Supporting Information

Structure and Dynamics of Full Length HIV-1 Capsid Protein in Solution

Lalit Deshmukh[1], Charles D. Schwieters[2], Alexander Grishaev[1], Rodolfo Ghirlando[3], James L. Baber[1], and G. Marius Clore[1,4]

[1]Laboratory of Chemical Physics and [3]Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520, U.S.A.

[2]Division of Computational Biosciences, Center for Information Technology, National Institutes of Health, Bethesda, MD 20892-5624

[4]To whom correspondence should be addressed. E-mail: mariusc@mail.nih.gov

2 Tables and 5 figures

Table S1. SVD fits of the bicelle RDC data for the C-terminal domain of full-length capsid (CA_{FL}) at three concentrations simultaneously using a mixture of monomer and dimer.[a]

Subunit Concentration	R-factor (%)[b]		target R-factor (%)[c]
	2KOD orientation	1A43 orientation	
(a) CA_{FL} only[d,e]			
0.2 mM	15.9	17.1	14.3
0.1 mM	19.1	19.0	18.1
0.05 mM	15.5	16.1	15.4
(b) CA_{FL} + CA_{FL}^{V181C} [f]			
CA_{FL}^{V181C}	20.4	32.4	18.9
CA_{FL} (0.2 mM)	16.3	17.6	14.3
CA_{FL} (0.1 mM)	19.1	19.1	18.1
CA_{FL} (0.05 mM)	15.7	16.6	15.4

[a]The fractions of monomer and dimer at the various concentrations are computed using the K_{dimer} values obtained from analytical ultracentrifugation (see Table3).

[b]The C-terminal domain dimer coordinates are obtained by fitting the C-terminal domain coordinates from the crystal structure of CA_{FL} (PDB id 3NTE)[S1] onto the NMR structure of $CA_{144-231}$ (PDB id 2KOD)[S2] or the crystal structure of $CA_{146-231}$ (PDB id 1A43)[S3] dimer structures. SVD is performed on all data simultaneously using two alignment tensors, one for the dimer and one for the monomer.

[c]The target values are the R-factors obtained by SVD against a single C-terminal domain subunit.

[d]The D_a^{NH} and rhombicity values obtained with the dimer in the 2KOD orientation are 11.8 Hz and 0.04, respectively, for the dimer, and 15.8 Hz and 0.5, respectively, for the monomer. The corresponding values obtained with the 1A43 orientation are 8.84 Hz and 0.18, respectively, for the dimer, and -14.6 Hz and 0.62 for the monomer. At a concentration of 0.2 mM (where the fraction of subunits in the monomeric state is 36%), the RDC R-factor obtained with the C-terminal domain dimer in the 2KOD orientation is significantly better than that in the 1A43 orientation; the much smaller differences in R-factor for the two alternative C-terminal domain dimer orientations at the lower concentrations is due to the larger fraction of subunits in the monomeric state (58% at 0.05 mM and 47% at 0.1 mM).

[e]The values of D_a^{NH} and rhombicity obtained for the dimer in the 2KOD orientation are very similar to those obtained from the SVD fits for the C-terminal domain of the CA_{FL}^{V181C} dimer alone in bicelles (10.3 Hz and 0.17, respectively) indicating that the overall shape of the CA_{FL} and CA_{FL}^{V181C} ensembles is similar (i.e. the conformational space sampled by the N-terminal domains relative to the C-terminal domain dimer is broadly similar for the two constructs). The D_a^{NH} and rhombicity obtained for the monomer, however, are different from those obtained from the SVD fits for the C-terminal domain of the monomeric $CA_{FL}^{W184A/M185A}$ mutant (-11.7 Hz and 0.4, respectively; cf Table S3), indicating that the conformational space sampled by the N-terminal domain relative to the C-terminal domain is very different for the wild-type and mutant monomers.

