The probability of cryptic pocket opening controls functional tradeoffs in filovirus immune evasion.

Upasana L. Mallimadugula[1], Matthew A. Cruz[1], Neha Vithani[1], Maxwell I. Zimmerman[1] and Gregory R. Bowman[2*]

[1]Department of Biochemistry & Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110, USA
[2]Department of Biochemistry & Biophysics and Bioengineering, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[*]To whom correspondence should be sent: grbowman@seas.upenn.edu

Abstract
Cryptic pockets are of growing interest as potential drug targets. However, it remains unclear whether they contribute to protein function or if they are merely happenstantial features that can easily be evolved away to achieve drug resistance. Here, we explore whether a cryptic pocket in the Interferon Inhibitory Domain (IID) of viral protein 35 (VP35) of Zaire ebolavirus has a functional role. We use simulations and experiments to study the relationship between cryptic pocket opening and double-stranded RNA (dsRNA) binding of the IIDs of two other filoviruses, Reston and Marburg. These homologs have structures nearly identical to Zaire but block different interferon pathways because Marburg IID only binds the backbone, while Zaire and Reston IIDs bind both backbone and blunt ends of dsRNA. We hypothesized that these differences arise from alterations in their pocket dynamics. Simulations and thiol-labeling experiments demonstrate that Reston has a lower probability of opening the cryptic pocket while Marburg has a higher probability than Zaire. Subsequent dsRNA-binding assays with different length substrates suggest that closed conformations preferentially bind dsRNA blunt ends while open conformations prefer binding the backbone. Further, a point mutation that increases the probability of cryptic pocket opening in Reston shows that the open states prefer backbone binding, while a substitution that lowers the probability of pocket opening in Zaire improves binding to blunt ends. These results demonstrate the cryptic pocket controls a functional tradeoff, suggesting cryptic pockets are under selective pressure and may be difficult to evolve away to achieve drug resistance.

Introduction

Cryptic pockets have garnered significant attention, particularly for their potential as drug targets, but it remains unclear whether they play a role in protein function. Cryptic pockets are pockets that are not observed in the experimentally obtained structure of a protein but form due to thermal fluctuations in the native structure(1, 2). They are an interesting class of protein dynamics as they are increasingly being explored as drug targets(3, 4). Allosteric modulators that target cryptic pockets provide many advantages over orthosteric drugs(5). These include non-competitive regulation, higher specificity due to greater variation of pocket dynamics within protein families than variation in active or functional sites(6, 7), and the possibility of enhancing and not just inhibiting function(8). Further, cryptic pockets can provide a means to target proteins that appear to be undruggable due to a lack of potential binding pockets in experimentally-derived snapshots of the protein (9). However, successfully drugging a cryptic pocket would create a selective pressure for the organism to evolve protein variants that lack the pocket to achieve drug resistance. If cryptic pockets are happenstantial features that have no functional significance, then evolving them away could be trivial. On the other hand, it could be impossible to evolve away a cryptic pocket if the pocket is an inevitable consequence of the protein's topology. Alternatively, it could be difficult to evolve away a cryptic pocket if the open state plays a functional role. To explore these possibilities, it would be useful to study the conservation of cryptic pockets across protein variants.

However, studying cryptic pockets and assessing their functional relevance has been challenging. Cryptic pockets have largely been identified serendipitously when inhibitor-bound structures of proteins are solved and show the inhibitor binds in a cryptic pocket (10–12). While finding such structures proves the existence of cryptic pockets, it does not provide a facile means to quantify the probability of pocket opening or study the effects of sequence variation on pocket opening in the absence of a ligand.

Here, we explore the functional significance of a cryptic pocket that we recently discovered in VP35 protein from Zaire ebolavirus, which is the virus that causes the disease commonly known as Ebola. VP35 plays an essential role in filovirus immune evasion by binding dsRNA that is formed during replication of the viral genome to prevent these nucleic acids from being discovered by host innate immune receptors such as RIG-I and MDA5 (13–17). The dsRNA binding affinity of the IID is a significant determinant of virulence, making it an appealing therapeutic target(18, 19). Though a few studies have attempted to find small molecule drugs to target this protein(20–22), none have resulted in a successful drug discovery campaign. The IID interacts with RNA via flat surfaces that are often difficult to drug, with some going as far as calling these surfaces 'undruggable' (23). We discovered a cryptic pocket in VP35 that allosterically controls dsRNA binding, thereby providing a potential means to target this protein(24). However, targeting this pocket would be of little utility if the pocket is just happenstance and the virus can easily evolve resistance by acquiring mutations that prevent the pocket from opening without any fitness cost. Therefore, it is important to understand if the cryptic pocket has a functional role that may be under selective pressure and make the evolution of drug resistance more difficult.

To explore the functional relevance of this cryptic pocket, we used a combination of simulations and experiments to study the relationship between cryptic pocket opening and dsRNA binding in the IIDs of two other filoviruses, Reston ebolavirus (Reston) and Marburg marburgvirus (Marburg). Zaire binds to dsRNA blunt ends, thereby blocking RIG-I binding to the RNA. It can also bind to the backbone to block MDA5 binding (25–27). In contrast, Marburg binds the backbone but not the blunt ends, thereby blocking MDA5 binding but is unable to block RIG-I binding to shorter blunt ended RNAs (27). Reston is known to bind both blunt ends and the

backbone of the RNA but is a slightly weaker inhibitor of the interferon response than Zaire (25). However, it is difficult to explain these differences given that the crystal structures of all three variants are essentially identical in both their RNA-bound and free forms (Fig 1) and all three variants have high sequence similarity in residues directly interacting with dsRNA (Fig S1) (27). Therefore, we hypothesized that differences in cryptic pocket dynamics are responsible for the functional differences between VP35 homologs.

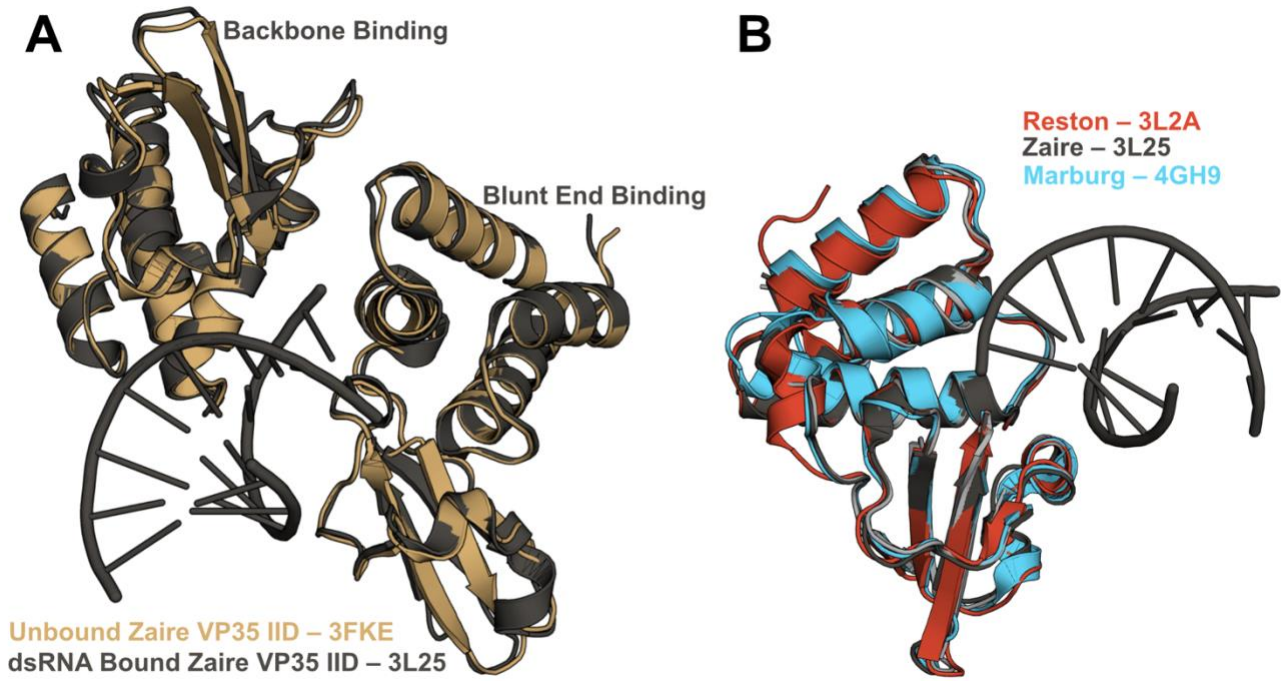

Figure 1: Crystal structures of the RNA-bound and unbound states of all three VP35 IIDs are nearly identical. A) Crystal structures of Zaire IID alone (3FKE, brown) and bound to 8-bp dsRNA (3L25, gray) show that both the blunt end and backbone binding poses of Zaire are nearly identical to the unbound (apo) structure. B) Overlay of Zaire IID bound to 8-bp dsRNA (3L25, gray) with unbound structures of Reston IID (3L2A, red) and Marburg IID (4GH9, blue)

