Supporting Information

Alexander et al. 10.1073/pnas.1308846110

SI Materials and Methods

Reagents. Urea and guanidinium hydrochloride (GdmCl) were purchased from MP Biomedicals. All other reagents were of analytical reagent grade and purchased from Sigma Chemical Co. or ThermoFisher Scientific. Core protein of hepatitis B virus comprising a structured capsid-forming region (residues 1–149) ($HBc_{1–149}$) mutants were generated using Quikchange site-directed mutagenesis (Strategene). Protein concentrations were determined using UV absorption measurements and the molar extinction coefficients calculated from the protein sequences of $HBc_{1–149}$ variants (1).

Purification of Capsid Assembly-Defective $HBc_{1–149}$ Variants. $HBc_{1–149}$ F23A and L42A were expressed and purified as per the WT, except the first size-exclusion column was replaced with an anion-exchange step (2). The resuspended ammonium sulfate pellet of $HBc_{1–149}$ was applied to a POROS HQ20 column (Applied Biosystems) before washing the column with 10 column volumes of 25 mM Tris·HCl buffer (pH 9.0) and 2 mM DTT. $HBc_{1–149}$ was eluted using a 0- to 550-mM sodium chloride gradient. $HBc_{1–149}$ fractions were identified by SDS/PAGE, pooled, dialyzed against 30 mM Tris·HCl buffer (pH 7.5), and purified using a Superdex 200 gel filtration column (GE Healthcare) as per WT $HBc_{1–149}$ (2). A perdeuterated ^{15}N-labeled $HBc_{1–149}$ L42A sample for NMR studies (see below) was expressed in minimal media as described previously (2).

$HBc_{1–149}$ Y132A with a hexahistidine tag was expressed as per WT $HBc_{1–149}$ (*Materials and Methods*). Cells were lysed by sonication in 50 mM Tris·HCl buffer (pH 9.0) containing 2 M urea, 250 mM sodium chloride, and 1 mM DTT. After clarification by centrifugation, $HBc_{1–149}$ Y132A was purified using an immobilized metal-affinity column equilibrated with lysis buffer. Tagged $HBc_{1–149}$ Y132A was eluted with lysis buffer containing 200 mM imidazole. EDTA was added to a final concentration of 10 mM in eluted protein fractions before dialyzing against 25 mM Tris·HCl buffer (pH 9.0) containing 2 mM DTT. Dialyzed protein was purified using a POROS HQ anion-exchange column (Applied Biosystems) and a 0- to 550-mM sodium chloride gradient. The hexahistidine tag was cleaved using an overnight digestion with tobacco etch virus (TEV) protease. The TEV protease and cleaved tag were removed with a second metal-affinity chromatography step. $HBc_{1–149}$ Y132A-containing fractions were identified by SDS/PAGE and pooled, and 10 mM EDTA was added to remove any metals that might have copurified with $HBc_{1–149}$ Y132A. For storage, pure protein was dialyzed into 25 mM Tris·HCl buffer (pH 9.0) containing 2 mM DTT. Electrospray ionization (ESI) ion mobility spectrometry (IMS) MS showed this material had an observed mass (34,946 ± 1 Da) that exactly matched the theoretical mass and contained no metal adducts (Fig. S3).

Protein Stability Measurements. A total of 51-point GdmCl titrations were performed in 30 mM sodium bicarbonate buffer (pH 9.5) and 2 mM DTT using $HBc_{1–149}$ at a final concentration of 7.5 μM. To ensure complete equilibration, solutions were incubated at 20 °C in a water bath for 5–18 h before experiments. For far-UV CD spectroscopy, the ellipticity at a selected wavelength was acquired over 1 min and averaged (using one measurement per second, a 0.5-nm bandwidth, and a 1-mm pathlength cuvette). For fluorescence emission spectroscopy, intrinsic fluorescence was excited at 280 nm (2.5-nm slit width and 4-mm pathlength), with an emission spectrum between 300 and 400 nm recorded twice and averaged (2.5-nm slit width, 60-nm/min scan rate, and 10-mm pathlength). Thereafter, the averaged intensity at a given emission wavelength was plotted vs. denaturant concentration.

