bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discovery of a cryptic allosteric site in Ebola's 'undruggable' VP35 protein using simulations and experiments

Matthew A. Cruz[1, ‡], Thomas E. Frederick[1, ‡], Sukrit Singh[1], Neha Vithani[1], Maxwell I. Zimmerman[1], Justin R. Porter[1], Katelyn E. Moeder[1], Gaya K. Amarasinghe[2], and Gregory R. Bowman[1,3,*]

[1]Department of Biochemistry & Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110, USA

[2]Department of Pathology & Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA

[3]Center for the Science and Engineering of Living Systems, Washington University in St. Louis, St. Louis, MO 63110, USA

[‡]These authors contributed equally to this work

[*]To whom correspondence should be sent: g.bowman@wustl.edu

Abstract

Many proteins are classified as 'undruggable,' especially those that engage in protein-protein and protein-nucleic acid interactions. Discovering 'cryptic' pockets that are absent in available structures but open due to protein dynamics could provide new druggable sites. Here, we integrate atomically-detailed simulations and biophysical experiments to search for cryptic pockets in viral protein 35 (VP35) from the highly lethal Ebola virus. VP35 plays multiple essential roles in Ebola's replication cycle, including binding the viral RNA genome to block a host's innate immunity. However, VP35 has so far proved undruggable. Using adaptive sampling simulations and allosteric network detection algorithms, we uncover a cryptic pocket that is allosterically coupled to VP35's key RNA-binding interface. Experimental tests corroborate the predicted pocket and confirm that stabilizing the open form allosterically disrupts RNA binding. These results demonstrate simulations' power to characterize hidden conformations and dynamics, uncovering cryptic pockets and allostery that present new therapeutic opportunities.

Introduction

Many proteins have proved so difficult to target with small molecule drugs that they are often classified as undruggable, greatly limiting the scope of drug design efforts. In fact, up to 85% of human proteins have been classified as undruggable because their folds are thought to lack binding pockets where small molecules can bind with the affinity and specificity required for drug design.[1] Many undruggable proteins predominantly participate in protein-protein interactions (PPIs) and protein-nucleic acid interactions (PNIs).[2,3] In contrast to the binding pockets that many enzymes and receptors use to bind their small molecule ligands, the large flat interfaces involved in PPIs and PNIs do not lend themselves to forming many favorable interactions with small drug-like molecules. As a result, PPIs and PNIs are often considered intractable targets even when there is strong evidence that disrupting these interactions would be of great therapeutic value.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cryptic pockets could provide new opportunities to target undruggable proteins,[4,5] but realizing this potential remains challenging. Such pockets are absent in available experimental structures because they only form in a subset of excited states that arise due to protein dynamics. Cryptic sites can serve as valuable drug targets if they coincide with key functional sites, as can cryptic allosteric sites that are coupled to distant functional sites. However, identifying cryptic pockets remains difficult. Most known cryptic sites were only identified after the serendipitous discovery of a small molecule that binds and stabilizes the open form of the pocket.[5,6] Experimental techniques for intentionally identifying and targeting cryptic pockets show great promise,[7-9] but they still leverage the simultaneous discovery of ligands that bind and stabilize the open pocket. To overcome this limitation, a number of computational methods have been developed to identify cryptic pockets without requiring the simultaneous discovery of small molecules that bind them.[10-17] These methods have proved capable of retrodicting a number of previously identified cryptic pockets. More importantly, applications to a variety of established drug targets and other enzymes have successfully identified novel cryptic pockets that have been corroborated by subsequent experiments.[11,18,19] In light of these successes, it is interesting to further investigate whether such approaches can reveal cryptic pockets in proteins that are currently considered undruggable, particularly proteins that primarily engage in PPIs and PNIs.

Here, we integrate atomically-detailed computer simulations and biophysical experiments to search for cryptic pockets in a target that has so far proved undruggable: the interferon inhibitory domain (IID) of Ebola viral protein 35 (VP35). Ebola virus causes a hemorrhagic fever that is often lethal, with case fatality rates approaching 90% in past outbreaks.[20,21] While recent progress in vaccine development and use of biologics, such as antibodies, for therapeutic and prophylactic purposes show promise,[22] small molecule drugs still offer many advantages, including ease of delivery, lower cost, and longer shelf life. The ~120 residue IID of VP35 is a particularly appealing drug target for combating Ebola and other viruses in the *Filoviridae* family given that it has a well-conserved sequence and plays multiple essential roles in the viral lifecycle.[23] One of its primary functions is to antagonize a host's innate immunity, particularly RIG-I-like receptor (RLR)-mediated detection of viral nucleic acids, to prevent an interferon (IFN) response and signaling of neighboring cells to heighten their antiviral defenses.[24-26] Crystal structures have provided a foundation for understanding much about the mechanism of VP35-mediated IFN antagonism.[27,28] For example, they have revealed that VP35's IID binds both the blunt ends and backbone of double-stranded RNA (dsRNA), and that there is a PPI between these dsRNA-binding modes (Fig. 1).[28] Disrupting any of these interactions could potentially render Ebola susceptible to a host's innate immunity. In particular, binding to dsRNA blunt ends plays a dominant role in IFN suppression by Ebola.[29] Indeed, mutations that reduce the IID's affinity for dsRNA blunt ends are sufficient to mitigate IFN antagonism, ultimately attenuating Ebola's pathogenicity.[29-32] So, disrupting this single binding mode could dramatically reduce the impact of an Ebola infection on the host and potentially reduce deleterious effects, including lethality. However, both dsRNA-binding interfaces are large flat surfaces that are difficult for small molecules to bind tightly (Fig. 1). As a result, attempts

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to rationally design small molecules that bind these interfaces have not yielded sufficiently strong leads to warrant clinical development.[33,34] The discovery of cryptic pockets in VP35 could provide new opportunities for drugging this essential viral component.