Results

Simulations predict the probability of cryptic pocket opening is low in Reston and high in Marburg compared to Zaire.

To assess the probability of cryptic pocket opening in the three filoviruses, we first used molecular dynamics simulations guided by adaptive sampling. We first ran 8 μs of simulation of each variant using a goal-oriented adaptive sampling algorithm called FAST (28) that balances between broad exploration of conformational space and focusing data acquisition on conformations with more open pockets. We then built a Markov State Model (MSM) of each homolog from these simulations and used the centers as seeds for long MD simulations on the Folding@home distributed computing platform (29). We collected a total of 126 and 124 μs of data for Reston and Marburg IIDs. Then we built a new MSM for each variant that incorporates the data from Folding@home. The cryptic pocket observed in Zaire IID occurs as the helix spanning residue 305 to 309 moves away from the alpha helical domain. Therefore, we characterize pocket opening based on the distance between residues 236 and 306 in Zaire, which corresponds to residues 225 and 295 in Reston and Marburg. For the sake of brevity,

unless mentioned otherwise, we will use Reston and Marburg numbering going forward. We thus obtain the probability distribution of the distance between residues 225 and 295 based on the equilibrium probability of each structure in the MSM (Fig 2).

We find that Marburg has a higher probability of opening than Zaire, but that Reston has a significantly lower probability (Fig 2). Furthermore, Marburg opens more widely than the other variants, as judged by reaching larger distances between residues 225 and 295 (Fig 2 insets). The differences in pocket opening have an interesting correspondence to the fact that Marburg prefers to bind the dsRNA backbone, while Reston and Zaire bind both blunt ends and the backbone. We hypothesized that open pocket conformations preferentially bind the backbone while closed conformations preferentially bind blunt ends.

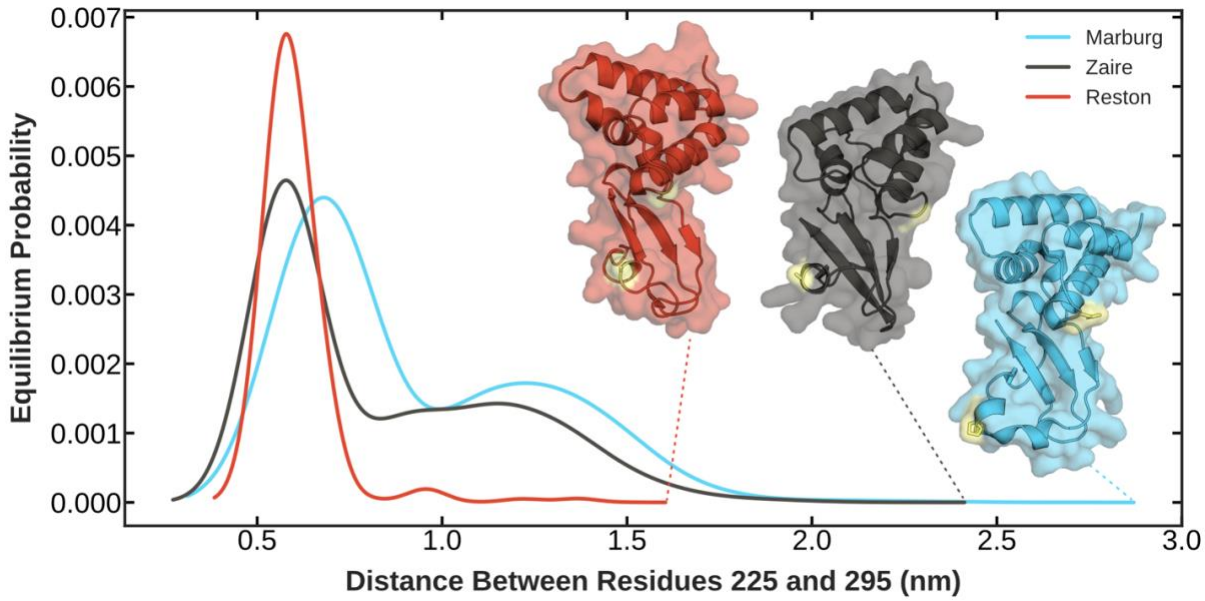

Figure 2. Simulations predict the probability of pocket opening is lowest in Reston and highest in Marburg. Each curve is the probability distribution of the distance between two residues that serves as a proxy for pocket opening. Reston VP35 IID (red) shows the least probability of opening the pocket. Zaire VP35 IID shows a greater probability of opening the pocket as well as an increased maximum distance of pocket opening. Marburg VP35 IID shows the highest probability of pocket opening and the highest maximum distance of pocket opening. The structure with the largest pocket (i.e. largest distance between the two residues) for each variant is shown in ribbon with a transparent surface, using the same color scheme as for the probability distributions. Residues 225 (236 in Zaire) and 295 (306 in Zaire) are shown in yellow sticks.

Thiol labeling experiments confirm Marburg has the highest probability of opening and Reston the lowest.

As a first test of our computational predictions, we used thiol labeling to experimentally measure the probability that the cryptic pocket is open in each IID homolog. In these experiments, we measure the rate of covalent modification of a cysteine in the cryptic pocket, as we have done previously with β-lactamases (30, 31) and Zaire IID (24). In these experiments, 5,5'-dithiobis-(2-nitrobenzoic acid), or DTNB, is added to the protein sample. In the presence of an oxidized (solvent exposed) thiol group of a cysteine residue, the disulfide bond between the two TNB molecules breaks, and one TNB molecule attaches to the cysteine via a disulfide bond. The free

TNB molecule left from this reaction absorbs light at 412 nm. Therefore, when a cysteine buried inside a pocket is exposed to solvent as the pocket opens, we observe an exponential increase in absorbance at a rate that depends on the opening and closing rates of the pocket. We quantify this using the Linderstrøm-Lang model (see Materials and Methods). We focus on thiol labeling of cysteine 296 as its solvent exposure is correlated with the extent of pocket opening (as measured by the distance between residue 225 and 295) in our simulations (Fig 3A, B and C), whereas the solvent exposure of other cysteines are not correlated with pocket opening (Fig 3E, F and G, and Fig S3).

