Long-distance correlations of rhinovirus capsid dynamics contribute to uncoating and antiviral activity

Amitava Roy and Carol Beth Post

Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN 47907

Edited by Peter J. Rossky, University of Texas at Austin, Austin, TX, and approved February 3, 2012 (received for review November 23, 2011)

Human rhinovirus (HRV) and other members of the enterovirus genus bind small-molecule antiviral compounds in a cavity buried within the viral capsid protein VP1. These compounds block the release of the viral protein VP4 and RNA from inside the capsid during the uncoating process. In addition, the antiviral compounds prevent "breathing" motions, the transient externalization of the N-terminal regions of VP1 and VP4 from the inside of intact viral capsid. The site for externalization of VP1/VP4 or release of RNA is likely between protomers, distant to the binding cavity for antiviral compounds. Molecular dynamics simulations were conducted to explore how the antiviral compound, WIN 52084, alters properties of the HRV 14 capsid through long-distance effect. We developed an approach to analyze capsid dynamics in terms of correlated radial motion and the shortest paths of correlated motions. In the absence of WIN, correlated radial motion is observed between residues separated by as much as 85 Å, a remarkably long distance. The most frequently populated path segments of the network were localized near the fivefold symmetry axis and included those connecting the N termini of VP1 and VP4 with other regions, in particular near twofold symmetry axes, of the capsid. The results provide evidence that the virus capsid exhibits concerted long-range dynamics, which have not been previously recognized. Moreover, the presence of WIN destroys this radial correlation network, suggesting that the underlying motions contribute to a mechanistic basis for the initial steps of VP1 and VP4 externalization and uncoating.

Results and Discussion

Long-distance correlations of rhinovirus capsid dynamics contribute to uncoating and antiviral activity

Human rhinovirus (HRV) and other members of the enterovirus genus bind small-molecule antiviral compounds in a cavity buried within the viral capsid protein VP1. These compounds block the release of the viral protein VP4 and RNA from inside the capsid during the uncoating process. In addition, the antiviral compounds prevent "breathing" motions, the transient externalization of the N-terminal regions of VP1 and VP4 from the inside of intact viral capsid. The site for externalization of VP1/VP4 or release of RNA is likely between protomers, distant to the binding cavity for antiviral compounds. Molecular dynamics simulations were conducted to explore how the antiviral compound, WIN 52084, alters properties of the HRV 14 capsid through long-distance effect. We developed an approach to analyze capsid dynamics in terms of correlated radial motion and the shortest paths of correlated motions. In the absence of WIN, correlated radial motion is observed between residues separated by as much as 85 Å, a remarkably long distance. The most frequently populated path segments of the network were localized near the fivefold symmetry axis and included those connecting the N termini of VP1 and VP4 with other regions, in particular near twofold symmetry axes, of the capsid. The results provide evidence that the virus capsid exhibits concerted long-range dynamics, which have not been previously recognized. Moreover, the presence of WIN destroys this radial correlation network, suggesting that the underlying motions contribute to a mechanistic basis for the initial steps of VP1 and VP4 externalization and uncoating.

Along with release of RNA, the uncoating process is also associated with the loss of VP4 from the viral capsid (13) as well as externalization of the N terminus of VP1 (14). Antibody recognition (15) and results from limited proteolysis of isolated virus (16) provide evidence that the N termini of VP1 and VP4, which are regions located inside the viral capsid (Fig. 1C), are reversibly exposed in the mature capsid, and suggest this externalization is part of the initial stages of uncoating. Notably, in the presence of WIN compounds these regions are not proteolyzed, and thus the motions leading to the exposure of these N termini are dampened, consistent with the activity of WIN compounds being to inhibit uncoating.

The reversible externalization of interior regions of VP1 and VP4, viral uncoating, and the inhibition of these processes by WIN together infer large-scale, concerted motions of the capsid. Further, RNA, VP4, and the N terminus of VP1 are suggested to exit the capsid at the fivefold symmetry axis (17–21) or an interface near the twofold axis (22–24), so that WIN binding in a pocket distant to these sites indicates a long-range activity. These observations motivated MD studies of the solvated HRV 14 capsid, with and without WIN 52084 bound, to probe functional protein dynamics that might underlie the initial stages of the externalization process and uncoating and to further elucidate the physical basis of the long-distance antiviral activity. The atomic fluctuations observed from MD studies were analyzed for radially correlated motions between capsid residues. Radial cross-correlation was examined rather than the more commonly used displacement vector cross-correlation (DCC) to overcome the dependence of the later on directionality (25) (see SI Appendix). We introduce a framework to analyze long-distance, concerted motions using a network based on radial cross-correlation coefficients. Potential pathways underlying the motions were traced in the radial correlation network by using a modification of Dijkstra’s graph searching algorithm (26). The resulting important paths involve residues at spontaneous drug-resistant mutation sites and were sensitive to the presence of the WIN compound. This framework based on a radial correlation network has allowed the previously undescribed identification of correlated motion over a long distance from an MD simulation, and allowed insights into the molecular mechanism of antiviral activity.

Results and Discussion

To investigate large scale, concerted motions of the HRV 14 capsid, trajectories were computed using rotational boundary
Long-Range Correlations. Reasoning that protein externalization and uncoating involve motions with a radial directionality, we examined the correlations in the distance from the capsid center among pairs of \( C_a \) atoms. A breathing motion is reflected in such a radial correlation coefficient in contrast to the DCC, which is generally diminished by angular averaging (25) (see SI Appendix). The absolute radial correlation coefficient, or absolute normalized radial covariance, \( \alpha_{i; j} \) (see Materials and Methods), and the associated 95% CI were determined for pairs of \( C_a \) atoms from \( T_{\text{nowin}} \) and \( T_{\text{win}} \).