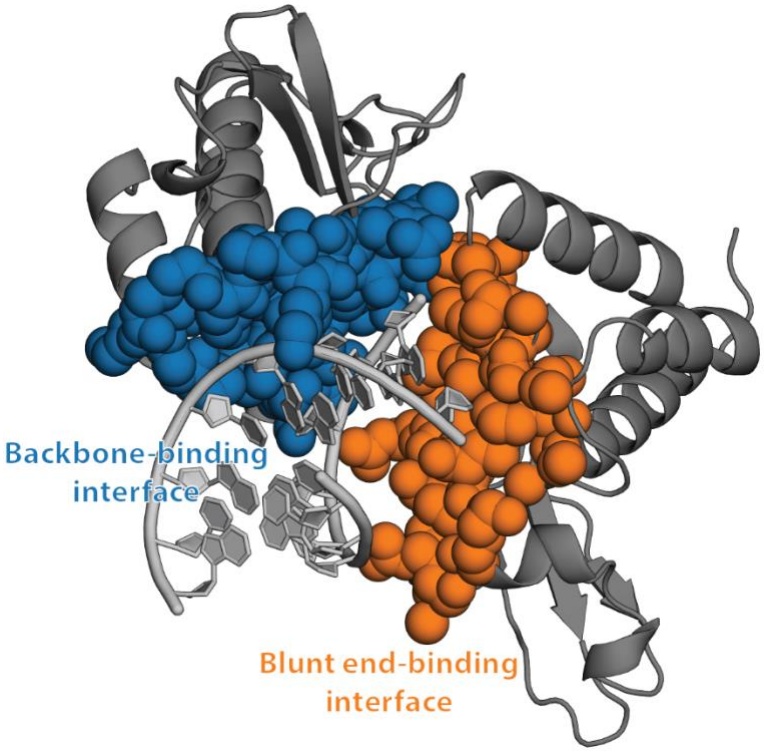

Figure 1. Crystal structure of two copies of VP35's IID (dark gray) bound to dsRNA (light gray) via two flat interfaces (PDB ID 3L25). The backbone-binding interface (blue) and blunt end-binding interface (orange) are shown as spheres to highlight that they lack deep pockets amenable to binding small molecules.

Results

Computer simulations reveal a potentially druggable cryptic pocket.

We applied our fluctuation amplification of specific traits (FAST) simulation algorithm[35] to enhance sampling of structures with large pocket volumes that may harbor cryptic pockets. FAST is a goal-oriented adaptive sampling algorithm that exploits Markov state model (MSM) methods to focus computational resources on exploring regions of conformational space with user-specified structural features. An MSM is a network model of a protein's energy landscape which consists of a set of structural states the protein adopts and the rates of hopping between them.[36,37] Adaptive sampling algorithms enable efficient construction of MSMs by iteratively 1) running a batch of simulations, 2) building an MSM, and 3) selecting a subset of the states that have been identified so far as starting points for the next batch of simulations to maximize the chances of improving the model.[38,39] FAST selects which states to further simulate in a manner that balances exploration/exploitation tradeoffs by considering 1) how well each state optimizes a user defined structural criterion (in this case maximizing the total pocket volume) and 2) the likelihood of discovering new conformational states.[35] After

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

running FAST, we gathered additional statistics by running simulations from each state on the Folding@home distributed computing environment, which brings together the computing resources of tens of thousands of citizen scientists who volunteer to run simulations on their personal computers. Our final model has 11,891 conformational states, providing a detailed characterization of the different structures the IID adopts but making manual interpretation of the model difficult.

To identify cryptic pockets within the large ensemble captured by our MSM, we applied our exposons analysis pipeline.[11] An exposon is a cluster of residues that undergo cooperative changes in their solvent exposure. Coupling between the solvent exposure of every pair of residues is quantified using a mutual information metric, as described in Methods. Exposons are often associated with cryptic sites because the opening/closing of such pockets gives rise to cooperative increases/decreases in the solvent exposure of surrounding residues. Importantly, once an exposon has been identified, our MSM framework provides a facile means to identify the conformational changes that give rise to that exposon.

The IID has two significant exposons, one of which corresponds to a large cryptic pocket. The blue exposon (Fig. 2A and 2B) consists of a set of strongly-coupled residues in helix 7 and adjacent loops and secondary structure elements. Visualizing the conformational change that gives rise to this exposon reveals a substantial displacement of helix 7, creating a large cryptic pocket between it and the helical domain (Fig. 2C, Supplementary Movie 1). A number of residues that are displaced along with helix 7 (i.e. A306, K309, and S310) make Van der Waals contacts with the dsRNA backbone in the dsRNA-bound crystal structure,[28] so targeting this cryptic pocket could directly disrupt this binding mode. Retrospective analysis of other validated drug targets suggests cryptic sites created by the movement of secondary structure elements, such as the displacement of helix 7, are often druggable.[40] The potential druggability of this cryptic site is also supported by application of the FTMap algorithm,[41,42] which predicts a number of hotspots within the pocket where small molecules could form a variety of energetically-favorable interactions (Supplementary Fig. 1). Unfortunately, disrupting backbone binding is of less therapeutic utility than disrupting blunt end binding and it is unknown whether the contacts between A306, K309, and S310 are essential for backbone binding. Therefore, it is unclear from this analysis alone whether drugging this newly discovered cryptic pocket would be useful.

The second exposon (orange in Fig. 2) encompasses portions of both dsRNA-binding interfaces, but it does not correspond to a cryptic pocket. This exposon includes residues that bind dsRNA's backbone (i.e. S272) and residues that interact with both the blunt ends and backbone of dsRNA (i.e. F239, Q274, and I340).[28] Therefore, altering the conformational preferences of the second exposon could potentially disrupt the blunt end-binding mode and its crucial role in Ebola's ability to evade an immune response. However, the largest conformational change involved in the formation of this exposon is a displacement of the loop between helices 3 and 4 (Fig. 2D, Supplementary Movie 2). This rearrangement does not create a cryptic pocket that is large enough to

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

accommodate drug-like molecules, so it is not obvious how to directly manipulate the orange exposon.