As expected from the MSMs, we observe the highest probability of pocket opening in Marburg, followed by Zaire and Reston in that order. Thiol labeling of Reston IID and Marburg IID fit to one exponential per cysteine, each with the same amplitudes, indicating that all the cysteines in each variant get labeled (Fig S4). We performed point mutations of individual cysteines to serines to assign observed labeling rates to the cysteines as we have previously done for Zaire IID (Fig S5). We compare these results to the labeling rates we previously observed for Zaire (24). The cysteines in Reston's cryptic pocket label far more slowly than those in Zaire, while the labeling rates of the cysteines in Marburg are intermediate between Zaire and Reston (Fig 3D and H). These data were fit to the Linderstrøm-Lang model (see Materials and Methods) to quantify the probability and kinetics of pocket opening (Supplementary Table 1). In all three homologs, C296 labels faster than expected from labeling of the unfolded fraction (Fig S6), suggesting that the pocket does exist in all three homologs. The equilibrium constants for C296 exposure (Eq 3 in Materials and Methods) for Reston, Zaire and Marburg IIDs are 0.078 $\pm$ 0.001, 0.413 $\pm$ 0.009, 5.2 $\pm$ 0.5. These correspond to probabilities of pocket opening of 0.072, 0.292, and 0.839 for Reston, Zaire, and Marburg, respectively. Therefore, in agreement with our computational prediction, we observe the highest probability of pocket opening in Marburg, followed by Zaire and Reston in that order.

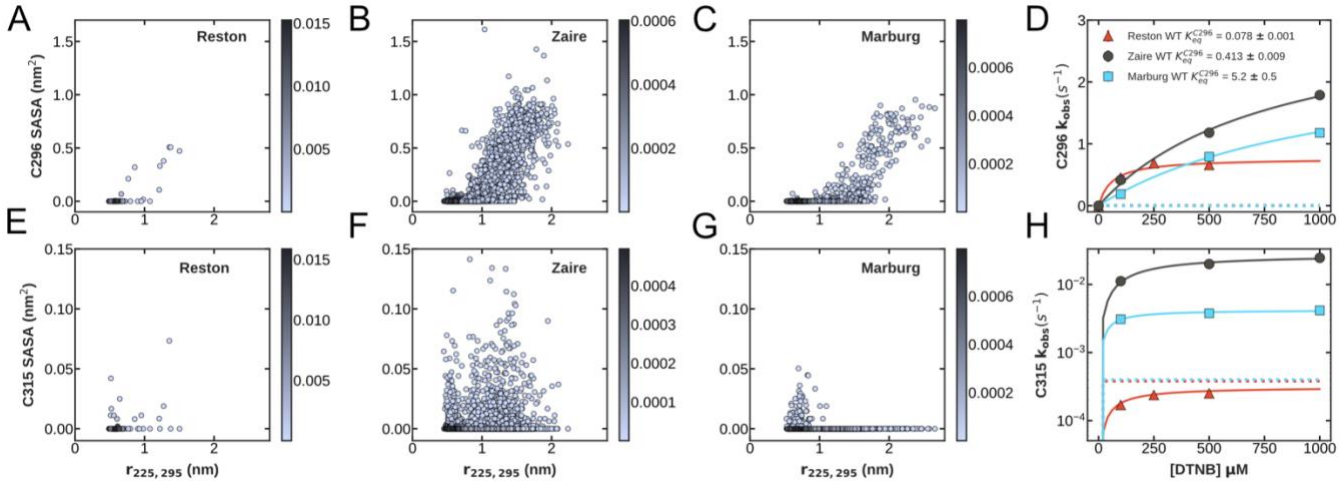

Figure 3: Thiol labeling of C296 confirms that the cryptic pocket in Marburg has the highest probability of being open while Reston has the lowest probability of opening.
(A-C) Plots of the distance between residues 225 and 295 vs. the Solvent Accessible Surface Area (SASA) of (A) C296 of Reston IID, (B) C307 of Zaire IID, (C) C296 of Marburg IID. Each point on the plot represents an MSM center and is colored according to its equilibrium probability.
(D) Observed thiol labeling rates for C296/C307 of Zaire IID (black circles), Marburg IID (blue squares), and Reston IID (red triangles) at a range of DTNB concentrations.
(E-G) Plots of the distance between residues 225 and 295 vs. the SASA of (E) C315 of Reston IID, (F) C326 of Zaire IID, and (G) C315 of Marburg IID calculated from our MSMs.

The SASA of C296 is more correlated with the opening of the cryptic pocket than the SASA of C315, so we focus on thiol labeling of C296 to experimentally characterize pocket opening.
(H) Observed thiol labeling rates for C315/C326 of Zaire IID (black circles), Marburg IID (blue squares), and Reston IID (red triangles) at a range of DTNB concentrations. Fits to the Linderstrøm–Lang model are shown in colored lines and the expected labeling rate from the unfolded state is shown as blue dotted lines for Marburg IID and red dotted lines for Reston IID. This rate is estimated from the stability and unfolding rate measured for these homologs shown in Fig S6. The mean and standard deviation from three measurements are shown for D and H. Error bars are smaller than the marker used.

Binding to different length RNAs suggests closed conformations preferentially bind dsRNA blunt ends while open conformations prefer binding the backbone

To test our prediction that more open homologs prefer the backbone while more closed ones preferentially bind the blunt ends, we used a fluorescence polarization assay to quantify the affinity of the homologs to different length dsRNA substrates, thereby varying the number of backbone binding sites available. Briefly, we use dsRNA labeled with fluorescein on one 5' end and titrate in varying concentrations of the proteins into a fixed concentration of the RNA (100 nM). Free RNA emits depolarized light upon excitation with polarized light due to its fast rotation whereas bound RNA emits polarized light. Recording the fluorescence polarization throughout the titration gives us the fraction of RNA bound to the protein. Particularly, we used a short 8bp RNA to get an accurate measure of binding to the blunt-ends and the same 8bp RNA with a 2 nucleotide overhang on the 3' ends to get an accurate measure of the binding to the backbone alone. Structural studies on these homologs suggest that the backbone binding mode has a footprint of 3 nucleotides (19, 25, 27). Previous RNA binding studies also suggest multiple IID molecules bind to each RNA molecule in solution (19, 26). Therefore, we also performed measurements with a longer 25bp RNA with and without a 2 nucleotide overhang on the 3' ends to get an accurate measure of any cooperativity in the backbone binding. To accurately account for multiple protein molecules binding to a single molecule of dsRNA, we used a one-dimensional lattice binding model to fit the experimental data (see Materials and Methods). Based on available structural information, the model assumes that each protein molecule binding to the end occludes 1 nucleotide and that each backbone binding protein occludes 3 nucleotides. It allows nearest neighbor cooperativity between all binding modes. Globally fitting these four binding curves for each homolog to this model (Fig 4A, B and C) generates estimates for the equilibrium constant for opening the pocket (Fig S7A), the dissociation constants of the closed and open conformations to each site on the backbone ($K_{D,backbone}^{closed}$ and $K_{D,backbone}^{open}$ respectively), the dissociation constant of closed conformations to the blunt-end of the RNA ($K_{D,end}^{closed}$), and the cooperativity between these interactions (Fig 4D and Fig S7B).

In support of our proposed model, we find that the equilibrium constant for pocket opening (K_{oc} , where oc stands for open vs closed) from our fits to dsRNA-binding data are in good agreement with those from our thiol labeling experiments (K_{eq}^{C296}). If the pocket opening observed in the thiol-labeling assays is the same conformational change affecting the RNA binding, we would expect the K_{oc} obtained from the fits to be comparable to K_{eq}^{C296}. Indeed, K_{oc} for all three homologs obtained from the fits to the binding model agree very well with the equilibrium constant for the exposure of C296 (C307 in Zaire IID) obtained from the thiol labeling assay (Fig S7A).