[f]When the data for the dimeric CA_{FL}^{V181C} mutant and the CA_{FL} wild-type at three concentrations are fitted by SVD simultaneously, it can be seen that the 2KOD orientation is consistent with all 4 data sets while the 1A43 orientation is inconsistent with the dimeric CA_{FL}^{V181C} mutant data set and yields a significantly poorer fit for the CA_{FL} wild-type at a concentration of 0.2 mM. The values of D_a^{NH} and rhombicity obtained with the 4 data sets in the 2KOD orientation are very similar to those obtained using the RDC data for the three CA_{FL} concentrations alone, namely 10.5 Hz and 0.2, respectively for the dimer, and 14.5 Hz and 0.6, respectively, for the monomer. The corresponding values of D_a^{NH} and rhombicity using the 1A43 orientation are 7.1 Hz and 0.4 for the dimer, and -15.0 Hz and 0.5, respectively, for the monomer.

Table S2. Force constants used during simulated annealing.

Xplor-NIH potential term	Restraints	Force constant	Units
SARDCPot	steric alignment RDCs	0.1	$kcal.mol^{-1}$
diNCS	ensemble non-crystallographic symmetry restraints	0.1…10	$kcal.mol^{-1}.Å^{-2}$
dSymm	ensemble distance symmetry restraints	100	$kcal.mol^{-1}.Å^{-2}$
solnScatPot	X-ray scattering potential	400	$kcal.mol^{-1}$
residueAff	low resolution contact potential	1	
rcon	quartic atom-atom repulsion	0.004…4	$kcal.mol^{-1}.Å^{-4}$
repel	vdw radius scale factor	0.9…0.8	
torsionDB	(i) general multi-dimensional torsion angle database potential	0.002…1	
	(ii) linker multi-dimensional torsion angle database potential	2	
bond	bond length	1000	$kcal.mol^{-1}Å^{-2}$
angle	bond angle	200…500	$kcal.mol^{-1}.rad^{-2}$
improper	improper dihedral	50…500	$kcal.mol^{-1}.rad^{-2}$