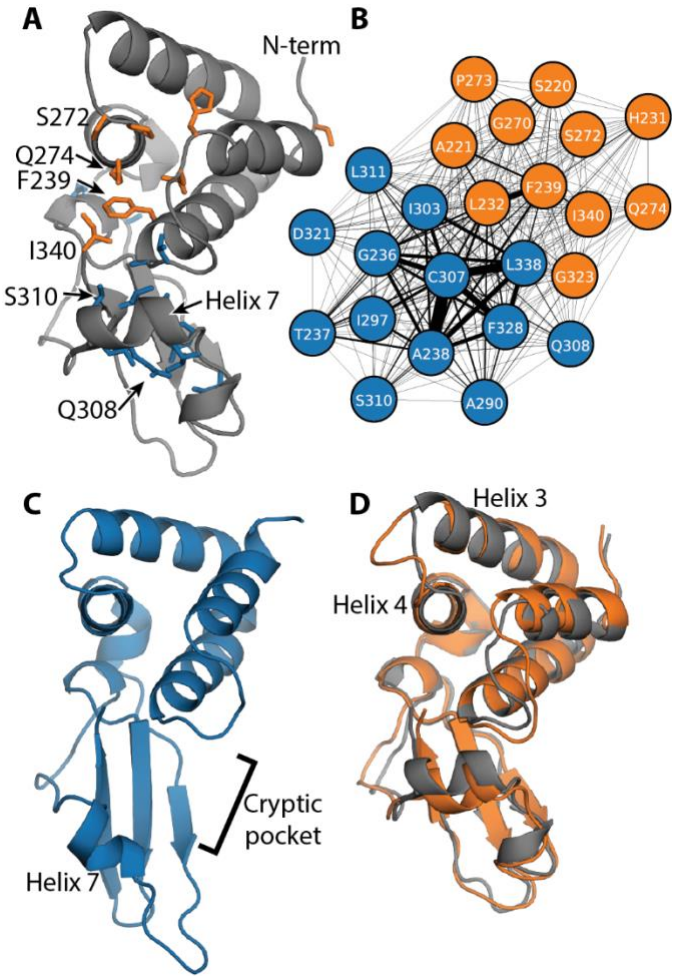

Figure 2. Exposons identify a large cryptic pocket and suggest potential allosteric coupling. A) Structure of VP35's IID highlighting residues in two exposons (blue and orange), the N-terminus (N-term), and C-terminus (I340). B) Network representation of the coupling between the solvent exposure of residues in the two exposons. The edge width between residues is proportional to the mutual information between them. C) Structure highlighting the opening of a cryptic pocket via the displacement of helix 7 that gives rise to the blue exposon. D) Structure highlighting the conformational change that gives rise to the orange exposon overlaid on the crystal structure (gray) to highlight that the rearrangements are subtler than in the blue exposon.

The cryptic pocket is allosterically coupled to the blunt end-binding interface. Even though the cryptic pocket does not coincide with the interface of VP35's IID that binds dsRNA blunt ends, it could still serve as a cryptic allosteric site that allosterically controls RNA binding. Indeed, the physical proximity of the two exposons and the coupling between them both hint at the possibility for allosteric coupling. Furthermore, our exposons analysis could easily underestimate this coupling given that it focuses on correlated transitions of residues between solvent exposed and completely buried

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

states, leaving it blind to more subtle conformational fluctuations and the coupling of residues that are always buried (or always exposed).

To explore the potential for a broader allosteric network, we applied the correlation of all rotameric and dynamical states (CARDS) algorithm.[43] CARDS classifies each dihedral in each snapshot of a simulation as being in one of three rotameric states (gauche+, gauche-, or trans) and one of two dynamical states (ordered or disordered). A dihedral is said to be disordered if it is rapidly hopping between different structural states, and it is classified as ordered if it appears to be locked into a single rotameric state for a prolonged time. The mutual information metric is then used to quantify how strongly coupled the structural and dynamical states of each pair of dihedrals are, enabling CARDS to capture the roles of both concerted structural changes and conformational entropy in allosteric communication. Importantly, CARDS accounts for the potential role of residues that are always buried or always exposed to solvent and subtle conformational changes that do not alter the solvent exposure of residues.

CARDS reveals a broader allosteric network than that identified by our exposons analysis and suggests strong coupling between the cryptic pocket and blunt end-binding interface (Fig. 3). This network consists of five communities of strongly coupled residues, four of which coincide with large portions of the two dsRNA-binding interfaces. One of these communities (orange) is a hub in the network, having significant coupling to all the other communities. It encompasses part of the orange exposon, particularly residues around the loop between helices 3 and 4. The orange CARDS community and exposon both capture Q274, which engages in both dsRNA-binding interfaces, and S272, which contacts the backbone.[28] However, the CARDS community includes many additional residues not captured by exposons analysis. Examples include I278, which engages in both dsRNA-binding interfaces, and D271, which is part of the PPI between the two binding modes.[28] One of the orange community's strongest allosteric connections is to the green community. This community encompasses the rest of the residues in the orange exposon, including F239 and I340, which are part of both dsRNA-binding interfaces.[28] The green community also captures additional residues, reaching deep into the helical domain. The orange community is also strongly coupled to the blue community, which includes much of helix 7 and nearby residues that move to give rise to the cryptic pocket that was captured by the blue exposon. Notably, the orange and blue communities are both coupled to a cyan cluster that was not hinted at by our exposons analysis because the residues involved are always solvent exposed. It includes R322, which is part of the blunt end-binding interface and the PPI between the two binding modes, and K282, which also contacts dsRNA blunt ends.[28] In addition, this community includes K339, which is an important determinant of the electrostatic favorability of dsRNA binding.[28] Together, these results suggest that opening of the cryptic pocket could strongly impact residues involved in both dsRNA-binding interfaces, as well as the PPI between the two binding modes.

To understand the potential impact of targeting the cryptic pocket on the blunt end-binding mode, we performed a dimensionality reduction based on the orange community. Since this community is a hub in the allosteric network, we reasoned that

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

performing a dimensionality reduction based on the structural/dynamical preferences of this community and examining representative structures would report on what is happening throughout the protein. To determine the relative importance of the structural and dynamical preferences of this community, we compared the magnitudes of the structural and dynamical components of CARDS. This analysis revealed that concerted structural changes are the dominant mode of allosteric communication in the IID, rather than conformational entropy and dynamical allostery (Supplementary Fig. 2). Therefore, examining structures where the orange community undergoes large conformational changes might reveal the perturbations these motions induce elsewhere in the protein. To understand what sort of conformational changes are present, we performed a dimensionality reduction on our simulation dataset by applying principal component analysis (PCA) to the distances between the Cβ atoms of every pair of residues in the orange community. Projecting our MSM onto the first two principal components (PC1 and PC2) reveals one dominant free energy minimum and a broad excited state (Fig. 3C).