In support of our hypothesis, our global fits show that the probability of pocket opening is the main difference between the variants. As describe above, the equilibrium constants for pocket opening vary over a range of about 60-fold between the three homologs. In contrast, each of the dissociation constants between either open or closed protein and the blunt ends or backbone of dsRNA varies by no more than a factor of about 6-fold between the homologs. We also do not observe large differences in the magnitude of the cooperativity parameters between the homologs (Fig S7B). The interaction between the closed state and the blunt ends of dsRNA is far stronger than the interaction of either the closed state or the open state for the backbone (Fig 4D). There is also negative cooperativity between two closed structures binding the backbone at adjacent positions (Fig S7B). As a result, homologs with low probabilities of pocket opening predominantly bind the blunt ends and their binding curves are left-shifted compared to more open homologs. Interestingly, even in this case, most of the proteins that do bind the backbone are in the open state (Fig 4E and F). For example, about 60% of the backbone is covered by Reston proteins in the open state even though the equilibrium constant for pocket opening (K_{oc}) is 0.067 $\pm$ 0.004, which corresponds to a probability of cryptic pocket opening of 0.063, the lowest of the three homologs. Homologs with a higher probability of pocket opening are less likely to bind the blunt ends since the open state cannot bind blunt ends in our model. Interestingly, the interaction between the open state and the backbone ($K_{D,backbone}^{open}$) is about two orders of magnitude stronger than that of the closed state ($K_{D,backbone}^{closed}$) for all three homologs (Fig 4D). There is also strong positive cooperativity between a protein in the open state binding the backbone alongside either a closed or open state (Fig S7B). As a result, homologs with high probabilities of pocket opening, like Marburg, predominantly bind the backbone and have right-shifted binding curves compared to more closed variants.

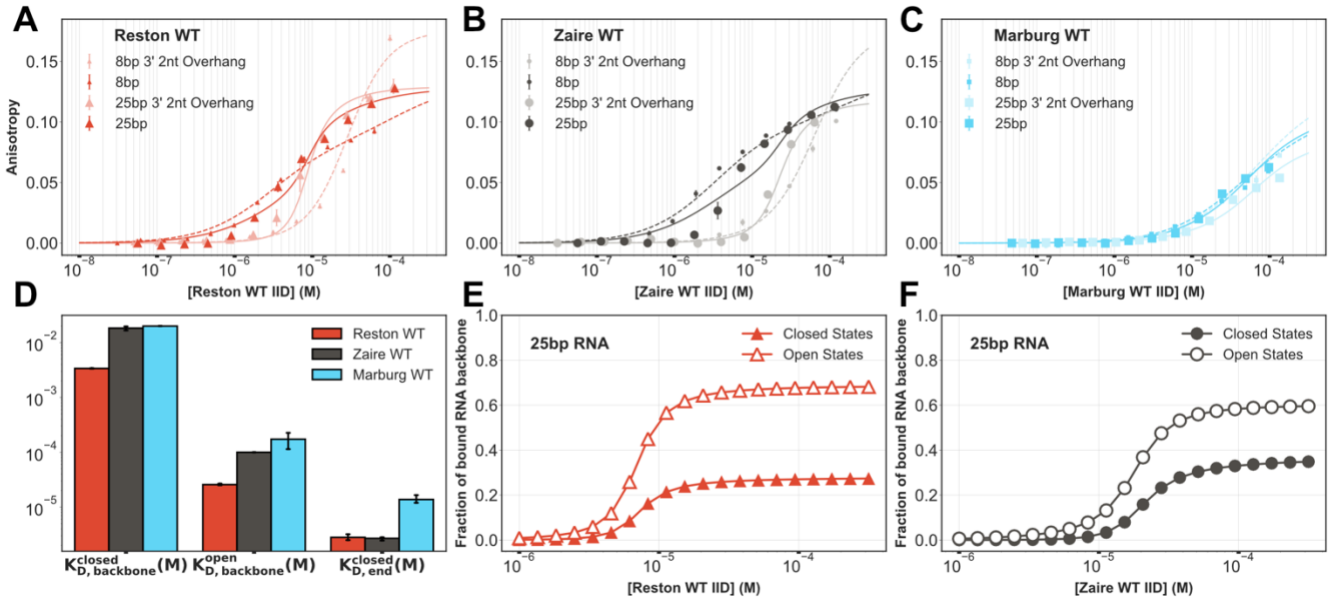

Figure 4: Binding to different length dsRNAs suggests closed conformations preferentially bind dsRNA blunt ends while open conformations prefer binding the backbone. (A-C) Binding to fluorescently labeled 8bp (dashed lines) and 25bp RNA (solid lines) with and without a 3' 2 nucleotide overhang of (A) Wild Type (WT) Reston IID (B) WT Zaire IID (C) WT Marburg IID.
The anisotropy was measured via a fluorescence polarization assay, converted to anisotropy, and fit to a one-dimensional lattice binding model. The mean and standard deviation from three replicates is shown but error bars are generally smaller than the symbols.

(D) Comparison of binding affinities obtained from the global fits. The mean and standard deviation from fits to each of the three replicates are shown.
(E-F) Fraction of the backbone of 25 bp RNA covered by the open states (empty markers) and closed states (colored markers) calculated from the binding parameters obtained from the fits for (E) Reston WT IID and (F) Zaire WT IID.

Point mutations that alter the probability of pocket opening also induce differential binding to blunt ends versus the backbone

As a further test of our model, we next sought to identify point mutations that modulate the probability of pocket opening and assess if they have the expected impact on backbone versus blunt end binding. We started by comparing the sequences of the homologs. Zaire and Reston IIDs are particularly good candidates for this as there are only 14 amino acid substitutions between these homologs and they are all distal to the RNA-binding interfaces. Of these substitutions, one causes an interesting structural difference between Zaire and Reston IIDs (25). Residue P280 in Reston IID results in a formation of an alpha helix where an alanine in the same structural position (A291) in Zaire IID results in a disordered loop. This disordered loop forms a hinge that rotates as the pocket opens, and we reasoned this motion may be inhibited by the structure the proline induces in Reston. Therefore, we hypothesized that mutating the alanine in Zaire IID to a proline would result in a reduced probability of opening the pocket and a stronger preference for binding dsRNA blunt ends. This hypothesis is supported by our past work showing that the substitution A291P in Zaire IID reduces the probability of pocket opening(24). Furthermore, we propose that mutating the proline in Reston to an alanine should increase the probability of pocket opening and enhance the binding to the dsRNA backbone.

As expected, introducing A291P into the Zaire IID leads to a reduced probability of pocket opening in both simulations and thiol labeling experiments, while introducing P280A into the Reston IID leads to an increased probability of pocket opening. FAST simulations performed on these variants indeed showed an increased probability of opening the pocket in Reston IID P280A and a decreased probability in Zaire IID A291P (Fig 5A). The pocket also opens more widely in Reston IID P280A than Wild Type (WT) Reston IID. As before, examining the solvent exposure of the cysteines observed in simulation shows that while exposure of C296 (C307 in Zaire) is correlated with pocket opening, that of other cysteines is not Fig S8. We followed this up with the thiol labeling assay and assigned labeling rates to cysteines by labeling point mutations of individual cysteines to serines Fig S9. We observe that both cysteines in Reston IID P280A labeled faster than WT Reston IID (Fig 5B and S10A). In comparison, both cysteines in Zaire IID A291P labeled slower compared to WT Zaire IID Fig 5B and S10B. Fits to the Linderstrøm-Lang model show a 26-fold increase the in the equilibrium constant for C296 exposure in Reston IID P280A compared to Reston IID WT and a 36-fold decrease in Zaire IID A291P compared to Zaire IID WT (Fig 5B). We observed only moderate differences in the stability of Reston IID WT and Reston IID P280A under urea denaturation (Fig S6). Taken together, this indicates that the mutation modulates pocket opening without significantly impacting the stability of the protein.