Data Analysis and Curve Fitting. Equilibrium chemical denaturation data were fitted to the following equation using nonlinear regression:

$$f = \frac{\propto_N \beta_N \ [GdmCl] \ (1-\mathrm{D})}{1+K_1} + \frac{\propto_I \beta_I \ [GdmCl] \ K_1(1-\mathrm{D})}{1+K_1} + (\propto_D \beta_D \ [GdmCl] \ \mathrm{D}),$$

where f is the observed signal; α_N, α_I, and α_D represent the y-axis intercepts of native, intermediate, and denatured baselines, respectively; β_N, β_I, and β_D represent the native, intermediate, and denatured baselines slopes, respectively; [*GdmCl*] represents the GdmCl concentration; K_1 and K_2 represent the equilibrium constants for native to intermediate and intermediate to denatured transitions, respectively, where

$$K_x = \exp\left(\frac{-\Delta G_n^{H_2O}}{R \ T}\right);$$

and D represents the concentration of the denatured state as defined below:

$$\mathrm{D} = \frac{-K_1 K_2 + \sqrt{(K_1 K_2)^2 + 8 K_1 K_2 [P_T] (1+K_1)}}{4 [P_T] (1+K_1)},$$

where $[P_T]$ represents total protein concentration. The free-fitted m values for all WT-like mutants were averaged, and the data were refitted using the averaged m values. For a minority of mutants, the m values of the individual transitions deviated from the average, and these titrations were fitted using freely floating m values or with the total m_{D-N} value constrained to the WT value (but where m_{I-N} and m_{D-I} were unconstrained for fitting). These are standard fitting procedures for such data, and the validity of this approach has been discussed elsewhere at length (3, 4).

For calculating the free energy change caused by mutations, the below formalisms were used:

$$\Delta\Delta G_{X-Y}^{H_2O} = m_{ave}\left([GdmCl]_{50\%}^{\mathrm{WT}} - [GdmCl]_{50\%}^{\mathrm{Mut}}\right)$$

or

$$\Delta\Delta G_{X-Y}^{H_2O} = m_{ave}[GdmCl]_{50\%}^{\mathrm{WT}} - m_{Mut}[GdmCl]_{50\%}^{\mathrm{Mut}},$$

where m_{ave} is an averaged m value calculated from all WT-like mutants and m_{Mut} is the m value calculated from the free fit of the equilibrium curve for that mutant.

ESI-IMS-MS. All mass spectra and IMS-MS data were collected using a Synapt HDMS orthogonal acceleration quadrupole-traveling wave TOF mass spectrometer (Micromass Ltd.). All samples were analyzed using nano-ESI with platinum/gold-plated borosilicate capillaries fabricated in-house using a P-97 micropipette puller (Sutter Instrument Company) and a Polaron SC7620 sputter coater (Quorum Technologies Ltd.). Initial experimental parameters were a capillary voltage of 1.8 kV, cone

voltage of 20 V, trap voltage of 6 V, and transfer voltage of 4 V. The source pressure was maintained at 2.8 mbar, and we used a trap flow rate of 0.1 mL/min. $HBc_{1–149}$ WT, C61A, and Y132A (typically 8 μM) were electrosprayed from 20 mM ammonium acetate (pH 7.5) buffers. All *m/z* spectra were calibrated using cluster ions generated from a separate infusion of cesium iodide solution (40 mg/mL), and data were analyzed using MassLynx 4.1 software (Micromass Ltd.).

To construct gas-phase unfolding profiles, the 10+ charge states of WT, C61A, and Y132A $HBc_{1–149}$ were mass-selected using the quadrupole analyzer and subjected to increasing collision voltage in the trap region of the Synapt HDMS mass spectrometer. Subsequent arrival time data were collected in the traveling-wave ion mobility device. Collision-induced unfolding was monitored by observing changes in arrival time distribution, with increasing collision energy for the ions traversing the traveling-wave device. Collision voltages of 6–54 V were used (in 2-V increments).