Comparing representative structures for the orange community's two dominant states suggests the cryptic pocket is indeed a cryptic allosteric site, targeting of which could allosterically disrupt binding of VP35's IID to dsRNA blunt ends. Most importantly, conformational changes of the orange community are associated with opening of the cryptic pocket (Fig. 3D). Therefore, targeting the cryptic pocket could modulate the entire allosteric network in addition to its potential direct effect on the backbone-binding mode. Comparing the structures also reveals that the end of helix 4 frays and the preceding loop, which sits at the PPI between the two dsRNA-binding modes, is displaced. So, targeting the cryptic pocket could allosterically modulate this PPI. Finally, we note a substantial reshuffling of residues F239, H231, and P273 and modest displacements of R322 and K339. Previous work has demonstrated that F239A, R322A, and K339A substitutions are each sufficient to disrupt dsRNA binding and IFN suppression.[28] CARDS analysis suggests targeting the cryptic pocket could allosterically alter the structures of these residues and have a similar impact on dsRNA binding.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

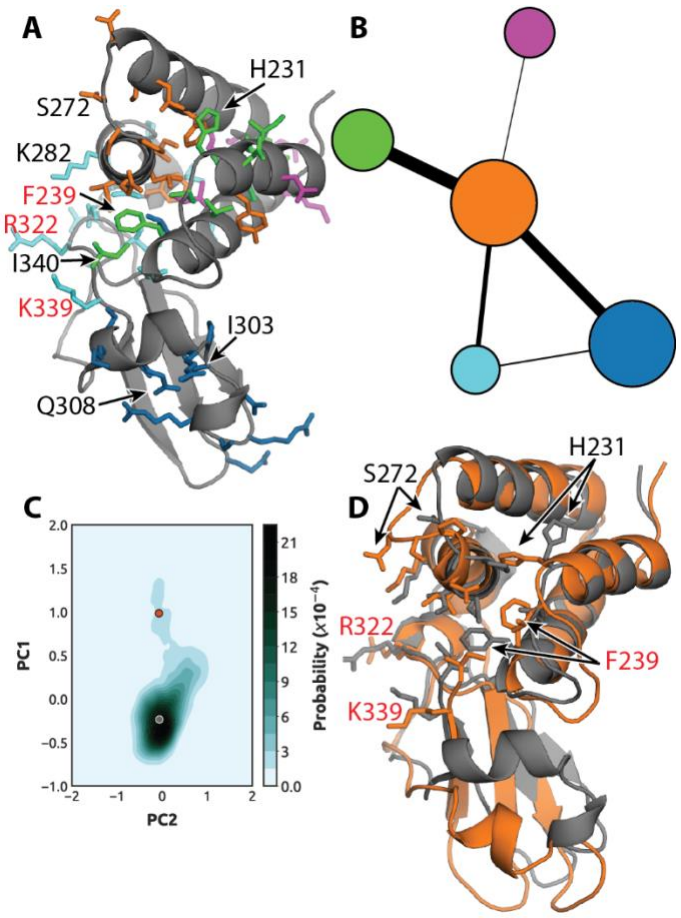

Figure 3. Allosteric network revealed by the CARDS algorithm. A) Structure of VP35's IID with residues in the allosteric network shown in sticks and colored according to which of five communities they belong to. Substitution of residues labeled in red with alanine disrupts binding to dsRNA blunt ends and results in a dramatic reduction in immune suppression. B) Network representation of the coupling between communities of residues, colored as in A. Node size is proportional to the strength of coupling between residues in the community, and edge widths are proportional to the strength of coupling between the communities. C) Free energy landscape of the orange exposon projected onto the first two principal components, PC1 and PC2, highlighting the centroid structures of the free energy minimum (gray circle) and excited state (orange circle). D) Structures of the centroids (colored as in panel C) capture opening of the cryptic pocket and rearrangements involving key residues for PPIs and PNIs.

Thiol labeling experiments corroborate the predicted cryptic pocket.

One way to experimentally test our prediction of a cryptic pocket is to probe for solvent exposure of residues that are buried in available crystal structures but become exposed to solvent upon pocket opening. Cysteines are particularly appealing candidates for such experiments because 1) they have a low abundance and 2) their thiol groups are highly reactive, so it is straightforward to detect exposed cysteines by introducing labeling reagents that covalently bind accessible thiols. Fortuitously, VP35's IID has two cysteines (C307 and C326) that are buried in available crystal structures but become

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

exposed to solvent when the cryptic pocket opens (Fig. 4A). There is also a cysteine (C275) that is on the surface of the apo crystal structure[27] and a fourth cysteine (C247) that is buried in the helical bundle. C275 is typically solvent exposed in our simulations, as expected based on the crystallographic data. Examining the solvent exposure of C247 revealed it is sometimes exposed to solvent via an opening of helix 1 relative to the rest of the helical bundle (Supplementary Fig. 3), but FTMap did not identify any hotspots that are likely to bind drug-like molecules in this region. Therefore, we expect to observe labeling of all four cysteines on a timescale that is faster than global unfolding of the protein.

To experimentally test our predicted pocket, we applied a thiol labeling technique that probes the solvent exposure of cysteine residues.[44] For these experiments, 5,5'-Dithiobis-(2-Nitrobenzoic Acid) (also known as DTNB or Ellman's reagent, Fig. 4B) is added to a protein sample. Upon reaction with the thiol group of an exposed cysteine, DTNB breaks into two TNB molecules, one of which remains covalently bound to the cysteine while the other is released into solution. The accumulation of free TNB can be quantified based on the increased absorbance at 412 nm. We have previously applied this technique to test predicted pockets in β-lactamase enzymes.[11,45]

As expected from our computational model, the observed signal from our thiol labeling experiments is consistent with opening of the cryptic pocket (Fig. 4C). Absorbance curves are best fit by four exponentials, each with an approximately equivalent amplitude that is consistent with expectations based on the extinction coefficient for DTNB (Supplementary Fig. 4). To assign these labeling rates to individual cysteines, we systematically mutated the cysteines to serines, performed thiol labeling experiments, and assessed which rates disappeared and which remained (Supplementary Table 1). For example, labeling of the C275S variant lacks the very fastest rate for wild-type, consistent with the intuition that a residue that is surfaced exposed in the crystal structure (i.e. C275) should label faster than residues that are generally buried. To test whether the observed labeling could be due to an alternative process, such as global unfolding, we determined the population of the unfolded state and unfolding rate of VP35's IID under native conditions (Supplementary Table 2) and the intrinsic labeling rate for each cysteine (Supplementary Table 3). As shown in Fig. 4C, the observed labeling rates are all considerably faster than the expected labeling rate from the unfolded state at a range of DTNB concentrations. This result confirms that labeling of all four cysteines arises from fluctuations within the native state, consistent with our computational predictions. Furthermore, the exposure of C247 is far rarer than C307 or C326 (equilibrium constants for the exposure of C247 and C307 are $5.4 \times 10^{-4} \pm 8.1 \times 10^{-6}$ and $8.5 \times 10^{-2} \pm 2.8 \times 10^{-3}$, respectively). Therefore, a ligand would have to pay a greater energetic cost to stabilize the conformational change that exposes C247 than to stabilize the open state of the cryptic allosteric site created by the motion of helix 7.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