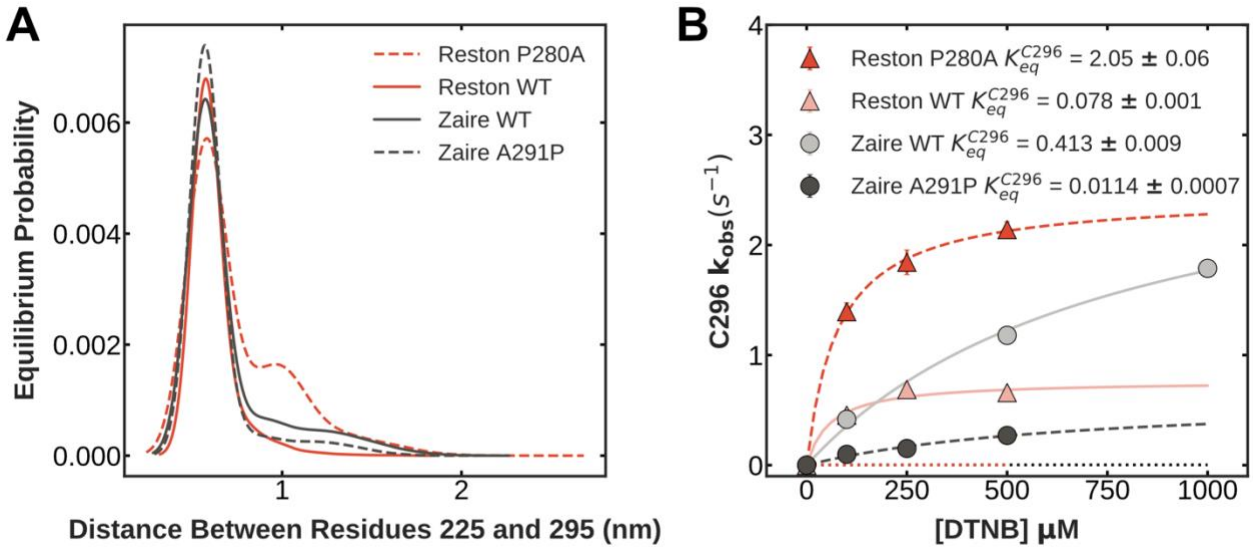

Figure 5: Single amino acid substitutions at residue 280 (291 in Zaire) modulate the probability of pocket opening. A) Probability distribution of the distance between residues 225 and 295 obtained from MSMs built from FAST adaptive sampling simulations of Reston IID WT (solid red), Reston IID P280A (dashed red), Zaire IID WT (solid black), and Zaire IID A291P (dashed black). B) Observed labeling rates of C296 for Reston IID WT (transparent solid red) and Reston IID P280A (dark dashed red) Zaire IID WT (transparent solid black), and Zaire IID A291P (dashed black).

In further support of our model, the point mutation that increases Reston's probability of pocket opening also shifts the protein towards backbone binding while the mutation that reduces the probability of pocket opening in Zaire shifts the balance towards blunt end binding. We performed fluorescence polarization binding experiments to the four RNA substrates we used for the WT homologs (Fig 6A-D). We analyzed these experiments using global fits of the binding of each mutant to all four RNAs to the one-dimensional lattice model described for the WT homologs (Fig 6A-D and Fig S11). Importantly, the equilibrium constants for pocket opening (K_{oc}) obtained from the fits to the binding model continue to agree very well with the equilibrium constants for the exposure of C296 obtained from the thiol labeling assay (Fig 6E). While there are large differences between the equilibrium constant for pocket opening amongst these variants, the rest of the fit parameters are very similar to those from the WT proteins. We observe that the dissociation constant of the closed conformations to each site on the backbone ($K_{D,backbone}^{closed}$) is two orders of magnitude weaker than that of the open conformations ($K_{D,backbone}^{open}$) for the mutants as well (Fig 6F), further increasing our confidence in the model. We do not observe large differences in the magnitude of the binding and cooperativity parameters between the WTs and the mutants except for the cooperativity between two neighboring molecules in the open state binding to the backbone σ_{oo} Fig (6F, G and H). This parameter seems to buffer the effect of the increased affinity of the open state to the backbone. This inherent trade-off also suggests that shifting the equilibrium too drastically in favor of the open conformation would not be beneficial for binding to the backbone. Particularly, Reston P280A is more open than WT Reston, but its binding curves for the overhang substrates are right-shifted compared to WT. This suggests that though the open pocket conformations bind better to the backbone, increasing the fraction of open conformation in solution has a negative effect on the overall backbone binding. To understand this, we calculated the fraction of the RNA backbone covered by the open and closed states of the various homologs and mutants for different

lengths of RNA based on the binding parameters obtained from the fit (Fig S12). We see that in all cases, a much larger proportion of the backbone is covered by the open states than by the closed states. This suggests that while it is necessary to have some ability to open the pocket, increasing the probability of opening has no added benefit for backbone binding. Meanwhile, Zaire A291P is more closed and prefers to bind the blunt ends more than the WT protein. This is reflected by its binding curve for the blunt ended substrates being left-shifted compared to WT. Overall, these results show a strong correspondence between the probability of pocket opening and the relative affinities of the VP35 IID for different binding sites on dsRNA.

Conclusion and Discussion

We have shown that the opening and closing of a cryptic pocket in VP35 toggles the protein between two different functions. Specifically, we have shown that VP35 preferentially binds the blunt ends of dsRNA when the cryptic pocket is closed but preferentially binds the backbone when the pocket is open. As a result, variants where the pocket tends to be closed are better at blocking RIG-I from binding dsRNA, whereas VP35 variants that are open more often are better at blocking MDA5. This model explains differences in the RNA binding behaviors of the Zaire, Reston, and Marburg variants of VP35 that were difficult to explain based on crystal structures of the proteins. Moreover, it predicts how point mutations that alter the probability of cryptic pocket opening will alter the relative affinity of a VP35 protein for the backbone versus the blunt ends of dsRNA. More importantly, cryptic pockets are not just happenstance, as they can have a functional role. This finding suggests that cryptic pockets are under selective pressure. As a result, it shouldn't be trivial for viruses or cells to evolve resistance to drugs that target cryptic pockets by acquiring mutations that prevent pocket opening. Future studies on other proteins will help to establish if these findings are generally true or are limited to a subset of cryptic pockets. In the meantime, efforts to target the cryptic pocket in VP35 are warranted given the evidence that this pocket has a functional role that should prevent Zaire ebolavirus from acquiring mutations that give rise to drug resistance by preventing the pocket from opening