After data collection, the arrival time distributions were analyzed to monitor the population of unfolding states as a function of collision energy. This was achieved by fitting Gaussian curves to the arrival time distribution at each energy to determine the minimum number of components that would, when summed, fit to the experimental arrival time distribution (5, 6). The process optimizes the mean, height, and SD of each of the Gaussian components until a solution is reached in which the sum of all components matches the experimental arrival time distribution with the lowest least-squares error. The components (unfolding states) were then correlated between collision energies and identified as N, I_1, I_2, or D. The area under the curve for each component was calculated by integration of its Gaussian profile, and this was plotted for each component as a function of center-of-mass collision energy, E_{cent} (electronvolt), calculated using the equation below, to produce an unfolding trace:

$$E_{cent} = zV_{MS} \times \left(\frac{M_{Argon}}{M_{Argon} + M_{ion}}\right),$$

where z is the charge state of the ion, V_{MS} is the trap collision voltage used, and M_{Argon} and M_{ion} are the molecular masses of the collision gas (argon) and the ion under scrutiny, respectively.

Isothermal Titration Calorimetry. Isothermal titration calorimetry (ITC) measurements were performed using a MicroCal ITC_{200} calorimeter (GE Healthcare). Before measurements, samples were dialyzed extensively against 30 mM carbonate buffer (pH 9.5). During the experiment, 800 μM Oct2 peptide (MHRSLLGRMKGA) (2) was titrated into 40 μM $HBc_{1–149}$ dimer using 25 individual injections of 1.5 μL each (with 160 s between injections and a syringe stirring speed of 1,000 rpm). Data were analyzed using Origin 7.0 (OriginLab Corp.).

X-Ray Crystallography. $HBc_{1–149}$ Y132A was concentrated to ∼17 mg/mL in 20 mM Tris·HCl buffer (pH 9.0) and 2 mM DTT and was crystallized at 20 °C in 100 mM citrate (pH 5.0), 15% (vol/vol) isopropanol, and 1% (wt/vol) PEG 10,000. Two hundred-micron crystals grew within 4 d. Crystals were cryocooled in mother liquor using 30% (vol/vol) glycerol as the cryoprotectant. Data were collected to 3 Å at beamline PXIII (Swiss Light Source). Data were processed using XDS and Scala (7, 8). Initial phases were obtained by molecular replacement using the program Phaser and chains E and F from Protein Data Bank (PDB) ID code 3KXS (9). The final structure (PDB ID code 4BMG) was iteratively built and refined using Coot and Phenix (10, 11) and had good geometry and fit to the electron density with final R_{work} and R_{free} values of 25.6% and 31.7%, respectively. R_{work} and R_{free} were calculated from the following formalism: $R = (\sum ||F_{obs}| - |F_{calc}||)/\sum |F_{obs}|$. 5% of the reflections were excluded from refinement and used to calculate R_{free}, with the remaining 95% used to calculate R_{work}.

NMR Spectroscopy. ^{15}N-^{1}H transverse relaxation-optimized spectroscopy (TROSY) NMR spectra (12) were acquired at 293 K using a Bruker Avance 700-MHz NMR spectrometer equipped with a triple-resonance inverse detection cryoprobe (Bruker BioSpin Corp.). WT and L42A $HBc_{1–149}$ variants were used at a concentration of ∼150 μM dimer in 30 mM sodium carbonate buffer (pH 9.5) [containing 2 mM DTT, 5% (vol/vol) deuterium oxide, and 0.1 μM dimethyl-4-silapentane-1-sulfonic acid as the referencing standard]. In this analysis, we conservatively binned L42A-induced chemical shift perturbations (CSPs) of amide resonances [relative to WT $HBc_{1–149}$, Biological Magnetic Resonance Bank entry 15969 (2)] into two categories: (*i*) amide resonances that lost all signal intensity due to suspected exchange broadening (red residues in Fig. S6*C*) and (*ii*) amide resonances with very large CSPs (0.07–0.35 ppm in the ^{1}H dimension or >1 ppm in the ^{15}N dimension; blue residues in Fig. S6*C*).