To identify concerted motions over large spatial scales, we consider \( \alpha_{i; j} \) values for \( C_a \) atoms of noninteracting residue pairs—i.e., those with a minimum distance between residues greater than the nonbond cutoff distance 14 Å. The distributions for the \( \alpha_{i; j} \) values with 95% CI greater than 0.6 from \( T_{\text{nowin}} \) and \( T_{\text{win}} \) are shown in Fig. 2, blue and red curves, respectively, as a function of the distance between the two \( C_a \) atoms. The time-average values for \( \alpha_{i; j} \) are well converged within 1 ns, and only a small fraction of the total \( C_a-C_a \) radial correlations are greater than 0.6 (see SI Appendix). We find that a large number of radially correlated \( C_a-C_a \) pairs exist for HRV 14 in \( T_{\text{nowin}} \) with the majority from pairs separated by 14 to 30 Å. In striking contrast, the corresponding \( \alpha_{i; j} \) distribution for HRV 14-WIN from \( T_{\text{win}} \) is overall greatly reduced. Moreover, many of the \( T_{\text{nowin}} \) pairs with large radial correlation were separated by distances greater than 45 Å and some with distances as large as 85 to 90 Å. \( \alpha_{i; j} \) values \( \geq 0.8 \) were observed for \( C_a-C_a \) pairs at distances greater than 55 Å. This was not the case for \( T_{\text{win}} \); many fewer \( C_a-C_a \) pairs separated by long distances have a 95% CI for \( \alpha_{i; j} \) greater than 0.6. In addition, the radially correlated \( C_a-C_a \) pairs separated by distances greater than 45 Å are largely nonoverlapping sets for \( T_{\text{nowin}} \) and \( T_{\text{win}} \) as shown by the dotted black line in Fig. 2. This dotted line is the distribution of \( C_a-C_a \) pairs present in the \( T_{\text{nowin}} \) distribution of Fig. 2 (blue curve) and have a 95% CI for \( \alpha_{i; j} \) that is less than 0.6 in \( T_{\text{win}} \). The dotted black curve nearly overlaps the blue curve, demonstrating that these correlated \( C_a-C_a \) pairs in \( T_{\text{nowin}} \) are not as correlated in \( T_{\text{win}} \).

It has long been suggested that RNA, VP4, and the N terminus of VP1 exit the capsid at the fivefold symmetry axis (17–21), although recent cryo-electron microscopy results on the closely related poliovirus, another member of the enterovirus genus, suggest RNA exits at an interface near the twofold symmetry axes (22–24) and the N terminus of VP1 exits though an opening at the base of the canyon (26). To ask if the long-distance radial correlations shown in Fig. 2 might reflect dynamics associated with breathing, uncoating, and WIN-binding effects, we identify radially correlated pairs that have one \( C_a \) atom in the region of the WIN-binding pocket or within 35 Å of the fivefold axis (yellow ring in Fig. 1 B and C). This distribution is shown in Fig. 3A for \( T_{\text{nowin}} \) and \( T_{\text{win}} \). While the \( T_{\text{win}} \) distribution (red curve) is overall reduced and shows little structure beyond 40 Å, the long-distance radially correlated pairs from \( T_{\text{nowin}} \) (blue curve) fall into three groups centered at 46 Å, 64 Å, and 84 Å, designated \( i \), \( ii \), and \( iii \), respectively. These groups are mapped onto the structure in Fig. 3 B and C; for each pair in the group, the \( C_a \) near the fivefold axis is indicated with a blue sphere drawn on one of the five protomers, and the second \( C_a \) is a sphere colored green for group \( i \), gold for group \( ii \), and red for group \( iii \). Residues surrounding the fivefold axis are highly correlated with residues near the fivefold but in the neighboring protomer (green spheres). Also observed are radial correlation with residues in the vicinity of the closest twofold axis (green spheres) and all other twofold symmetry conditions (27–31) with a solvated viral pentamer as the asymmetric unit for simulation. Five copies of the protomer centered at the fivefold axis (Fig. 1 A and B) accurately models microscopic dynamics of not only the WIN-binding pocket but also the region around the fivefold symmetry axis. Trajectories were calculated starting from the crystallographic coordinates of either HRV 14 (32) or HRV 14 with WIN 52084 bound (33) (see SI Appendix). The last 2 ns of 10 3-ns trajectories were used for analysis, for a total time equal to 20 ns for HRV 14 trajectories, referred to as \( T_{\text{nowin}} \), and 20 ns for HRV 14-WIN 52084 trajectories, referred to as \( T_{\text{win}} \). Confidence intervals, CI, of time-averaged quantities were obtained from the bootstrapping method (see SI Appendix).
axes across the pentamer (gold and red spheres). The higher densities of long-range radially correlated motions in $T_{\text{nowin}}$ are therefore between residues in the WIN pocket and residues in regions near the fivefold and twofold symmetry axes. Moreover, as noted above, these correlations are lost in the presence of WIN, so that the behavior is suggestive of functional motions related to externalization and uncoating at the fivefold axis as long proposed (17–21) as well as near the twofold axes as more recently reported (22–24).

The radial correlation in groups $i$, $ii$, and $iii$ in $T_{\text{nowin}}$ largely involve residues from the reversibly externalized N-terminal regions of VP1 or VP4; however, these radial correlations are not present in $T_{\text{win}}$. Fig. 3D displays the subset of the distribution in Fig. 3A corresponding to $C_{\alpha}-C_{\alpha}$ pairs with one $C_{\alpha}$ from either VP1 (residues 1001–1015) or VP4 (residues 4001–4040). The curves are colored as in Fig. 3A. Because the dotted black curve in Fig. 3D closely follows the blue curve from $T_{\text{nowin}}$ for distances $>40$ Å, most of these long-distance $C_{\alpha}-C_{\alpha}$ pairs with large radial correlation coefficients in $T_{\text{nowin}}$ are not radially correlated in $T_{\text{win}}$. As visualized in Fig. 3E and F, these VP1 or VP4 residues with long-distance radial correlations in $T_{\text{nowin}}$ but not in $T_{\text{win}}$ are primarily correlated with residues near the neighboring twofold axis (green spheres) or more distant twofold axes (gold and red spheres). These radial correlations are indicative of concerted motions of these innermost regions of VP1 and VP4 spanning distances of 85 Å.