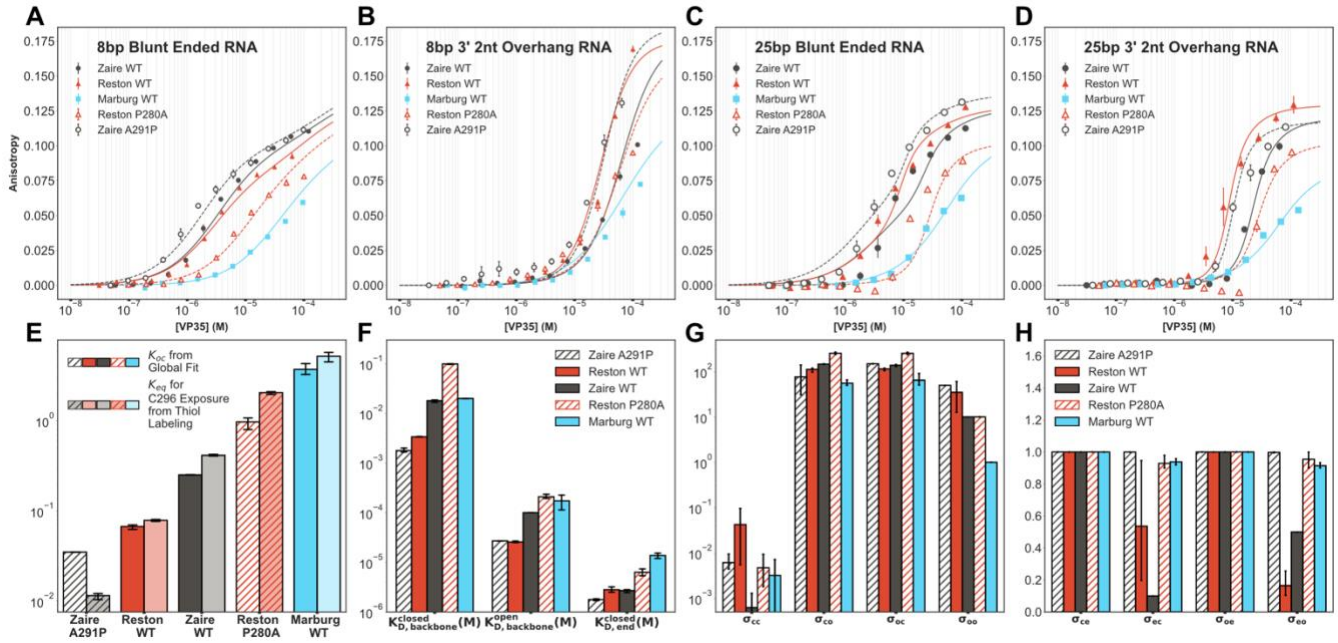

Figure 6: Single amino acid substitutions that alter the probability of pocket opening also induce differential binding to blunt ends and the backbone. (A-D) Binding of Zaire A291P IID (empty black circles), WT IID (solid red triangles), WT Zaire IID (solid black circles), Reston P280A IID (empty red triangles), WT Marburg IID (blue squares) to fluorescently labeled (A) 8bp blunt-ended RNA (B) 8bp RNA with two nucleotide overhangs on 3' ends (C) 25bp blunt-ended RNA (D) 25bp RNA with two nucleotide overhangs on 3' ends.
The anisotropy was measured via a fluorescence polarization assay, converted to anisotropy, fit to a one-dimensional lattice binding model. The mean and standard deviation from three replicates is shown but error bars are generally smaller than the symbols. Lines indicate the global fits.
(E) Comparison of K_{oc} obtained from the global fits to K_{eq} for C296 exposure obtained from DTNB labeling experiments shown in Fig 3D.
(F) Comparison of dissociation constants obtained from the global fits.
(G) Comparison of cooperativity between backbone binding modes obtained from the global fits. (H) Comparison of cooperativity between end and backbone binding modes obtained from the global fits.
The mean values of fits to three individual replicates are shown for E, F, G and H. Standard deviations are shown as error bars.

Materials and Methods

Molecular dynamics simulations and analysis

Simulations for Reston IID and Marburg IID were initiated from the apoprotein models of PDB 3L2A (Reston IID(25)) and 4GHL (Marburg IID(27)) and run with Gromacs(32) using the amber03 force field(33) and TIP3P explicit solvent(34) at a temperature of 300 K and 1 bar pressure, as described previously(35). We first applied our FAST-pockets algorithm(28) to replicate the simulation protocol of Zaire IID in (24). As done previously, we performed ten rounds of FAST simulations with 10 simulations/round and 80 ns/simulation. We then performed an RMSD-based clustering using a hybrid k-centers/k-medoids algorithm(36) implemented in

Enspara(37) to divide the data into 1000 clusters. Then we ran three simulations initiated from each cluster center on the Folding@home distributed computing environment, resulting in an aggregate simulation time of 126 μs for Reston IID and 124 μs for Marburg IID. We calculated the Solvent Accessible Surface Area (SASA) of all atoms in each frame using a 2.8Å probe. We summed up the SASA of all the atoms of the sidechain of each residue and clustered this sidechain SASA using kcenters upto a cluster radius of 3.5 nm^2 for Reston IID and 2.7 nm^2 for Marburg IID followed by five rounds of kmedoids. We then built Markov State Models (MSMs) at multiple lagtimes and used the implied timescales test (Fig S2) to decide on a 6ns lagtime for the final MSMs of all homologs.

Protein expression and purification

All variants of VP35's IID were purified from the cytoplasm of *E. coli* BL21(DE3) Gold cells (Agilent Technologies) following the protocols detailed in (25, 27, 38) . Variants were generated using the site-directed mutagenesis and confirmed by DNA sequencing. Cells were transformed using heat shock at 42°C. Transformed cells were grown in LB media (Fisher Scientific or Lambda Biotech, Ballwin, MO) at 37 °C until OD 0.3 then grown at 18 °C until induction at OD 0.6 with 1 mM IPTG (Gold Biotechnology, Olivette, MO). Cells were grown for 15 h then centrifuged. The pellet was resuspended in 20 mM sodium phosphate pH 8, 1 M sodium chloride, 5 mM Imidazole with 5.1 mM *β*-mercaptoethanol. Resuspended cells were subjected to sonication at 4 °C followed by centrifugation. The supernatant was then subjected to Ni-NTA affinity (BioRad Bio-Scale Mini Nuvia IMAC column) and eluted with 20 mM sodium phosphate pH 8, 1 M sodium chloride, 250 mM Imidazole with 5.1 mM *β*-mercaptoethanol. This was dialyzed to a final buffer of 20 mM sodium phosphate pH 8, 50mM NaCl with 5.1 mM *β*-mercaptoethanol. The dialyzed sample was subjected to TEV digestion (proTEV plus, Promega) at room temperature for 24-48 hours. We followed this by cation exchange (BioRad UNOsphere Rapid S column) and eluted using a slow gradient of 20 mM sodium phosphate pH 8, 1 M NaCl with 5.1 mM *β*-mercaptoethanol. Cleaved VP35 with the sequence shown in Fig S1 elutes at around 180 mM NaCl. We follow this by size-exclusion chromatography (BioRad Enrich SEC 70 column or Cytiva HiLoad 16/600 Superdex 75) into 10 mM HEPES pH 7, 150 mM NaCl, 1 mM $MgCl_2$, 2 mM TCEP.

Thiol Labeling

We monitored the change in absorbance over time of 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB, Ellman's reagent, Thermo Fisher Scientific). Various concentrations (100 -1000 μM) of DTNB were added to the 5 μM protein and change in absorbance was measured in an SX-20 Stopped Flow instrument (Applied Photophysics, Leatherhead, UK) at 412 nm until the reaction reached a steady state (300-1200 s). Each time course was fit with as many cysteines being labeled for that protein to obtain an observed labeling rate (k_{obs}) for each cysteine at each DTNB concentration. The k_{obs} as a function of DTNB concentration for each cysteine were fit with a Linderstrøm–Lang model, shown below, to extract the thermodynamics and/or kinetics of pocket opening, as described in detail previously (31).

$$Closed \underset{k_{close}}{\overset{k_{open}}{\rightleftharpoons}} Open \xrightarrow{k_{int}[DTNB]} Labeled \qquad (1)$$

Which leads to,

$$k_{obs} = \frac{k_{open} * k_{int}[DTNB]}{k_{close} + k_{open} + k_{int}[DTNB]} \qquad (2)$$

From this fit, we report the equilibrium constant for cysteine exposure calculated as:

$$K_{eq} = \frac{k_{open}}{k_{closed}} \tag{3}$$

As a control, the equilibrium constant for folding and the unfolding rate were measured (Fig S6) and used to predict the expected labeling rate from the unfolded state. The equilibrium constant was inferred from a two-state fit to urea melts monitored by fluorescence and unfolding rates were inferred from exponential fits to unfolding curves monitored by fluorescence after the addition of urea, as described previously (31, 39, 40). Fluorescence data were collected using a Jasco FP-8300 Spectrofluorometer with Jasco ETC-815 Peltier and Koolance Exos2 Liquid Coolant-controlled cuvette holder.

Fluorescence Polarization Assay

Binding affinities between homologs and variants of VP35's IID and dsRNA were measured using fluorescence polarization in 10 mM HEPES pH 7, 150 mM NaCl, 1 mM $MgCl_2$. 8bp and 25bp FITC-dsRNA (Integrated DNA Technologies) substrates with or without a 2 nucleotide 3' overhang of the following sequences was included at 100 nM.