1. Gill SC, von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 182(2):319–326.
2. Freund SM, Johnson CM, Jaulent AM, Ferguson N (2008) Moving towards high-resolution descriptions of the molecular interactions and structural rearrangements of the human hepatitis B core protein. *J Mol Biol* 384(5):1301–1313.
3. Serrano L, Kellis JT, Jr., Cann P, Matouschek A, Fersht AR (1992) The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability. *J Mol Biol* 224(3):783–804.
4. Ferguson N, Fersht AR (2011) Phi-value analysis of protein folding transition states. *Protein Engineering Handbook*, eds Lutz S, Bornscheuer UT (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany), pp 1–47.
5. Hopper JT, Oldham NJ (2009) Collision induced unfolding of protein ions in the gas phase studied by ion mobility-mass spectrometry: The effect of ligand binding on conformational stability. *J Am Soc Mass Spectrom* 20(10):1851–1858.
6. Freeke J, Bush MF, Robinson CV, Ruotolo BT (2012) Gas-phase protein assemblies: Unfolding landscapes and preserving native-like structures using noncovalent adducts. *Chem Phys Lett* 524:1–9.
7. Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125–132.
8. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62(Pt 1):72–82.
9. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Cryst* 40(Pt 4):658–674.
10. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
11. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.
12. Pervushin K, Braun D, Fernández C, Wüthrich K (2000) [15N,1H]/[13C,1H]-TROSY for simultaneous detection of backbone 15N-1H, aromatic 13C-1H and side-chain 15N-1H2 correlations in large proteins. *J Biomol NMR* 17(3):195–202.

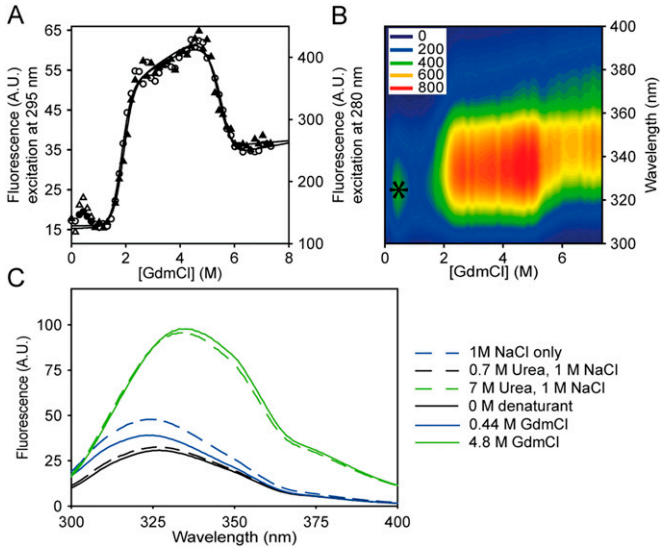

Fig. S1. Equilibrium denaturation of $HBc_{1–149}$ monitored by fluorescence emission spectroscopy (FES). (*A*) Equilibrium denaturation curves of WT $HBc_{1–149}$ were essentially identical when probed using excitation wavelengths of 280 nm (which excite tyrosine and tryptophan fluorophores; ○) and 295 nm (which excite solely tryptophan fluorophores; ▲). (*B*) Contour plot of $HBc_{1–149}$ fluorescence emission intensity vs. wavelength and GdmCl concentration. Different emission intensities are represented by different colors as shown. The signal associated with GdmCl-induced capsid formation is indicated with an asterisk. (*C*) FES spectrum of WT $HBc_{1–149}$ in 0.44 M GdmCl (blue solid line) is very similar to that of assembled capsids in 1 M NaCl (blue dashed line). However, the FES spectrum in 1 M salt with 0.7 M urea (black dashed line) closely resembles the spectrum of the $HBc_{1–149}$ dimers in the absence of salt or denaturant (black solid line), suggesting small amounts of denaturant are sufficient to prevent salt-induced capsid assembly. The $HBc_{1–149}$ spectrum in 7 M urea (green dashed line) is essentially identical to that in 4.8 M GdmCl (green solid line), suggesting that the folding intermediate thus populated has the same structural properties. A.U., arbitrary units.