**Identification of Correlation Paths.** What are the underlying origins of the radial correlations occurring over distances as large as 90 Å? That these correlations exist in free HRV 14 but are largely dampened in the presence of WIN 52084 suggests that identifying the origin of the radial correlations could provide information toward the mechanism for transient externalization, or breathing, and perhaps the uncoating process.

To answer the question, we developed a method for detecting the link between two $C_{\alpha}$ atoms based on a modified version of Figure 3.

Figure 3. (A) Distribution of nonneighbor $C_{\alpha}$ pairs from Fig. 2 with at least one $C_{\alpha}$ within 35 Å of the fivefold symmetry axis (the yellow circle in Fig. 1). (D) $C_{\alpha}$ pairs present in A with at least one $C_{\alpha}$ from N termini of VP1 (residues 1001–1015) or VP4 (residues 4001–4040) are included. Long-distance radially correlated pairs in A fall into three groups centered at 46 Å, 64 Å, and 84 Å, designated $i$, $ii$, and $iii$, respectively. These groups are mapped onto the structure in B and C. For each pair in the group, the $C_{\alpha}$ near the fivefold axis is represented by a blue sphere drawn on one of the five protomers, and the second $C_{\alpha}$ is a sphere colored green for group $i$, gold for group $ii$, and red for group $iii$. D–F are similarly formatted for regions $i$, $ii$, and $iii$ in B. Long-distance radial correlation for residues surrounding the fivefold axis with residues near fivefold but in other protomers and with residues in the vicinity of twofold axes as marked with 5f and 2f, respectively.

Figure 4. A simplified network. Each node in the network corresponds to a $C_{\alpha}$ atom of a residue. Two nodes are connected by an edge if atoms in the corresponding residues are within 14 Å during the simulation. Node $i$ is connected with its neighbors through orange color edges and node $j$ through red color edges. All other edges are shown in gray. Weight of an edge $i\rightarrow j$, $W_{ij} = 1 - C_{\alpha}(i, a)$, is written next to the edge. Weight of path $i \rightarrow j$, $W_{i:j} = W_{ij} \otimes W_{j:i} = 1 - C_{\alpha}(i, a) \cdot C_{\alpha}(a, b)$. Shortest path between two nodes $i$ and $j$ is $i \rightarrow j$, the path with minimum $W_{i:j}$. 

Roy and Post PNAS Early Edition ∣ 3 of 6
Dijkstra’s graph searching algorithm (26). The nodes of the network correspond to $C_i$ atoms and edges connecting all $C_i$ atoms belonging to residues within the nonbonded distance of 14 Å (Fig. 4). To capture the radial correlation behavior, an edge between $C_i$ and $C_j$ was assigned a value equal to $1 - C_{ij}(i,j)$. The path between any two nodes of the network, $l$ and $m$, was determined from the minimum value of the weight $W_{ij} = C_{ij}(i,j)$ (see Materials and Methods). The weight of a path is specified as one minus the product of the radial correlation coefficients for all edges of the path so that the path with the minimum weight is the optimal path in terms of high radial correlation and fewest number of edges. This path is named hereafter the “shortest path.” The shortest path was determined for all $C_i - C_i$ pairs of the pentamer, or approximately $8 \times 10^8$ paths. The second step of the method to identify the origin of the radial correlations was to determine which path segments had the highest betweenness (34). Betweenness of an edge is the probability with which the edge occurred in all shortest paths, or the number of occurrences of the edge relative to the total number of paths. From the distribution of betweenness values, shown in SI Appendix, a small number of edges are observed to lie at the far edge of the distribution for $T_{nowin}$ by having considerably higher betweenness than other edges. In contrast, the betweenness through these edges is substantially reduced in $T_{win}$.

The component of the radial correlation network with highest betweenness is visualized by mapping the edges with the largest betweenness values onto the structure in Fig. 5A. The edges having a betweenness with a 95% CI $>0.0005$ are shown with blue lines of thickness proportional to betweenness. It is noted that all edges within the pentamer asymmetric unit are treated independently so that the near fivefold symmetry apparent in the blue segments of Fig. 5A is evolved from microscopic sampling. High-betweenness edges from $T_{nowin}$ using the same cutoff value of 0.0005 are shown with red lines in Fig. 5B. The presence of WIN 52084 very clearly alters the observed patterns of betweenness; fewer red lines are observed and the overall thinner widths indicate smaller betweenness in general for $T_{win}$.

The patterns in Fig. 5 are striking. Each node of the network has close to a hundred or more edges so that the number of possible paths between two $C_i$ atoms is large and increases dramatically as the distance of separation increases. Thus, an edge with highest betweenness, and more importantly, sequential edges of high betweenness are definite outliers. Sequential edges of highest betweenness are a linked set of radially correlated $C_i$ atoms and suggest a mechanism for the long-range connectivity in Fig. 3. Numerous blue path segments are observed from $T_{nowin}$ (Fig. 5A) and found to be concentrated around the canyon and the fivefold axis. The set of blue lines encircling the fivefold axis corresponds to the N-terminal region of VP3, which forms a well-ordered $\beta$-annulus through interprotomer interactions. This annulus is absent from the pattern of highest betweenness paths calculated from $T_{win}$ when WIN is bound (Fig. 5B).

The paths connect radially correlated $C_i - C_i$ pairs separated by long distances (Fig. 3) and are a likely origin for these radial correlations. An example of one such path is given in Fig. 5C, which shows sequential edges connecting VP4 residue 4008 near the fivefold symmetry axis with the VP1 residue 1254 across the canyon. As such, this path provides a linkage between residues of $C_i - C_i$ pairs in group I Fig. 3C (green with blue spheres Fig. 3 E and F). Moreover, this path does not exist in the presence of WIN 52084.