RNA	Sequence
8bp Blunt Ended	5'-/56-FAM/CGCAUGCG-3' 5'-CGCAUGCG-3'
8bp 2nt 3' Overhang	5'-/56-FAM/CGCAUGCGCU-3' 5'-CGCAUGCGCU-3'
25bp Blunt Ended	5'-/56-FAM/AAACUGAAAGGGAGAAGUGAAAGUG-3' 5'-CACUUUCACUUCUCCCUUUCAGUUU-3'
25bp 2nt 3' Overhang	5'-/56-FAM/AAACUGAAAGGGAGAAGUGAAAGUGCU-3' 5'-CACUUUCACUUCUCCCUUUCAGUUUCU-3'

The sample was equilibrated for one hour before data collection. Data were collected on a BioTek Synergy2 Multi-Mode Reader as polarization and were converted to anisotropy (r) as in (41):

$$r = \frac{2p}{3-p}$$

Where p in the polarization calculated from the observed parallel and perpendicular intensities, $I_{\parallel}$ and $I_{\perp}$ as:

$$p = \frac{I_{\parallel} - \mathrm{G} I_{\perp}}{I_{\parallel} + \mathrm{G} I_{\perp}}.$$

The total fluorescence intensities were checked at each well for fluorescence anomalies such as quenching.

Analysis of dsRNA binding experiments.

To obtain the contributions from the end-binding and backbone-binding to the increase in anisotropy observed in the fluorescence polarization experiments, we used a one-dimensional lattice model for proteins competing to bind to nucleic acids. Here we consider our labeled dsRNA to be the macromolecule and the IIDs to be the ligands. Specifically, our model has the

open and the closed conformations of the IIDs competing for binding to the dsRNA. The total concentration of the ligand is known for every point in the titration and relates to the relative concentration of the open and the closed states as:

$$[IID]_{total} = (1 + K_{oc})[IID]_{closed} \tag{4}$$

$$[IID]_{open} = K_{oc}[IID]_{closed} \tag{5}$$

To treat the simplest case first, we assume that only the closed state binds to the ends of the dsRNA and that such binding occludes one nucleotide. We infer this from the structural data available for Reston and Zaire IIDs. We refer to the intrinsic association constant for this interaction as K_A^{end}.

We assume that the open and closed states bind to the dsRNA backbone with intrinsic association constants K_A^{open} and K_A^{closed}. We assume that these interactions occlude three nucleotides each in the backbone of the dsRNA.

To further test these assumptions of binding site sizes, we repeated the fitting with models with backbone binding site sizes ranging from 3-5 nucleotides (Fig S13-16). We note that the major conclusions of the study hold true for varying backbone binding site sizes. This indicates that the fluorescence polarization assay does not have the sensitivity to capture the binding site size effectively. Future RNA binding studies with a technique with this sensitivity will help in confidently estimating these parameters.

The dissociation constants referred to in the text are inverses of the association constants described above. Therefore,

$$K_{D,backbone}^{closed} = \frac{1}{K_A^{closed}} \tag{6}$$

$$K_{D,backbone}^{open} = \frac{1}{K_A^{open}} \tag{7}$$

$$K_{D,end}^{closed} = \frac{1}{K_A^{end}} \tag{8}$$

We treat the backbone binding of the closed state, the backbone binding of the open state, and the end-binding of the closed state as three ligand species competing for binding the dsRNA. We denote the ligand type by $s = 1,2\ or\ 3$ for each of those cases respectively.

We use the transfer matrix method detailed in (42) to calculate the probability, $p_s(i)$, that a base-pair i is occupied by a ligand of type s given by:

$$p_s(i) = \frac{Z_s(i)}{Z} \tag{9}$$

Where Z is the partition function and $Z_s(i)$ is a sum over all allowed Boltzmann-weighted states consistent with base pair i, for $i \in (1, N)$ where N is the length of the dsRNA, being covered by a type s ligand. $Z_s(i)$ and Z are calculated using transfer matrices.

To construct the transfer matrix for each nucleotide i, we enumerate all the possible states a given nucleotide i can be in and assign each state a statistical weight according to the method in (42). We need to make two major changes to this model to describe our system.

The first is to account for the fact that the bulk concentrations of the ligand species in our system are not independent. In the model, the statistical weight for the binding of ligand of type s is given by $c_s K_s$ where c_s is the bulk concentration of the ligand and K_s is its intrinsic association constant. In our system, the bulk concentrations of the ligands are the concentration of the open and closed conformations of the protein. These concentrations are related as shown in equations (4) and (5). Therefore, the statistical weights for each of the interactions in our system are given by K_{eff}^{s} where,

$$K_{eff}^{1} = \frac{K_A^{closed}[IID]_{total}}{(1 + Koc)} \tag{10}$$

$$K_{eff}^{2} = \frac{Koc\ K_A^{open}[IID]_{total}}{(1 + Koc)} \tag{11}$$

$$K_{eff}^{3} = \frac{(K_A^{end} - K_A^{closed})[IID]_{total}}{(1 + Koc)} \tag{12}$$

The second change we need to make is to account for the fact that end-binding can only occur at nucleotides $i = 1\ or\ N$. Therefore, the value of K_{eff}^{3} given by equation (12) is only included in transfer matrices for nucleotides $1\ or\ N$ and we set $K_{eff}^{3} = 0$ for all other nucleotides.

Further, we allow cooperativity between all binding modes denoted by the constants $\sigma_{ss'}$ where s and s' can be $1,2\ or\ 3$ denoting the open state binding to the backbone, closed state binding to the backbone and the closed state binding to the ends respectively. For example, the cooperativity between one closed conformation binding to the backbone and an open conformation binding to the backbone to its right is given by σ_{12}. For ease of understanding we refer to ligand types $s = 1,2 \text{ and } 3$ as o, c and e in the text. Therefore, the cooperativity between one closed conformation binding to the backbone and an open conformation binding to the backbone to its right is given by σ_{co}. Examples of these statistical weights are shown in Fig S17.

With these statistical weights, we construct the transfer matrices for each nucleotide i as below, with $K_{eff}^{3} = 0$ for all nucleotides other than 1 and N.

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & \sigma_{11} & 0 & 0 & \sigma_{12} & \sigma_{13} \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & K_{eff}^{1} & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & \sigma_{21} & 0 & 0 & \sigma_{22} & \sigma_{23} \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & K_{eff}^{2} & 0 & 0 \\ K_{eff}^{3} & 0 & 0 & \sigma_{31} & 0 & 0 & \sigma_{32} & 0 \end{pmatrix}$$

We use these to calculate the probability, $p_s(i)$, that a base-pair i is occupied by a ligand of type s given by equation (9).

From this, we calculate the average probability of the backbone being bound as

$$p_b = \frac{\sum_{i=1}^{N}\left(p_1(i) + p_2(i)\right)}{N} \tag{13}$$

and the average probability of the ends being bound as

$$p_{end} = \frac{p_3(1) + p_3(N)}{2} \tag{14}$$

We then convert this average probability to observed anisotropy (r_{obs}) as follows:

$$r_{obs} = \left(r_0 - r_{max}^{end}\right)p_{end} + \left(r_0 - r_{max}^{b}\right)p_b \tag{15}$$

Where, r_o is the anisotropy of the free RNA.
r_{max}^{end} is the maximum anisotropy for blunt-end binding,
r_{max}^{b} is the maximum anisotropy for backbone binding,

Here, r_{max}^{end} and r_{max}^{b} are parameters obtained from the fit and are shown Fig S18.

Fits were performed using Scipy 1.8.0, NumPy 1.22.2 and lmfit 1.2.2.

Acknowledgments

We are grateful to the citizen scientists who participate in Folding@home for volunteering to run simulations on their personal computers. We thank Dr. Gaya Amarasinghe for providing the plasmids for Reston and Marburg IIDs. This work was supported by NIH NIGMS R35GM152085 and NSF MCB 2218156. M.A.C. was supported by the NIH grants 5R25GM103757 to WUSTL IMSD program, and NIH F31AI157079.