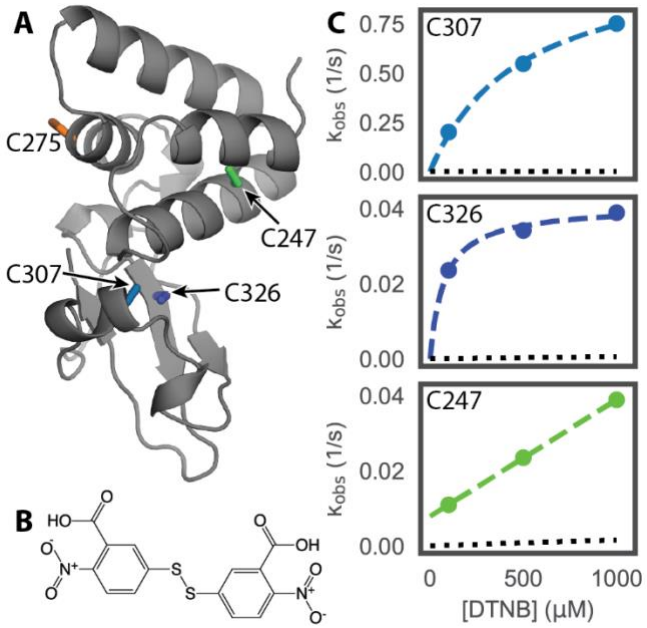

Figure 4. Thiol labeling supports the existence of the predicted cryptic pocket. A) Structure of VP35's IID highlighting the locations of the four native cysteines (sticks). C307 and C326 are both buried and point into the proposed cryptic pocket. B) Structure of the DTNB labeling reagent. C) Observed labeling rates (circles) at a range of DTNB concentrations. Fits to the Linderstrøm-Lang model are shown in dashed colored lines and the expected labeling rate from the unfolded state is shown as black dotted lines. The mean and standard deviation from three replicates are shown but error bars are generally smaller than the symbols. Labeling for C275 is not shown because it is surface exposed in both the available crystal structures and our simulations, and it behaves as expected (labeling rate greater than 1 s^{-1} with a linear dependence on [DTNB]).

Stabilizing the open cryptic pocket allosterically disrupts binding to dsRNA blunt ends.

We reasoned that covalent attachment of TNB to C307 and C326 would provide a means to capture the open pocket and assess the impact of stabilizing this state on dsRNA binding. Addition of TNB to these cysteines is sterically incompatible with the closed conformation of VP35's IID that has been observed crystallographically. TNB's mass of ~198 Da is also similar to many drug fragments used in screening campaigns, making it a reasonable surrogate for the type of effect one might achieve with a fragment hit. Given that we already know DTNB labels the IID's cysteines, a TNB-labeled sample is easily obtainable by waiting until the labeling reaction goes to completion. Finally, we have previously used this same strategy to identify cryptic pockets that exert allosteric control over the activity of β-lactamase enzymes.[11,45] To ensure that we primarily capture the effect of labeling on pocket opening, we used a C247S/C275S variant of VP35's IID that only has cysteines pointing into the cryptic pocket. As with the wild-type protein, thiol labeling of the C247S/C275S variant is consistent with the formation of the proposed cryptic pocket (Supplementary Fig. 5).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To measure the effect of TNB labeling on the IID's interaction with dsRNA, we developed a fluorescence polarization (FP) assay for monitoring dsRNA binding. Paralleling our past work on VP35-peptide interactions,[46] we added varying concentrations of C247S/C275S IID to a fixed concentration of 25-bp dsRNA with a fluorescein isothiocyanate (FITC) conjugation at one end (Supplementary Table 4). Free FITC-dsRNA emits depolarized light upon excitation with polarized light because of the molecule's fast rotation. Binding of one or more VP35 molecules restricts the motion of FITC-dsRNA, resulting in greater emission of polarized light, which is best monitored by the change in anisotropy.[47]

Monitoring the binding of unlabeled protein to 25-bp dsRNA with either blunt ends or 3' overhangs demonstrates that our FP assay is sensitive to both dsRNA-binding modes and gives affinities that are consistent with past work. Past work using a dot-blot assay to measure binding reported an apparent dissociation constant (K_d) for blunt-ended dsRNA of 3.4±0.07 μM.[29] Furthermore, sterically hindering binding of the IID to dsRNA blunt ends by adding 2-nucleotide overhangs to the 3' of the RNA reduces the apparent dsRNA-binding affinity by 10-fold.[48] This weaker interaction was attributed to the backbone-binding mode since it is still available to VP35's IID even when the presence of an overhang inhibits blunt end binding. Similarly, our FP assay gives an apparent K_d of 3.6±0.34 μM for blunt-ended dsRNA (Fig. 5A). Addition of 3' overhangs results in a strong rightwards shift of the binding curve, consistent with at least a 5-fold reduction in the apparent binding affinity (apparent K_d of 20.4±1.1 μM). However, an upper baseline could not be captured due to limitations in the protein's solubility, so this apparent K_d is a lower bound. The data are also fit well assuming an apparent K_d of 30.1±7.2 μM that was reported previously.[48]

Repeating our FP assay with TNB-labeled protein reveals that labeling allosterically reduces the affinity for blunt-ended dsRNA by at least 5-fold (Fig. 5A). Solubility limitations again prevented us from observing complete binding curves for labeled protein, but the data are sufficient to demonstrate that TNB-labeling has at least as strong an effect on binding as addition of a 3' overhang. As a control to ensure that labeling does not disrupt binding by simply unfolding the protein, we measured the circular dichroism (CD) spectra of labeled and unlabeled protein. The similarity between the CD spectra (Fig. 5B) demonstrates that the IID's overall fold is not grossly perturbed. Since the cryptic pocket does not coincide with the blunt end-binding interface, our results suggests the impact on dsRNA binding is allosteric. Furthermore, past work demonstrated that reducing the blunt end-binding affinity by as little as 3-fold is sufficient to allow a host to mount an effective immune response,[28] so targeting our cryptic pocket could be of great therapeutic value.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