The 10 edges with the highest betweenness from $T_{nowin}$ are listed in Table 1 by the residue numbers of the nodes delimiting the edge. These edges have a 95% CI of betweenness greater than 0.001. In contrast, their values from $T_{win}$ are approximately two orders of magnitude smaller. All the the residues in Table 1,
other than 2241, 2242, 3005, and 3006, are conserved or vary by one amino-acid type substitution among HRV-B serotypes, of which HRV 14 is a member. The underlined residues in Table 1 are similarly conserved in HRV-A. Of the 15 residues in Table 1, two (1199 and 1219) are in the WIN-binding pocket of the capsid, and an additional eight (1071, 1173, 1249, 1250, 3005, 3006, and 3007) are within 14 Å of WIN 52084 bound in the cavity.

**Conclusion**

The uncoating process, a critical part of enveloped picornaviral infection, requires large changes in the capsid conformation. WIN antiviral compounds disrupt uncoating and protect virus particles against thermal loss of VP4 (7) and thus are presumed to prevent these conformational changes. Several experimental observations of the native-state capsid detect large-scale reversible fluctuations that are likely relevant to this uncoating process (16, 19). Antibody epitope sites (15) and limited proteolysis (16) indicate that the N-terminal regions of VP1 and VP4 are reversibly exposed from inside the viral capsid and provide evidence that the mature intact viral capsid undergoes large spatial-scale dynamics. The uncoating process involves the externalization of these same protein regions; the small VP4 protein is lost in its entirety during infection (13, 20), and the N terminus of VP1 becomes exposed (14). The exposed peptides are thought to facilitate membrane permeability (14), and the N terminus of VP1 was shown to insert into membranes and considered important for tethering the virus particle during endocytosis (35). Further the WIN antiviral compound diminishes the large-scale motions of the capsid detected by proteolysis, which also links the breathing motion to uncoating. Together, these results argue that the reversible large-scale, breathing motions of the intact native virus particle are relevant to the uncoating process and that WIN compounds can be used to interrogate these dynamics.

Here, an analysis of MD simulations of HRV 14, without and with WIN 52084 bound, revealed large-scale concerted motions of the capsid that reasonably contribute to the initial events that lead to breathing and potentially uncoating. Radial motions of residues of HRV 14 without WIN separated by distances as large as 90 Å were found to be highly correlated (Fig. 2). Many of these radial correlations occur between residues within 35 Å of the fivefold axis, including the canyon region, and either a residue in the vicinity of the fivefold axis or a residue near a twofold axis that is nearby or across the pentameric unit. Sequential edges of the correlation paths with highest betweenness were found to connect these distant regions in HRV 14 (blue lines in Fig. 5A). These pathways are concentrated around the fivefold axis and the canyon region and extend to regions near the twofold symmetry axes regions where the RNA and VP4 are considered likely to exit the capsid (17–24).

Importantly, the presence of WIN 52084 dampens the radial correlations and destroys many of the highest-betweenness paths surrounding the fivefold axis (red lines, Fig. 5B), as would be predicted from the antiviral activity of inhibiting the uncoating process and blocking the breathing motion. The results therefore suggest that the long-range basis for the WIN antiviral activity is to interrupt local interactions that are necessary for the highest-betweenness correlation paths. This effect of WIN on the large-scale capsid dynamics could act in concert with an entropic effect on the virus capsid predicted previously (36) from MD simulation studies of the HRV capsid. That is, binding the hydrophobic WIN compound in the buried pocket of VP1 was predicted to entropically stabilize the capsid, a prediction later supported by experiment (37). With the reasoning that long-range concerted motions are lower entropy and that disruption of the correlation paths by WIN could lead to an increase in entropy, the possibility of an entropic effect upon WIN binding was examined using a quasi-harmonic analysis of the configurational entropy (38) of $T_{\text{bound}}$ and $T_{\text{win}}$. Although the error from such an estimate is large, we find that the HRV14 capsid with WIN bound is higher entropy (SI Appendix, Fig. G1). Thus, WIN binding could both stabilize the intact capsid by increasing the configurational entropy and also disrupt capsid interactions essential for the concerted dynamics leading to uncoating.

The paths of sequential edges with highest betweenness illustrated by the blue lines in Fig. 5A encompass a number of conserved residues (Table 1). Two of these, 1199 and 1219, are associated with antiviral activity and uncoating (39) as well as known to alter breathing (16). Together, the pattern of radial correlations and highest-betweenness paths composed of conserved residues, and the loss of this pattern when WIN is bound strongly suggest these paths are the origins of the radial correlations and part of the functional motion leading to uncoating.

Long-distance, correlated motions are a candidate mechanism for allostery in proteins. Nonetheless, such motions are often elusive in simulation studies (40); DCC are a function of the angle between centered vector variables (25) and insensitive to orthogonal displacements (41) (see SI Appendix A). Cross-correlation coefficients derived from dihedral angles are even less sensitive for detecting correlated motion (42). To circumvent this problem, we used radial cross-correlation coefficients, which defined numerous very long-distance correlations in capsid protein dynamics.

The insensitivity of displacement vector cross-correlation to identify long-distance effects motivated the development of network-based approaches. Networks built from protein structures alone provide topological analysis that can give information on structural stability and binding sites (43). Another approach that has been used to investigate allostery and long-distance communication is a network using low-frequency fluctuations from elastic network normal mode analysis (ENMA) (44, 45). However ENMA is most useful for such analysis of binding effects in the presence of large structural changes (46). This is not the case with WIN binding HRV 14; the root mean square deviation between $C_a$ atoms of average structures from $T_{\text{bound}}$ and $T_{\text{win}}$ is only 1.32 Å. Indeed, our efforts to use elastic network normal mode calculations were not fruitful and found no significant difference in low-frequency modes of HRV and HRV-WIN 52084. Finally, a network derived from interaction energy correlations was reported to identify coupling over long-distance in signaling proteins (47, 48). Correlations in energy reflect interactions that reasonably lead to concerted motion but do not provide any direct information on that motion. The radial correlation network approach reported here offers an alternative method to query long-distance correlation paths from concerted motions.