References

1. G. R. Bowman, P. L. Geissler, Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 11681–11686 (2012).

2. C. R. Knoverek, G. K. Amarasinghe, G. R. Bowman, Advanced Methods for Accessing Protein Shape-Shifting Present New Therapeutic Opportunities. *Trends in Biochemical Sciences* **44**, 351–364 (2019).

3. J. R. Horn, B. K. Shoichet, Allosteric Inhibition Through Core Disruption. *Journal of Molecular Biology* **336**, 1283–1291 (2004).

4. C. J. Wenthur, P. R. Gentry, T. P. Mathews, C. W. Lindsley, Drugs for Allosteric Sites on Receptors. *Annu. Rev. Pharmacol. Toxicol.* **54**, 165–184 (2014).

5. E. Guarnera, I. N. Berezovsky, Allosteric drugs and mutations: chances, challenges, and necessity. *Current Opinion in Structural Biology* **62**, 149–157 (2020).

6. A. Meller, *et al.*, Drug specificity and affinity are encoded in the probability of cryptic pocket opening in myosin motor domains. *eLife* **12**, e83602 (2023).

7. C. M. Chio, C. S. Lim, A. C. Bishop, Targeting a Cryptic Allosteric Site for Selective Inhibition of the Oncogenic Protein Tyrosine Phosphatase Shp2. *Biochemistry* **54**, 497–504 (2015).

8. K. M. Hart, *et al.*, Designing small molecules to target cryptic pockets yields both positive and negative allosteric modulators. *PLoS ONE* **12**, e0178678 (2017).

9. A. Meller, D. Kelly, L. G. Smith, G. R. Bowman, Toward physics-based precision medicine: Exploiting protein dynamics to design new therapeutics and interpret variants. *Protein Science* **33**, e4902 (2024).

10. J. S. Allingham, R. Smith, I. Rayment, The structural basis of blebbistatin inhibition and specificity for myosin II. *Nat Struct Mol Biol* **12**, 378–379 (2005).

11. P. Cimermancic, *et al.*, CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites. *Journal of Molecular Biology* **428**, 709–719 (2016).

12. S. Vajda, D. Beglov, A. E. Wakefield, M. Egbert, A. Whitty, Cryptic binding sites on proteins: definition, detection, and druggability. *Current Opinion in Chemical Biology* **44**, 1–8 (2018).

13. C. F. Basler, *et al.*, The Ebola Virus VP35 Protein Inhibits Activation of Interferon Regulatory Factor 3. *J Virol* **77**, 7945–7956 (2003).

14. A. L. Hartman, J. S. Towner, S. T. Nichol, A C-terminal basic amino acid motif of Zaire ebolavirus VP35 is essential for type I interferon antagonism and displays high identity with the RNA-binding domain of another interferon antagonist, the NS1 protein of influenza A virus. *Virology* **328**, 177–184 (2004).

15. W. B. Cárdenas, *et al.*, Ebola Virus VP35 Protein Binds Double-Stranded RNA and Inhibits Alpha/Beta Interferon Production Induced by RIG-I Signaling. *J Virol* **80**, 5168–5178 (2006).

16. D. W. Leung, C. F. Basler, G. K. Amarasinghe, Molecular mechanisms of viral inhibitors of RIG-I-like receptors. *Trends in Microbiology* **20**, 139–146 (2012).

17. K. A. Dilley, *et al.*, The Ebola virus VP35 protein binds viral immunostimulatory and host RNAs identified through deep sequencing. *PLoS ONE* **12**, e0178717 (2017).

18. C. Woolsey, *et al.*, A VP35 Mutant Ebola Virus Lacks Virulence but Can Elicit Protective Immunity to Wild-Type Virus Challenge. *Cell Rep* **28**, 3032-3046.e6 (2019).

19. D. W. Leung, *et al.*, Structural basis for dsRNA recognition and interferon antagonism by Ebola VP35. *Nat Struct Mol Biol* **17**, 165–172 (2010).

20. G. L. Daino, *et al.*, Identification of Myricetin as an Ebola Virus VP35–Double-Stranded RNA Interaction Inhibitor through a Novel Fluorescence-Based Assay. *Biochemistry* **57**, 6367–6378 (2018).

21. J. G. Glanzer, *et al.*, In silico and in vitro methods to identify ebola virus VP35-dsRNA inhibitors. *Bioorganic & Medicinal Chemistry* **24**, 5388–5392 (2016).

22. C. S. Brown, *et al.*, In Silico Derived Small Molecules Bind the Filovirus VP35 Protein and Inhibit Its Polymerase Cofactor Activity. *Journal of Molecular Biology* **426**, 2045–2058 (2014).

23. A. L. Hopkins, C. R. Groom, The druggable genome. *Nature Reviews Drug Discovery* **1**, 727–730 (2002).

24. M. A. Cruz, *et al.*, A cryptic pocket in Ebola VP35 allosterically controls RNA binding. *Nature Communications* **13**, 2269 (2022).

25. D. W. Leung, *et al.*, Structural and Functional Characterization of Reston Ebola Virus VP35 Interferon Inhibitory Domain. *Journal of Molecular Biology* **399**, 347–357 (2010).

26. M. R. Edwards, *et al.*, Differential Regulation of Interferon Responses by Ebola and Marburg Virus VP35 Proteins. *Cell Reports* **14**, 1632–1640 (2016).

27. P. Ramanan, *et al.*, Structural basis for Marburg virus VP35–mediated immune evasion mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 20661–20666 (2012).

28. M. I. Zimmerman, G. R. Bowman, FAST Conformational Searches by Balancing Exploration/Exploitation Trade-Offs. *J. Chem. Theory Comput.* **11**, 5747–5757 (2015).

29. V. A. Voelz, V. S. Pande, G. R. Bowman, Folding@home: Achievements from over 20 years of citizen science herald the exascale era. *Biophysical Journal* **122**, 2852–2863 (2023).

30. C. R. Knoverek, *et al.*, Opening of a cryptic pocket in β-lactamase increases penicillinase activity. *Proceedings of the National Academy of Sciences* **118**, e2106473118 (2021).

31. J. R. Porter, *et al.*, Cooperative Changes in Solvent Exposure Identify Cryptic Pockets, Switches, and Allosteric Coupling. *Biophysical Journal* **116**, 818–830 (2019).

32. D. Van Der Spoel, *et al.*, GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **26**, 1701–1718 (2005).

33. Y. Duan, *et al.*, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry* **24**, 1999–2012 (2003).

34. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926–935 (1983).

35. K. M. Hart, C. M. W. Ho, S. Dutta, M. L. Gross, G. R. Bowman, Modelling proteins' hidden conformations to predict antibiotic resistance. *Nature Communications* **7**, 12965 (2016).

36. K. A. Beauchamp, *et al.*, MSMBuilder2: Modeling Conformational Dynamics on the Picosecond to Millisecond Scale. *J. Chem. Theory Comput.* **7**, 3412–3419 (2011).

37. J. R. Porter, M. I. Zimmerman, G. R. Bowman, Enspara: Modeling molecular ensembles with scalable data structures and parallel computing. *The Journal of Chemical Physics* **150**, 044108 (2019).

38. D. W. Leung, *et al.*, Structure of the Ebola VP35 interferon inhibitory domain.

39. G. R. Bowman, E. R. Bolin, K. M. Hart, B. C. Maguire, S. Marqusee, Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. *Proceedings of the National Academy of Sciences* **112**, 2734–2739 (2015).

40. M. I. Zimmerman, *et al.*, Prediction of New Stabilizing Mutations Based on Mechanistic Insights from Markov State Models. *ACS Cent. Sci.* **3**, 1311–1321 (2017).

41. J. L. Keck, Ed., *Single-Stranded DNA Binding Proteins: Methods and Protocols* (Humana Press, 2012).

42. A. N. Nilsson, *et al.*, Competitive binding-based optical DNA mapping for fast identification of bacteria - multi-ligand transfer matrix theory and experimental applications on Escherichia coli. *Nucleic Acids Research* **42**, e118–e118 (2014).