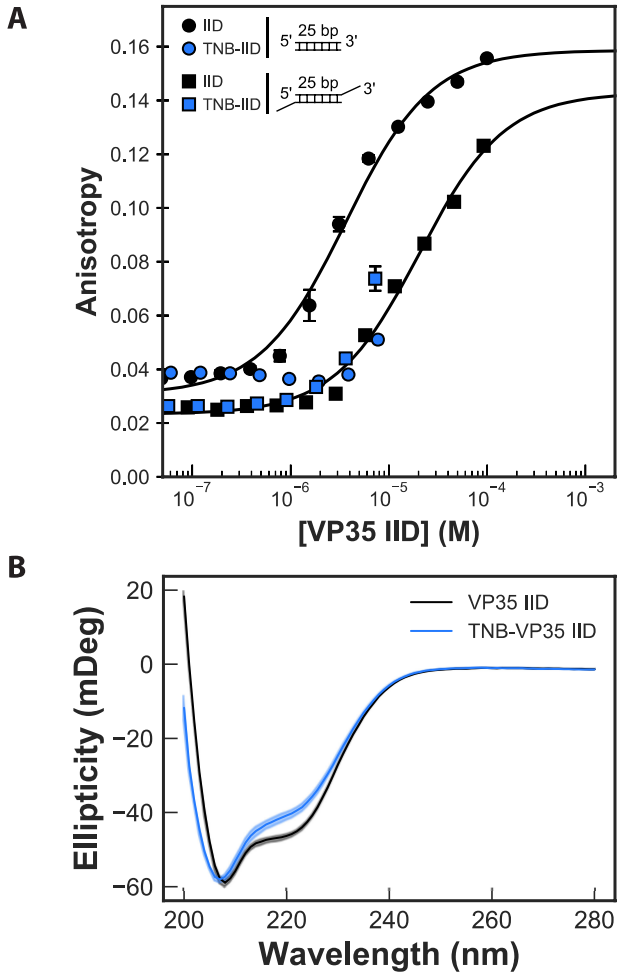

Figure 5. Stabilizing the open cryptic pocket in VP35's IID disrupts dsRNA binding. A) Binding of both TNB-labeled and unlabeled C247S/C275S variants of the IID to two different dsRNA constructs. This protein variant only has cysteines in the cryptic pocket. The RNA constructs both have a 25-bp double-stranded segment, and one has 2 nucleotide overhangs on the 3' ends. The anisotropy was measured via a fluorescence polarization assay and fit to a single-site binding model (black lines). The mean and standard deviation from three replicates are shown but error bars are generally smaller than the symbols. B) Circular dichroism (CD) spectra of labeled and unlabeled protein demonstrate that labeling does not unfold the protein. The opaque and semi-transparent lines represent the mean and standard deviation, respectively, from three replicates.

Discussion

We have identified a cryptic allosteric site in the IID of the Ebola VP35 protein that provides a new opportunity to target this essential viral component. Past work identified several sites within the VP35 IID that are critical for immune evasion and viral replication,[23,26,31,32] but structural snapshots captured crystallographically lacked druggable pockets.[27,28] We used adaptive sampling simulations to access more of the ensemble of conformations that VP35 adopts, uncovering an unanticipated cryptic pocket. While the pocket directly coincides with the interface that binds the backbone of dsRNA, it was not clearly of therapeutic relevance since binding dsRNA's blunt ends is

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

more important for Ebola's immune evasion mechanism.[29] However, our simulations also suggested the cryptic pocket is allosterically coupled to the blunt end-binding interface and, therefore, could modulate this biologically-important interaction. Subsequent experiments confirmed that fluctuations within the folded state of the IID expose two buried cysteines that line the proposed cryptic pocket to solvent. Moreover, covalently modifying these cysteines to stabilize the open form of the cryptic pocket allosterically disrupts binding to dsRNA blunt ends by at least 5-fold. Previous work demonstrated that reducing the binding affinity by as little as 3-fold is sufficient to allow a host to mount an effective immune response.[28] Therefore, it may be possible to attenuate the impact of viral replication and restrict pathogenicity by designing small molecules to target the cryptic allosteric site we report here.

More generally, our results speak to the power of simulations to provide simultaneous access to both hidden conformations and dynamics with atomic resolution. Such information is extremely difficult to obtain from single structural snapshots or powerful techniques that report on dynamics without directly yielding structures, such as NMR and hydrogen deuterium exchange. As a result, simulations are a powerful means to uncover unanticipated features of proteins' conformational ensembles, such as cryptic pockets and allostery, providing a foundation for the design of further experiments. We anticipate such simulations will enable the discovery of cryptic pockets and cryptic allosteric sites in other proteins, particularly those that are currently considered undruggable. Furthermore, the detailed structural insight from simulations will facilitate the design of small molecule drugs that target these sites.

Methods

Molecular dynamics simulations and analysis

Simulations were initiated from chain B of PDB 3L25[28] and run with Gromacs[49] using the amber03 force field[50] and TIP3P explicit solvent[51] at a temperature of 300 K and 1 bar pressure, as described previously.[52] We first applied our FAST-pockets algorithm[35] to balance 1) preferentially simulating structures with large pocket volumes that may harbor cryptic pockets with 2) broad exploration of conformational space. For FAST, we performed 10 rounds of simulations with 10 simulations/round and 80 ns/simulation. To acquire better statistics across the landscape, we performed an RMSD-based clustering using a hybrid k-centers/k-medoids algorithm[53] implemented in Enspara[54] to divide the data into 1,000 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 122 μs.