**Materials and Methods**

Initial coordinates were from crystallography (PDB ID code 4RHV for HRV 14 and 1RUD for HRV 14-WIN). Modeling of missing coordinates and protocol of MD simulation are detailed in SI Appendix.

Each $C_a$ atom was represented by a node in the network. WIN 52084 was represented by the first carbon atom on the right of the benzene ring in Fig. 1C. Two nodes are connected by an edge if any atom of the corresponding residues are within 14 Å in at least one frame of the 10 trajectories. An edge is assigned a weight calculated from the radial correlation of the $C_a$ atoms, $C_R(i, j)$,

$$C_R(i, j) = \frac{|e_R(i, j)|}{\sqrt{(R_i - R_j)^2}}$$

where

$$e_R(i, j) = (R_i - R_j)(R_i - R_j).$$

where, $R_i$ is the distance of the $C_a$ atom of the $i$th residue from the origin at an instance and $\langle ... \rangle$ denotes the ensemble average. The weight of an edge connecting two neighboring nodes $i$ and $j$, $W_{ij}$, is $1 - C_R(i, j)$.
A path in the network between two $C_i$ atoms, or nodes, consists of one or more edges (Fig. 4). If nodes $i$ and $j$ are not neighbors and there exists a path $a_1 ... a_d$, where $a_i,a_d$ are nodes between $i$ and $j$ and $n \geq 1$, then the weight of the path $a_1 ... a_d$ is

$$W'(a_1 ... a_d) \equiv 1 - C_R(i, a_1) \cdot \ldots \cdot C_R(a_{n-1}, a_n) \cdot C_R(a_n, j). \quad [2]$$

The shortest path or path with least weight, between all nonneighbor pairs of nodes is found using an algorithm developed by Dijkstra (26) with edge weights modified to include the radial correlation defined by Eq. 2. The betweenness of an edge is the fraction of all possible shortest-paths that include the edge.

ACKNOWLEDGMENTS. We thank the Rosen Center for Advanced Computing at Purdue University for providing computing resources. We also thank Professor Jayanta Ghosh of the statistics department at Purdue University and his student Jyotishka Dutta for insightful discussion about graph theory and bootstrapping. This work was supported by National Institutes of Health Grant AI039639.

Supporting information: Long-distance correlations of rhinovirus capsid dynamics contribute to uncoating and antiviral activity.

Amitava Roy and Carol Beth Post
Department of Medicinal Chemistry and Molecular Pharmacology
Purdue University, West Lafayette, USA

Appendix A: Cross-correlation

Angular dependence of displacement vector cross-correlation

Figure A1: Two particles, \(i\) and \(j\), are fluctuating in a two dimensional space. \(\mathbf{R}_i\) and \(\mathbf{R}_j\) are their position vectors at an instance. \(\mathbf{R}_i\) makes an angle \(\theta_i\) with \(X\) axis and \(\mathbf{R}_j\) makes an angle \(\theta_{ij}\) with \(\mathbf{R}_i\).

\[
\mathbf{R}_i = X_i \hat{x} + Y_i \hat{y} = R_i \cos \theta_i \hat{x} + R_i \sin \theta_i \hat{y} \\
similarly, \quad \mathbf{R}_j = R_j \cos (\theta_{ij} + \theta_i) \hat{x} + R_j \sin (\theta_{ij} + \theta_i) \hat{y}
\]

where \(\hat{x}\) and \(\hat{y}\) are unit vectors along \(X\) and \(Y\) axes respectively and \(X_i\) and \(Y_i\) are projections of \(\mathbf{R}_i\) along \(X\) and \(Y\) axes respectively. Similarly \(X_j\) and \(Y_j\) are projections of \(\mathbf{R}_j\) along \(X\) and \(Y\) axes respectively. Then covariance of displacement \(c(i, j)\) between \(i\) and \(j\) is
\[ c(i,j) = \left( (R_i - \langle R_i \rangle) \cdot (R_j - \langle R_j \rangle) \right) \]
\[ = (X_i X_j + Y_i Y_j - \langle X_i \rangle \langle X_j \rangle - \langle Y_i \rangle \langle Y_j \rangle) \]
\[ = \langle R_i \cos \theta_i R_j \cos (\theta_{ij} + \theta_i) \rangle + \langle R_i \sin \theta_i R_j \sin (\theta_{ij} + \theta_i) \rangle - \langle R_i \cos \theta_i \rangle \langle R_j \cos (\theta_{ij} + \theta_i) \rangle - \langle R_i \sin \theta_i \rangle \langle R_j \sin (\theta_{ij} + \theta_i) \rangle \]
\]
\[ (a2) \]

and cross-correlation coefficient of displacement, \( C(i,j) \), is

\[ C(i,j) = \frac{c(i,j)}{\sqrt{\langle (R_i - \langle R_i \rangle)^2 \rangle \langle (R_j - \langle R_j \rangle)^2 \rangle}} \]
\[ (a3) \]

where \( \langle .. \rangle \) denotes ensemble average. If \( R_i \) and \( R_j \) are statistically independent of \( \theta_i \) and \( \theta_{ij} \) then we can rewrite eqn a2 as

\[ c(i,j) = \langle R_i R_j \rangle \langle \cos \theta_i \cos \theta_{ij} \rangle - \langle R_i \rangle \langle R_j \rangle \left\{ \langle \cos \theta_i \rangle \langle \cos \theta_{ij} \rangle - \langle \sin \theta_i \rangle \langle \sin \theta_{ij} \rangle \right\} \]
\[ (a4) \]