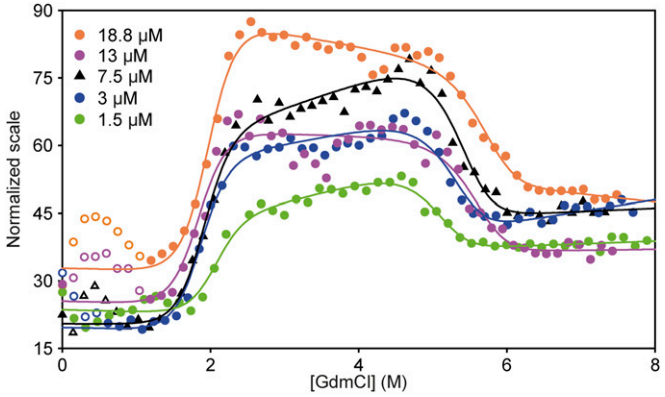

Fig. S2. GdmCl-induced equilibrium denaturation of $HBc_{1–149}$ signals over a 12.5-fold protein concentration range. These data are the raw data corresponding to the normalized data presented in Fig. 3. Each dataset is fitted to an equation (solid lines) describing a linear three-state transition with a dimeric intermediate (*Materials and Methods* and *SI Materials and Methods*).

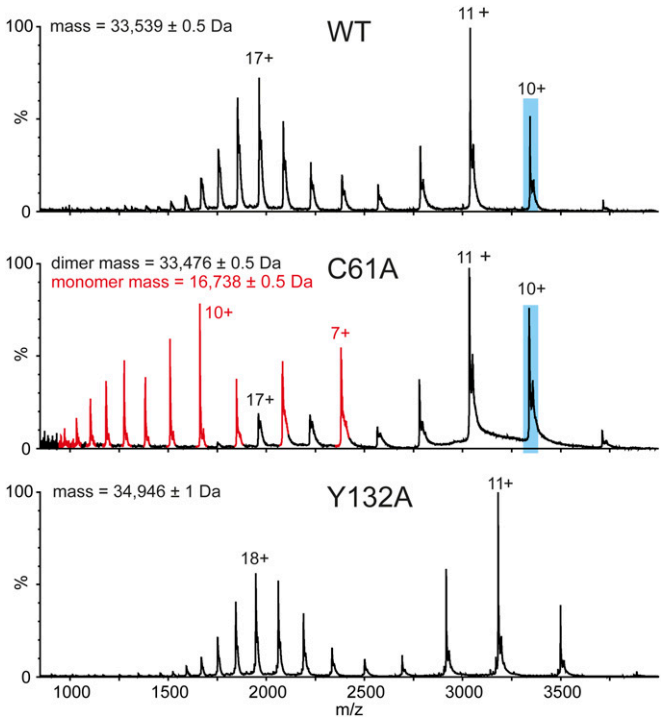

Fig. S3. *m/z* Distribution of $HBc_{1–149}$ ions in the gas phase. WT $HBc_{1–149}$ (*Top*), C61A (*Middle*), and Y132A (*Bottom*) showed expected masses and no evidence of bound metals or ligands in ESI-IMS-MS experiments. WT and Y132A had very similar *m/z* distributions. C61A showed additional peaks at higher charges corresponding to an $HBc_{1–149}$ monomer (red). The 10+ charge states (highlighted in blue) were used for analysis of unfolding mechanisms.

