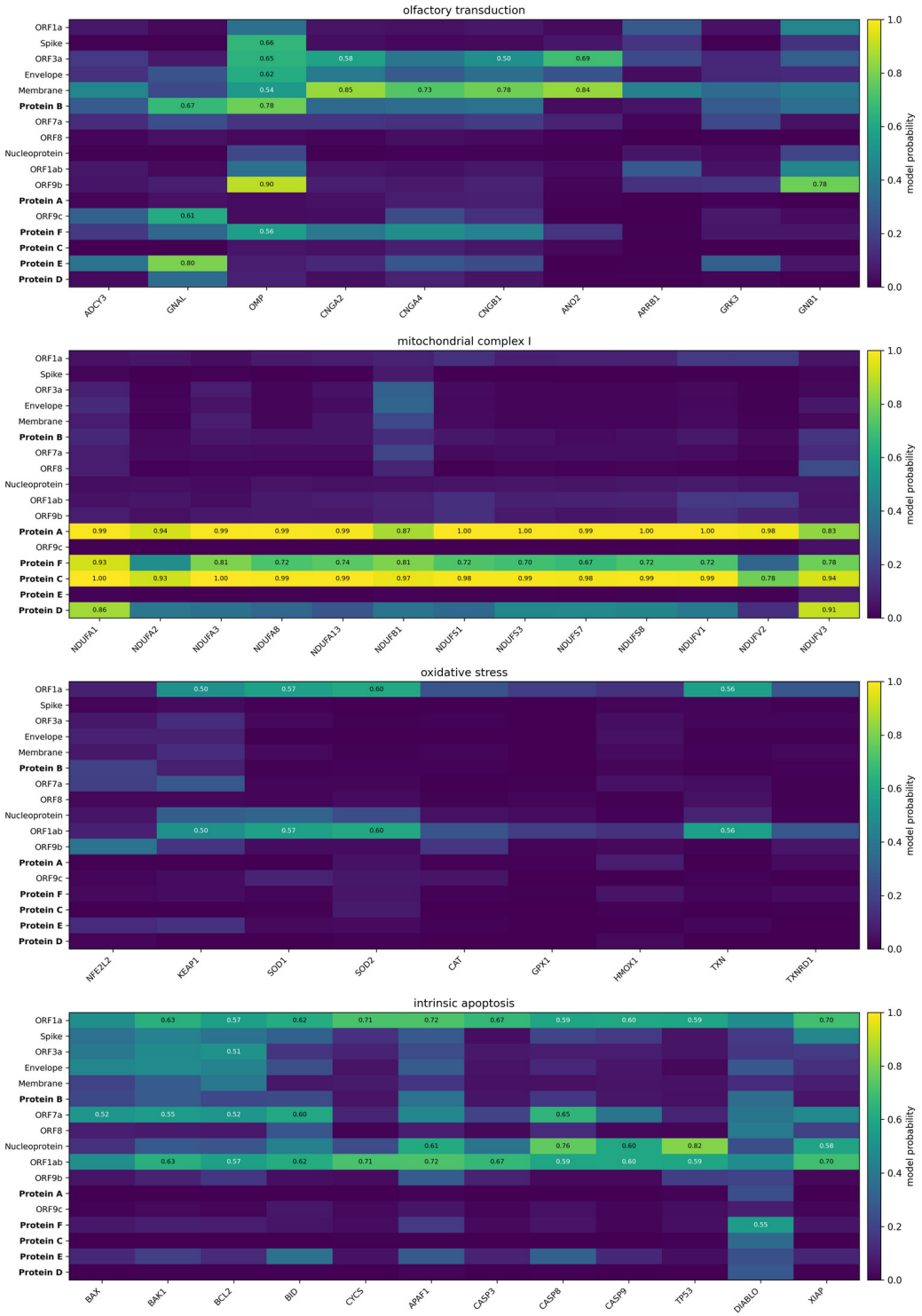


Fig. S11. Direct SARS-CoV-2 × human pathway interaction heatmap. Raw Atlas-Sars scores (no threshold) for all 17 SARS-CoV-2 proteome entries (rows) against individual human proteins in four pathway gene sets (columns). Color scale: dark (0) to yellow (1); cells at or above 0.5 are annotated. The six chemosensory leads are labeled Protein A to F in bold and their accessions withheld; calibration proteins recovering already-published mechanisms keep their common names. Protein A saturates the mitochondrial Complex I panel (0.83 to 1.00 across all 13 subunits), and Proteins C, D, and F show the same Complex I elevation, which is the family co-saturation signature discussed in the main text. The olfactory transduction panel contains the downstream signal-transduction cascade (ADCY3, GNAL, OMP, CNG channels) rather than the olfactory receptor family itself, so the 409 of 433 olfactory receptor overlap reported for Protein A in the main text is not displayed here; in this cascade panel Protein A is not elevated, whereas Membrane, ORF9b, and Protein B are. Spike shows no elevated olfactory signal, providing a within-screen specificity control.

Protein F, a putative ORF, produced 290 partners, with leading enrichments in olfactory receptor activity (GO:0004984, adj$p = 2.8 \times 10^{-23}$, overlap 45), serotonin-gated cation-selective signaling pathway (GO:0140227, adj$p = 2.0 \times 10^{-7}$, overlap 5), KEGG Olfactory transduction (adj$p = 6.3 \times 10^{-18}$, overlap 39). Independent direct-RNA-sequencing and ribosome-profiling studies argue that Protein F is not translated at biologically meaningful levels: "[Protein F] is represented by only one read in DNB data (0.000009% of viral junction-spanning reads) and it was not supported at all by DRS data" (82), and its "overall ribosome occupancy was lower than for every

other canonical accessory gene and was similar in occupancy to the 3' untranslated region" (83). Any Protein F prediction should be treated as the binding profile of a polypeptide whose existence in vivo is contested.

Protein C, a putative ORF, produced 452 partners, with leading enrichments in olfactory receptor activity (GO:0004984, adj$p = 0.0$, overlap 287), detection of chemical stimulus involved in sensory perception of smell (GO:0050911, adj$p = 4.7 \times 10^{-190}$, overlap 124), mitochondrial inner membrane (GO:0005743, adj$p = 1.4 \times 10^{-83}$, overlap 116). Same OR and mitochondrial double signal as Protein A. Partner-count inflation applies here (see main-text caveats).

Protein D produced 616 partners, the highest count among the SARS-CoV-2 proteins screened, with leading enrichments in G protein-coupled receptor activity (GO:0004930, adj$p = 5.3 \times 10^{-283}$, overlap 297), KEGG Olfactory transduction (adj$p = 1.6 \times 10^{-225}$, overlap 223), Reactome expression and translocation of olfactory receptors (adj$p = 1.1 \times 10^{-252}$, overlap 234), GPCRs Class A Rhodopsin Like (WP455, adj$p = 3.0 \times 10^{-70}$, overlap 92). A third independent COVID accessory protein converging on ORs/Class A GPCRs.

Protein E, a putative ORF, produced 293 partners, with leading enrichments in G protein-coupled receptor signaling pathway (GO:0007186, adj$p = 3.0 \times 10^{-7}$, overlap 27), G protein-coupled receptor activity (GO:0004930, adj$p = 4.3 \times 10^{-5}$, overlap 29), GPCR ligand binding (R-HSA-500792, adj$p = 5.8 \times 10^{-8}$, overlap 28). Broad-GPCR rather than specifically OR-enriched; a more diffuse signal than Protein A/Protein C/Protein D.