From eqn a2 it is clear that \( c(i,j) \) vanishes if we fix \( \theta_{ij} \) to 90°. Furthermore if \( \theta_{ij} \) and \( \theta_i \) are statistically independent then

\[ c(i,j) = \langle R_i R_j \rangle \langle \cos \theta_i \rangle - \langle R_i \rangle \langle R_j \rangle \left\{ \langle \cos \theta_i \rangle \langle \cos \theta_{ij} \rangle - \langle \sin \theta_i \rangle \langle \sin \theta_{ij} \rangle \right\} + \langle \sin \theta_i \rangle \langle \sin \theta_{ij} \rangle \]
\[ (a5) \]

Covariance, hence hence cross-correlation coefficient, of displacement depends on \( \theta_{ij} \), the angle between two displacement vectors.

**Radial correlation**

Cross-correlation of displacement in Cartesian coordinate (DCC) is most sensitive when \( \theta_{ij} \) is around 0° or 180° and not suitable to describe correlation in fluctuation of a spherically symmetric assembly of particles. In a spherically symmetric assembly of particles \( C(i,j) \) will fail to capture correlation between perpendicular fluctuations and overemphasize correlation between linear fluctuations. To overcome insensitivity of \( C(i,j) \) to radial fluctuation we define radial correlation, \( C_R(i,j) \), as
Figure A2: Distribution of non-neighbor Cα pairs with 95% confidence interval (CI) of radial correlation coefficient, $C_R(i,j)$, >0.6 as a function of intra-pair Cα distance in $T_{\text{nowin}}$ (solid). Distribution of same pairs with cross correlation of displacement > 0.6 for $T_{\text{nowin}}$ is shown with dotted line. Cross-correlation of displacement fails to catch long-distance correlation.

\[
C_R(i,j) = \frac{|c_R(i,j)|}{\sqrt{\langle(R_i - \langle R_i \rangle)^2\rangle\langle(R_j - \langle R_j \rangle)^2\rangle}}
\]

where

\[
c_R(i,j) = \langle(R_i - \langle R_i \rangle)(R_j - \langle R_j \rangle)\rangle = \langle R_i R_j \rangle - \langle R_i \rangle\langle R_j \rangle.
\]

$a6$

$C_R(i,j)$ is most sensitive to radial as well as linear fluctuations but insensitive to tangential fluctuations. Distribution of DCCs of displacement and $C_R$s of non-neighboring Cα pairs in $T_{\text{nowin}}$, 20 ns trajectories each of HRV 14, against intra pair distance are shown in Fig A2 in dotted and solid lines respectively. Non-neighboring Cα pairs are from residues outside nonbond distance of 14 Å of each other. DCCs fail to capture long distance correlations.

Appendix B: Homology modeling of missing HRV 14 coordinates

The crystal structure of human rhinovirus (HRV) 14, PDBID: 4RHV [1], has missing coordinates for residues 1001 to 1016 of VP1, 2001 to 2007 of VP2 and 4001 to 4028 of VP4. The missing coordinates were modeled from HRV 16 (PDBID 1AYM [2]) and poliovirus (PDBID 1HXS [3]). VP1 residues were modeled from HRV 16 which has 43% sequence identity and 60% sequence similarity with VP1 of HRV 14. Secondary structure prediction of the missing coordinates of VP1 of HRV 14 matches exactly with the determined secondary structure of HRV 16. The crystal structure of polio virus, has no missing coordinates in VP4. It has 57% sequence identity and 75% sequence similarity with VP4 of HRV 14. Coordinates of residue 4001 to 4029 of VP4 of 4RHV were transferred from poliovirus. Missing coordinates of VP2 were not modeled as we could not find any template with known coordinates for the modeling.

The energy of the modeled residues was minimized while maintaining the known 4RHV crystallographic coordinates fixed. While still keeping the known atomic positions fixed, in vacuum NVE molecular dynamics (MD) simulations were performed to raise the temperature of the modeled regions over a 490 ps period from 100 K to 5000 K, the system was annealed at 5000 K for a 50 ps period, and then cooled to 300 K over a 980 ps period. Twenty modeled structures, ten each for viral capsid with and without drug, generated from twenty independent annealing and subsequent cooling MD calculations were used as initial coordinates for MD trajectory calculations.

Appendix C: Molecular dynamics
Starting from different annealed structures from modeling, ten 3 ns MD trajectories were calculated each for solvated viral capsid of HRV 14 without and with WIN 52084. The last 2 ns of the ten trajectories for HRV 14 and HRV 14-WIN 52084, referred as $T_{\text{nowin}}$ and $T_{\text{win}}$ respectively, were used for analysis. All MD calculations were performed using the CHARMM program [4] with constant volume and energy (NVE ensemble). A force switching function [5] was used to smoothly truncate electrostatic and van der Waals non-bonded forces with a cutoff of 14.0 Å. The covalent bonds to hydrogen atoms were constrained by the SHAKE algorithm. The equations of motion were integrated using the Verlet leap-frog algorithm with a time step of 1 fs. The unit cell of simulation was a pentamer containing five protomers solvated by an explicit shell of water constrained by outside and inside spherical boundaries. Water molecules were constrained within the outer and inner radii by applying spherical quadratic potentials referenced 170 Å and 85 Å, respectively. The spherical quadratic potential was set up with a well depth of -0.25 kcal/mol at 1 Å from the reference distance followed by a smoothly rising repulsion. Solvated pentamer contained $\approx 140,000$ atoms.