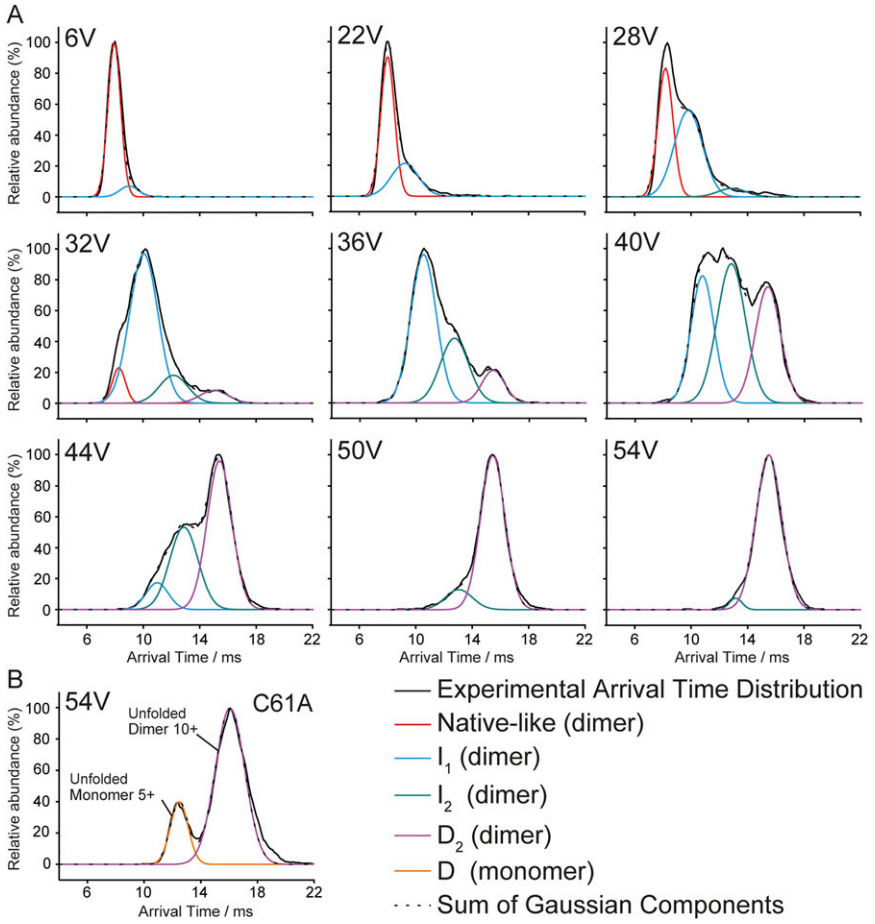

Fig. S4. Fitting of the arrival time distribution in ESI-IMS-MS unfolding experiments. (*A*) At each collision energy (ranging from 6 to 54 V), the data were fitted to the minimum number of Gaussian curves such that the sum of these curves accounted for the experimental arrival time distribution. With increasing collision energy, the arrival time of the detected ions increased, corresponding to an expansion in the collision cross-section surface area and consistent with the protein becoming unfolding due to an increase in internal energy. (*B*) Removal of the intermolecular disulfide bond by the C61A mutation caused the appearance of monomer at high collision energies, suggesting unfolding precedes dimer dissociation.

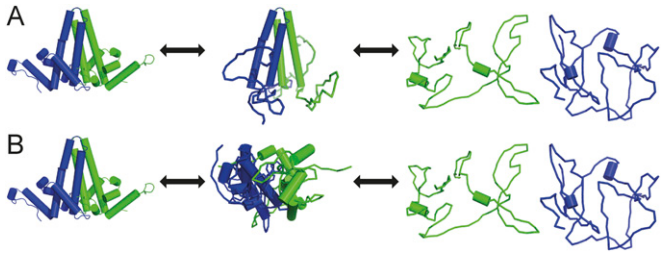

Fig. S5. Schematic of the three-state unfolding mechanisms for $HBc_{1–149}$ that can be excluded by our data (as discussed in the main text). (*A*) $HBc_{1–149}$ has independent foldons, where contact domains fold and unfold independently of the four-helix bundle (which remains structured in the intermediate state). (*B*) Native $HBc_{1–149}$ unfolds via a highly helical, molten globule-like state that is expanded but lacks extensive tertiary structure. The individual HBc_{1-149} monomers are shown in green and blue for clarity.