Exposons were identified using our previously described protocols,[11] as implemented in Enspara.[54] Briefly, the solvent accessible surface area (SASA) of each residue's side-chain was calculated using the Shrake-Rupley algorithm[55] implemented in MDTraj[56] using a drug-sized probe (2.8 Å sphere). Conformations were clustered based on the SASA of each residue using a hybrid k-centers/k-medoids algorithm, using a 2.5 $Å^2$ distance cutoff and 5 rounds of k-medoids updates. A Markov time of 6 ns was selected based on the implied timescales test (Supplementary Fig. 6). The center of each cluster was taken as an exemplar of that conformational state, and residues were classified as exposed if their SASA exceeded 2.0 $Å^2$ and buried otherwise. The mutual

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

information between the burial/exposure of each pair of residues was then calculated based on the MSM (i.e. treating the centers as samples and weighting them by the equilibrium probability of the state they represent). Finally, exposons were identified by clustering the matrix of pairwise mutual information values using affinity propagation.[57]

The CARDS algorithm[43] was applied to identify allosteric coupling using our established protocols,[58] as implemented in Enspara.[54] Briefly, each dihedral angle in each snapshot of the simulations was assigned to one of three rotameric states (gauche+, gauche-, or trans) and one of two dynamical states (ordered or disordered). The total coupling between each pair of dihedrals X and Y was then calculated as $I(X_R, Y_R) + I(X_R, Y_D) + I(X_D, Y_R) + I(X_D, Y_D)$, where I is the mutual information metric, X_R is the rotameric state of dihedral X, and X_D is the dynamical state of dihedral X. The term $I(X_R, Y_R)$ is the purely structural coupling, while the sum of the other three terms is referred to as the disorder-mediated coupling. The dihedral level couplings were coarse-grained into residue-level coupling by summing the total coupling between all the relevant dihedrals and the network was filtered to only retain significant edges.[59] Finally, communities of coupled residues were identified by clustering the residue-level matrix of total couplings using affinity propagation.[57]

Protein expression and purification

All variants of VP35's IID were purified from the cytoplasm of *E. coli* BL21(DE3) Gold cells (Agilent Technologies). Variants were generated using the site directed mutagenesis method and confirmed by DNA sequencing. Transformed cells were grown 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 hours then centrifuged after which the pellet was resuspended in 20 mM Sodium Phosphate pH 8, 1 M sodium chloride, 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, TEV digestion, cation exchange (BioRad UNOsphere Rapid S column), and size exclusion chromatography (BioRad Enrich SEC 70 column) 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 of DTNB were added to protein and change in absorbance was measured in either an SX-20 Stopped Flow instrument (Applied Photophysics, Leatherhead, UK), or an Agilent Cary60 UV-vis spectrophotometer at 412 nm until the reaction reached steady state (~300 s). Data were fit with a Linderstrøm-Lang model to extract the thermodynamics and/or kinetics of pocket opening, as described in detail previously.[11] As a control, the equilibrium constant for folding and the unfolding rate were measured (Supplementary Table 2) 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 single exponential fits to unfolding curves monitored by fluorescence after the addition of urea, as described previously.[11,45,60] Fluorescence data were collected using a Photon Technology

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

International Quanta- Master 800 rapid excitation spectrofluorometer with Quantum Northwest Inc. TC-125 Peltier-controlled cuvette holder.

Fluorescence polarization binding assay

Binding affinities between variants of VP35's IID and dsRNA were measured using fluorescence polarization in 10 mM Hepes pH 7, 150 mM NaCl, 1 mM $MgCl_2$. A 25 base pair FITC-dsRNA (Integrated DNA Technologies) substrate with and without a 2 nucleotide 3' overhang was included at 100 nM. 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 as described previously.[47] TNB-labeled samples were generated by allowing DTNB and VP35's IID to react for 3 minutes and then removing excess DTNB with a Zeba spin desalting columns (Thermo Fisher Scientific). A single-site binding model was sufficient to fit the data.

Acknowledgements

We are grateful to the citizen scientists who participate in Folding@home for volunteering to run simulations on their personal computers. This work was funded by NSF CAREER Award MCB-1552471 and NIH grant R01 GM124007 (Bowman), as well as NIH grants R01AI123926, P01AI120943, and R01AI143292 (Amarasinghe). GRB holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund and a Packard Fellowship for Science and Engineering from The David & Lucile Packard Foundation. MAC was supported by the 5R25GM103757 IMSD program and SS was supported by a MilliporeSigma Fellowship. We thank Drs. Timothy M. Lohman and Alexander G. Kozlov for advice on FP assays.

1 Hopkins, A. & Groom, C. The druggable genome. *Nat Rev Drug Discov* **1**, 727-730 (2002).

2 Arkin, M. R. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov* **3**, 301-317 (2004).

3 Johnson, D. K. & Karanicolas, J. Computational Screening and Design for Compounds that Disrupt Protein-protein Interactions. *Current topics in medicinal chemistry* **17**, 2703-2714 (2017).

4 Knoverek, C. R., Amarasinghe, G. K. & Bowman, G. R. Advanced Methods for Accessing Protein Shape-Shifting Present New Therapeutic Opportunities. *Trends Biochem Sci* **44**, 351-364 (2019).

5 Vajda, S., Beglov, D., Wakefield, A. E., Egbert, M. & Whitty, A. Cryptic binding sites on proteins: definition, detection, and druggability. *Current opinion in chemical biology* **44**, 1-8 (2018).

6 Hardy, J. A. & Wells, J. A. Searching for new allosteric sites in enzymes. *Curr Opin Struct Biol* **14**, 706-715 (2004).

7 Erlanson, D. A. *et al.* Site-directed ligand discovery. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 9367-9372 (2000).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

8 Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A. & Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503**, 548-551 (2013).

9 Keedy, D. A. *et al.* An expanded allosteric network in PTP1B by multitemperature crystallography, fragment screening, and covalent tethering. *Elife* **7** (2018).

10 Bowman, G. R. & Geissler, P. L. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 11681-11686 (2012).

11 Porter, J. R. *et al.* Cooperative changes in solvent exposure identify cryptic pockets, switches, and allosteric coupling. *Biophysical Journal* (2019).

12 Johnson, D. K. & Karanicolas, J. Druggable protein interaction sites are more predisposed to surface pocket formation than the rest of the protein surface. *PLoS Computational Biology* **9**, e1002951 (2013).

13 Oleinikovas, V., Saladino, G., Cossins, B. P. & Gervasio, F. L. Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations. *Journal of the American Chemical …*, jacs.6b05425 (2016).

14 Cimermancic, P. *et al.* CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites. *J Mol Biol* **428**, 709-719 (2016).

15 Schmidt, D., Boehm, M., McClendon, C. L., Torella, R. & Gohlke, H. Cosolvent-Enhanced Sampling and Unbiased Identification of Cryptic Pockets Suitable for Structure-Based Drug Design. *J Chem Theory Comput* **15**, 3331-3343 (2019).

16 Cuchillo, R., Pinto-Gil, K. & Michel, J. A Collective Variable for the Rapid Exploration of Protein Druggability. *J Chem Theory Comput* **11**, 1292-1307 (2015).