Neighboring protomers were generated using icosahedral boundary conditions implemented with the IMAGES facility within CHARMM. The detail implementation of icosahedral boundary condition is explained, in detail elsewhere [6]. Image atoms within 16.0 Å of a primary atom were included in the non-bonded pair list. Updates of image and non-bonded lists were made heuristically. In each of the MD calculations capsid was equilibrated for 1 ns before collecting data for another 2 ns period. During the 2 ns time period root mean square deviation (RMSD) values of the crystallographic coordinates w.r.t. the initial structures remained around 2 Å. RMSD values of trajectories of HRV 14 are shown in Fig C1.

**Figure C1:** $C_\alpha$ RMSD of coordinates of ten trajectories in $T_{\text{nowin}}$ with respect to the initial structures of each trajectory. For last 2 ns time period RMSD value remained around 2 Å.

**Appendix D: Bootstrap**

Bootstrap is a statistical procedure to learn about sampling variation using one set of observations [7]. Given ten independent 2 ns long trajectories, to find the variability in an observable $\mu$, such as $C_R(i,j)$ or betweenness, a new set of ten 2 ns long trajectories is made by randomly choosing ten trajectories from the original pool, where a single trajectory is allowed to appear more than once. A new estimate $\hat{\mu}^*$ is calculated from the mean value of the new set. This procedure is repeated $n$ times to generate the bootstrap distribution containing $n$ values of $\hat{\mu}^*$. The bootstrap distribution is then used to calculate the 95% confidence interval (CI) of the true mean $\mu$ from the mean $\hat{\mu}$ of the original ten trajectories and $n$ values of $\hat{\mu}^*$ from bootstrap distribution.

The $\rho\%$ CI of $\mu$, is then given by
\[ 2\hat{\mu} - \hat{\mu}_p(n+1)(1-\alpha/2) \leq \mu \leq 2\hat{\mu} - \hat{\mu}_p(n+1)\alpha/2 \]  
\[ \text{(D1)} \]

where \( \alpha \) is \( (1 - \rho/100) \) and \( \hat{\mu}_p \) is the \( p \)th ordered value of the bootstrap distribution \( \hat{\mu}_1 \leq \hat{\mu}_2 \leq ... \leq \hat{\mu}_n \) \([7]\). For 95\% CI, \( \alpha \) is \( (1 - 95/100) = 0.05 \). With \( n \) being 399 in our case \( \hat{\mu}_p(n+1)(1-\alpha/2) \) and \( \hat{\mu}_p(n+1)\alpha/2 \) are \( \hat{\mu}_{390} \) and \( \hat{\mu}_{10} \).

**Appendix E: Distribution of radial correlation coefficients and betweenness**

**Figure E1:** Distribution of \( C_R \) values for all possible \( C_\alpha \) pairs of \( T_{\text{nowin}} \) and \( T_{\text{win}} \) are plotted with blue and red curve respectively. \( T_{\text{win}} \) has lesser number of pairs with high value of \( C_R \) than \( T_{\text{nowin}} \). Values \( C_R > 0.6 \) are clear outliers. Inset shows a zoomed in region of the plot.

**Figure E2:** Number of edges as a function of betweenness in \( T_{\text{nowin}} \) and \( T_{\text{win}} \) are plotted in red and blue curve respectively. Betweenness of an edge is the number of occurrences of the edge relative to the total number of paths. Inset shows the same distribution with a magnified Y-axis. Edges with betweenness >0.003 present in \( T_{\text{nowin}} \) are completely missing in \( T_{\text{win}} \).

In this article we investigated radial correlation, \( C_R \), between \( C_\alpha \) pairs. Distribution of \( C_R \) values for all possible \( C_\alpha \) pairs of \( T_{\text{nowin}} \) and \( T_{\text{win}} \) are plotted in Fig E1 with blue and red curve respectively. \( T_{\text{win}} \) has lesser number of pairs with \( C_R \geq 0.4 \) than \( T_{\text{nowin}} \) indicating overall loss of correlation with introduction of WIN52084. Values \( C_R > 0.6 \) are clear outliers in the distribution and indicate unusual highly correlated radial motion. Hence we concentrated our analysis on \( C_R > 0.6 \) for non-neighboring \( C_\alpha \) pairs.

From a network, where the nodes are \( C_\alpha \) atoms with edges between \( i, j \) neighboring nodes are weighted as \( 1 - C_R(i, j) \), we calculate the betweenness of each edge. Betweenness is the frequency an edge occurs in all shortest paths between two non-neighboring \( C_\alpha \)s. The number of edges as a function of betweenness in \( T_{\text{nowin}} \) and \( T_{\text{win}} \) are plotted in Fig E2 with blue and red curve respectively. Edges with betweenness >0.003 present in \( T_{\text{nowin}} \) are completely missing in \( T_{\text{win}} \).
Figure F1: Average $C_R$, calculated from a window of varying width from the end of $T_{nowin}$ trajectories for $C_\alpha$ pairs with final $C_R$ value between 0.6 and 0.7. Neighboring and non-neighboring pairs are plotted with blue and red curve respectively. The value of $C_R$ is converged for window for widths $> 450$ ps.

Appendix F: Convergence of radial correlation coefficients

To investigate the convergence of values $C_R > 0.6$, we grouped all $C_\alpha$ pairs in to four groups, group a to d, according to whether their $C_R$ values calculated from $T_{nowin}$ is between interval $(0.9 - 1.0)$, $(0.8 - 0.9)$, $(0.7 - 0.8)$ or $(0.6 - 0.7)$ respectively.

To determine a window size over which the $C_R$ stabilizes, average value of $C_R$ for windows of different width from the end of 10 HRV 14 trajectories for neighboring and non-neighboring $C_\alpha$ pairs of group d are plotted in blue and red curve respectively in Fig F1. The initial increase of $C_R$ in Fig F1 indicates statistical dependence of the data for window width $< 450$ ps. The $C_R$ value is stabilizes for window width $> 450$ ps.