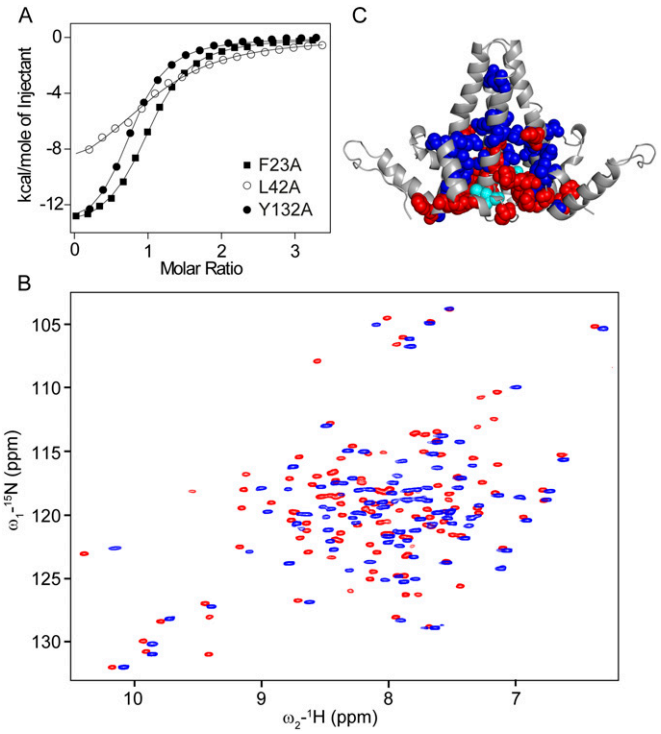

Fig. S6. Evidence of L42A inducing long-range structural rearrangements in $HBc_{1–149}$. (*A*) ITC experiments of $HBc_{1–149}$ variants titrated with an antiviral peptide (MHRSLLGRMKGA). Y132A and F23A had identical K_d values (both 3 μM), which agreed well with our reported values for WT (K_d of ∼2 μM) (2). Thus, these mutants did not significantly affect binding of this peptide to the top of the four-helix bundle (Fig. 5*F*). By contrast, the L42A mutation significantly reduced the binding affinity (K_d of ∼11 μM), even though L42A has a WT-like far-UV CD signal (Fig. 5*D*) and the mutation site is >37 Å from the binding site (Fig. 5*F*). (*B*) Transverse relaxation-optimized spectroscopy (TROSY) spectra of WT (red) and L42A (blue) $HBc_{1–149}$ under the experimental conditions used in this study. Although many L42A amide resonances had similar chemical shifts to WT $HBc_{1–149}$ (2), a number of cross-peaks had significant CSPs. (*C*) CSPs from *B* were mapped onto the structure of the native dimer and classified as very large or exchange-broadened CSPs (red) or large CSPs (blue, as described in *SI Materials and Methods*). The largest CSPs were close to the mutation site (L42 is shown in cyan) but also propagated through the four-helix bundle and regions around the base of this bundle. The data in this figure support L42A inducing large-scale structural rearrangements in $HBc_{1–149}$ (as discussed in the main text). A more comprehensive analysis of these NMR data will be reported in a separate paper.

Table S1. Equilibrium capsid assembly reactions for different $HBc_{1–149}$ variants

Mutation	Capsid, %	Dimer, %
WT	80	20
K7A	55	45
F23A	1	99
L30A	45	55
L42A	2	98
L60V	82	18
V72A	74	26
S87D	83	17
F97A	79	21
F97L	87	13

As per the data shown in Fig. 5*E*, samples containing 100 μM different $HBc_{1–149}$ variants were dialyzed overnight at 20 °C under capsid assembly conditions [0.1 M Hepes buffer (pH 7.5) with the ionic strength corrected to 250 mM using NaCl]. Thereafter, the different species populated at equilibrium were resolved using size-exclusion chromatography, and the relative proportion of capsids and dimers was determined from differential refractive index measurements. In an effort to provoke capsid formation in F23A and L42A, these samples were dialyzed at even higher protein concentrations (∼600 μM), conditions where 100% of WT $HBc_{1–149}$ molecules form capsids. However, L42A and F23A showed no significant population of capsids even under these conditions. The SD from triplicate measurements was ≤±2%.