17 Ghanakota, P. & Carlson, H. A. Moving Beyond Active-Site Detection: MixMD Applied to Allosteric Systems. *The Journal of Physical Chemistry B* **120**, 8685-8695 (2016).

18 Wassman, C. D. *et al.* Computational identification of a transiently open L1/S3 pocket for reactivation of mutant p53. *Nat Commun* **4**, 1407 (2013).

19 Schames, J. *et al.* Discovery of a novel binding trench in HIV integrase. *J. Med. Chem* **47**, 1879-1881 (2004).

20 Cross, R. W., Mire, C. E., Feldmann, H. & Geisbert, T. W. Post-exposure treatments for Ebola and Marburg virus infections. *Nat Rev Drug Discov* **17**, 413-434 (2018).

21 Keshwara, R., Johnson, R. F. & Schnell, M. J. Toward an Effective Ebola Virus Vaccine. *Annual review of medicine* **68**, 371-386 (2017).

22 Mulangu, S. *et al.* A Randomized, Controlled Trial of Ebola Virus Disease Therapeutics. *N Engl J Med* **381**, 2293-2303 (2019).

23 Messaoudi, I., Amarasinghe, G. K. & Basler, C. F. Filovirus pathogenesis and immune evasion: insights from Ebola virus and Marburg virus. *Nature reviews. Microbiology* **13**, 663-676 (2015).

24 Basler, C. F. *et al.* The Ebola virus VP35 protein inhibits activation of interferon regulatory factor 3. *J Virol* **77**, 7945-7956 (2003).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

25 Cárdenas, W. B. *et al.* Ebola virus VP35 protein binds double-stranded RNA and inhibits alpha/beta interferon production induced by RIG-I signaling. *Journal of virology* **80**, 5168-5178 (2006).

26 Hartman, A. L., Towner, J. S. & Nichol, S. T. 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).

27 Leung, D. W. *et al.* Structure of the Ebola VP35 interferon inhibitory domain. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 411-416 (2009).

28 Leung, D. W. *et al.* Structural basis for dsRNA recognition and interferon antagonism by Ebola VP35. *Nature structural & molecular biology* **17**, 165-172 (2010).

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

30 Hartman, A. L., Dover, J. E., Towner, J. S. & Nichol, S. T. Reverse genetic generation of recombinant Zaire Ebola viruses containing disrupted IRF-3 inhibitory domains results in attenuated virus growth in vitro and higher levels of IRF-3 activation without inhibiting viral transcription or replication. *J Virol* **80**, 6430-6440 (2006).

31 Prins, K. C. *et al.* Mutations abrogating VP35 interaction with double-stranded RNA render Ebola virus avirulent in guinea pigs. *J Virol* **84**, 3004-3015 (2010).

32 Prins, K. C. *et al.* Basic residues within the ebolavirus VP35 protein are required for its viral polymerase cofactor function. *Journal of virology* **84**, 10581-10591 (2010).

33 Brown, C. S. *et al.* In silico derived small molecules bind the filovirus VP35 protein and inhibit its polymerase cofactor activity. *J Mol Biol* **426**, 2045-2058 (2014).

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

35 Zimmerman, M. I. & Bowman, G. R. FAST Conformational Searches by Balancing Exploration/Exploitation Trade-Offs. *Journal of chemical theory and computation* **11**, 5747-5757 (2015).

36 Pande, V. S., Beauchamp, K. & Bowman, G. R. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **52**, 99-105 (2010).

37 Prinz, J.-H. *et al.* Markov models of molecular kinetics: generation and validation. *J Chem Phys* **134**, 174105 (2011).

38 Bowman, G. R., Huang, X. & Pande, V. S. Network models for molecular kinetics and their initial applications to human health. *Cell Research* **20**, 622-630 (2010).

39 Hruska, E., Abella, J. R., Nuske, F., Kavraki, L. E. & Clementi, C. Quantitative comparison of adaptive sampling methods for protein dynamics. *J Chem Phys* **149**, 244119 (2018).

40 Beglov, D. *et al.* Exploring the structural origins of cryptic sites on proteins. *Proc Natl Acad Sci U S A* **115**, E3416-E3425 (2018).

41 Ngan, C. H. *et al.* FTMAP: extended protein mapping with user-selected probe molecules. *Nucleic Acids Res* **40**, W271-275 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

42 Kozakov, D. *et al.* The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protoc* **10**, 733-755 (2015).

43 Singh, S. & Bowman, G. R. Quantifying Allosteric Communication via Both Concerted Structural Changes and Conformational Disorder with CARDS. *Journal of chemical theory and computation* **13**, 1509-1517 (2017).

44 Bernstein, R., Schmidt, K., Harbury, P. & Marqusee, S. Structural and kinetic mapping of side-chain exposure onto the protein energy landscape. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 10532-10537 (2011).

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

46 Liu, G. *et al.* A Sensitive in Vitro High-Throughput Screen To Identify Pan-filoviral Replication Inhibitors Targeting the VP35-NP Interface. *ACS infectious diseases* **3**, 190-198 (2017).

47 Kozlov, A. G., Galletto, R. & Lohman, T. M. SSB-DNA binding monitored by fluorescence intensity and anisotropy. *Methods Mol Biol* **922**, 55-83 (2012).

48 Ramanan, P. *et al.* Structural basis for Marburg virus VP35–mediated immune evasion mechanisms. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 20661-20666 (2012).

49 van der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *Journal of Computational Chemistry* **26**, 1701-1718 (2005).

50 Duan, Y. *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).

51 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926 (1983).

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

53 Beauchamp, K. A. *et al.* MSMBuilder2: Modeling Conformational Dynamics at the Picosecond to Millisecond Scale. *J Chem Theor Comp* **7**, 3412-3419 (2011).

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

55 Shrake, A. & Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J Mol Biol* **79**, 351-371 (1973).

56 McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **109**, 1528-1532 (2015).

57 Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science (New York, NY)* **315**, 972-976 (2007).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.09.940510; this version posted February 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

58 Sun, X., Singh, S., Blumer, K. J. & Bowman, G. R. Simulation of spontaneous G protein activation reveals a new intermediate driving GDP unbinding. *eLife* **7**, 19 (2018).

59 Dianati, N. Unwinding the hairball graph: Pruning algorithms for weighted complex networks. *Physical Review E* **93** (2016).

60 Zimmerman, M. I. *et al.* Prediction of New Stabilizing Mutations Based on Mechanistic Insights from Markov State Models. *ACS central science* **3**, 1311-1321 (2017).