With a window of 950 ps we calculated values of $C_R$ from ten trajectories of HRV 14 for a total average time of 9.5 ns for neighboring and non-neighboring $C_\alpha$ pairs. Starting from the initial step of the trajectory, the 950 ps window was shifted by 10 ps to calculate $C_R$. $\Delta C_R$, changes in average $C_R$ values, is plotted against the initial time of the moving window from 1 ps to 2050 ps, for group a to d in Fig F2a to Fig F2d respectively. $\Delta C_R$ for neighboring and non-neighboring $C_\alpha$ pairs are plotted with blue and red curve respectively. Convergence is assumed when $\Delta C_R$ values fluctuates around zero.

Non-neighboring residues are correlated via interactions through intermediate residues and as expected $C_R$ between non-neighboring pairs converges slower than $C_R$ between neighboring pairs. Also smaller $C_R$ values require longer averaging time to separate the correlated fluctuation from random thermal fluctuation. $\Delta C_R$ for non-neighboring pairs of group d reaches zero around 1000 ps. We conclude the convergence of $C_R$ values of interest has been achieved with 1 ns data from 10 HRV trajectories.
Appendix G: Entropy from vibrational modes

Vibrational frequency spectrum of a protein can be calculated from eigenvalues of mass weighted covariance matrix of the atomic Cartesian coordinates, \( \mathbf{A} \), of the protein calculated from MD simulations.

\[
\mathbf{AV} = \lambda \mathbf{V}
\]

where matrix \( \mathbf{V} \) is defined as the eigenvector matrix of \( \mathbf{A} \) and \( \lambda = \{\lambda_\alpha\} \) is a diagonal matrix of the eigenvalues whose elements represent mean-square fluctuation (MSF) along the direction of the corresponding eigenvectors. Matrix \( \mathbf{A} \) is a \( 3N \times 3N \) dimensional matrix where \( N \) is number of atoms present in the protein. An element \( a_{ij} \) of matrix \( \mathbf{A} \) is defined as

\[
a_{ij} = \langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \rangle
\]

where

\[
q_i(t) = \begin{bmatrix}
(m_n)^{1/2}x_n(t) \\
(m_n)^{1/2}y_n(t) \\
(m_n)^{1/2}z_n(t)
\end{bmatrix}
\]

The eigenvalues \( \lambda_\alpha \)s are related to the frequency of vibrational modes \( \nu_\alpha \)s as

\[
\lambda_\alpha = \frac{k_BT}{4\pi^2 \nu_\alpha^2}
\]

The contribution of a vibrational mode, \( S^{\text{vib}}_\alpha \) to the entropy of the protein is

\[
S^{\text{vib}}_\alpha = -R \ln(1 - e^{-\frac{h\nu_\alpha}{k_BT}}) + \frac{N_A h \nu_\alpha}{T(e^{\frac{h\nu_\alpha}{k_BT}} - 1)}
\]

where \( h \) is Planck’s constant, \( N_A \) is Avogadro’s number, \( R \) is gas constant and \( T \) is temperature of the protein. Vibrational entropy of a protein is then \( S^{\text{vib}} = \sum_\alpha S^{\text{vib}}_\alpha \). See Ref.[9] for short review on the topic.

To estimate change of entropy, upon binding of WIN to the capsid, from vibrations of tertiary structures we included only the Cartesian coordinates of \( C_\alpha \) atom of protein residues and one central carbon atom of WIN compounds in the covariance matrix \( \mathbf{A} \). We found change in entropy upon binding of WIN, \( \Delta S^{\text{vib}} = S^{\text{vib}}_{\text{win}} - S^{\text{vib}}_{\text{nowin}} \), is 75.64 cal/K/mol. Fig G1 shows contribution vibrational modes to \( S^{\text{vib}} \) as a function of frequency \( \nu \) of the modes. Fig G2 shows distribution of vibrational mode frequencies calculated from covariance matrix defined from the Cartesian coordinates of \( C_\alpha \) atoms. Most of the vibrational frequencies are less than 20 ps\(^{-1}\) which is a lower frequency than vibrational modes estimated for secondary structures of proteins [10]. Accordingly, the modes from HRV14 are more collective and related to tertiary structures of the capsid.

Experimental results show WIN like drug stabilizes poliovirus capsid, structurally similar to rhinovirus capsid, entropically by \( \sim \) 320 cal/K/mol[11]. From fluctuations of \( C_\alpha \) atoms we find
Figure G1: Vibrational entropies $S_{\text{vib \, \text{nowin}}}$ (blue), calculated from fluctuations of $C_\alpha$ atoms in $T_{\text{nowin}}$, and $S_{\text{vib \, \text{win}}}$ (red), calculated similarly from $T_{\text{win}}$, are plotted as a function of frequency of fluctuations. Difference in vibrational entropy, $\Delta S_{\text{vib}} = S_{\text{vib \, \text{win}}} - S_{\text{vib \, \text{nowin}}}$ is plotted as a function of frequency in black. Final value of $\Delta S_{\text{vib}}$ is 75.64 cal/K/mol.

Figure G2: Number of vibrational modes as a function of frequency of the modes calculated from fluctuations of $C_\alpha$ atoms in $T_{\text{nowin}}$ and $T_{\text{win}}$ are plotted in blue and red curve respectively. Most of the modes have frequencies less than 20 ps$^{-1}$ indicating that the observed difference in vibrational entropy is coming from vibrations of tertiary structures of capsid proteins.

$\Delta S_{\text{vib}}$ has a value of same order of magnitude. It leads us to speculate that vibrational modes contributing to $\Delta S_{\text{vib}}$ form part of the reaction coordinates of conformational change leading to the “breathing” motions.

The quasi-harmonic normal modes are also a useful approach to examine concerted motion, and therefore could be considered as a means to determine differences in long-distance correlated motions of HRV14 and HRV14-WIN capsids. While this information must be present in A, it exists in some combination of modes that is not readily apparent. How to extract a meaningful difference from the large set of normal modes for HRV14 and HRV14-WIN is unclear.

References