-

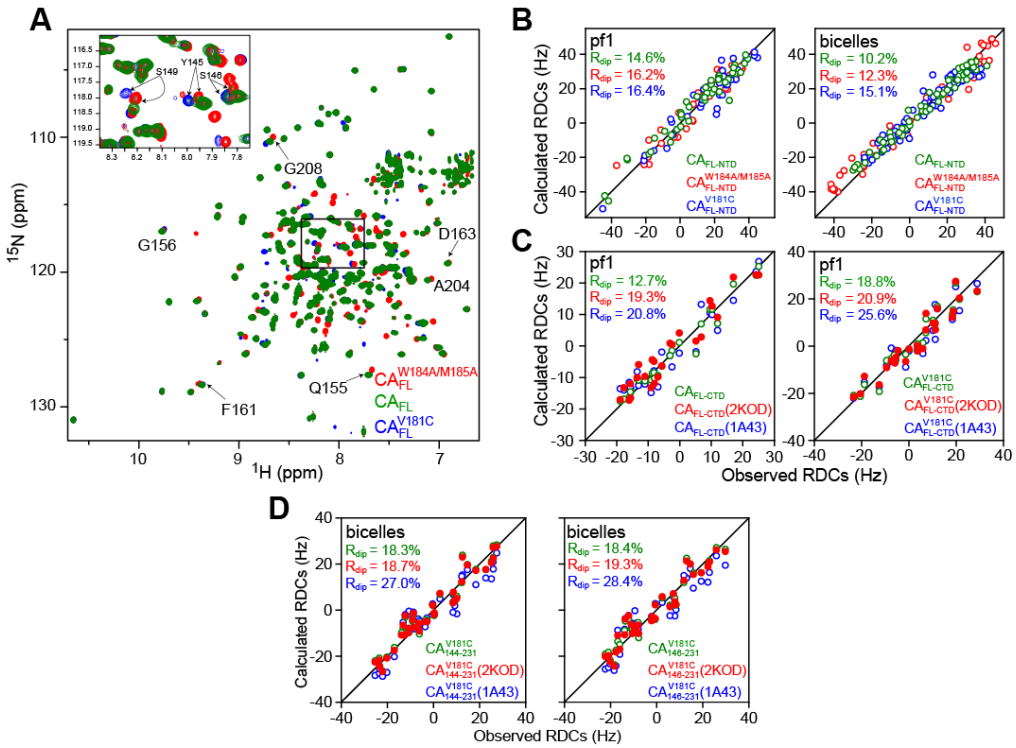

Figure S1. NMR analysis of HIV-1 capsid constructs. (*A*) Overlay of the ^{1}H-^{15}N TROSY spectrum of ^{2}H/^{15}N-labeled wild-type HIV-1 capsid, CA_{FL} (green), with those for the mutant CA_{FL}^{V181C} obligate dimer (blue) and the mutant $CA_{FL}^{W184A/M185A}$ monomer (red) (0.5 mM in subunits, 35°C, in buffer containing 20 mM sodium phosphate, pH 6.5, 50 mM NaCl, 1 mM EDTA, 7% D_2O / 93% H_2O and 1 mM DTT, with the latter omitted for the disulfide-linked mutant dimer). For all three constructs, the resonances of the N-terminal domain (residues 1-145) superimpose nicely; for the C-terminal domain (residues 150-221), however, some chemical shift differences are observed as a consequence of the presence of different oligomerization states: namely, monomer ($CA_{FL}^{W184A/M185A}$), dimer (CA_{FL}^{V181C}), and a monomer-dimer equilibrium (CA_{FL}). A few cross-peaks for the C-terminal domain are labeled. The insert shows an expanded region of the ^{1}H-^{15}N TROSY spectra with cross-peaks for some of the linker residues of the mutant CA_{FL}^{V181C} and $CA_{FL}^{W184/M185A}$ constructs labeled. For wild-type CA_{FL}, the $^{1}H_N/^{15}N$ cross-peaks for many residues located in close proximity to the indole ring of Trp184, including the dimer interface and the linker connecting the N-terminal domain to the C-terminal domain, are broadened beyond detection owing to chemical exchange between the monomer and dimer on a time scale that is intermediate on the chemical shift scale (i.e. $k_{ex} \sim \Delta\omega$ where k_{ex} is the overall exchange rate and $\Delta\omega$ the difference in chemical shifts between monomeric and dimeric forms of CA_{FL}): specifically, the cross-peaks for residues 145-154 and 169-193 are not observed in wild-type CA_{FL}. Nevertheless the overall quality of the NMR spectra of CA_{FL} are high making subsequent RDC analysis feasible. For the mutant CA_{FL}^{V181C} dimer, cross-peaks for residues 150-153, 170-177, 182, 186 and 189-191 are not observed; for the mutant $CA_{FL}^{W184/M185A}$ monomer only cross-peaks for residues 176 and 180-182 are not observed. (*B*) SVD analysis of RDCs measured in phage pf1 (left panel) and bicelles (right panel) for the N-terminal domain of wild-type CA_{FL} (green), the CA_{FL}^{V181C} obligate dimer (blue) and the $CA_{FL}^{W184A/M185A}$ monomer (red) using the coordinates of the 3NTE crystal structure.[S1] (*C*) SVD analysis of the RDCs measured in phage pf1 for CA_{FL} (left panel) and the CA_{FL}^{V181C} obligate dimer using the coordinates of the 3NTE[S1] crystal structure of capsid for a single C-terminal domain subunit (green) and for the dimer in the orientations observed for $CA_{144-231}$ by NMR (2KOD)[S2] (red) and $CA_{146-231}$ by crystallography (1A43)[S3,S4]. In the case of the wild-type CA_{FL}, the fits to the dimer in the NMR (2KOD) and X-ray (1A43) orientations are significantly worse (RDC factors 50-60% higher) than the fits to an individual subunit (green); for the CA_{FL}^{V181C} obligate dimer, however, the fits to an individual subunit (green) and to a dimer in the NMR (2KOD) orientation (red) are comparable, but significantly worse to a dimer in the X-ray (1A43) orientation (blue). (D) SVD analysis of the RDCs measured in bicelles for the disulfide linked C-terminal domain constructs $CA_{144-231}^{V181C}$ (left panel) and $CA_{146-231}^{V181C}$ (right panel) using the 3NTE[S1] crystal structure coordinates. In both instances the fits to an individual subunit (green) and to a dimer in the NMR (2KOD) orientation (red) are comparable, but the fits to a dimer in the X-ray (1A43) orientation (blue) are significantly worse (RDC R-factor about 50% higher).

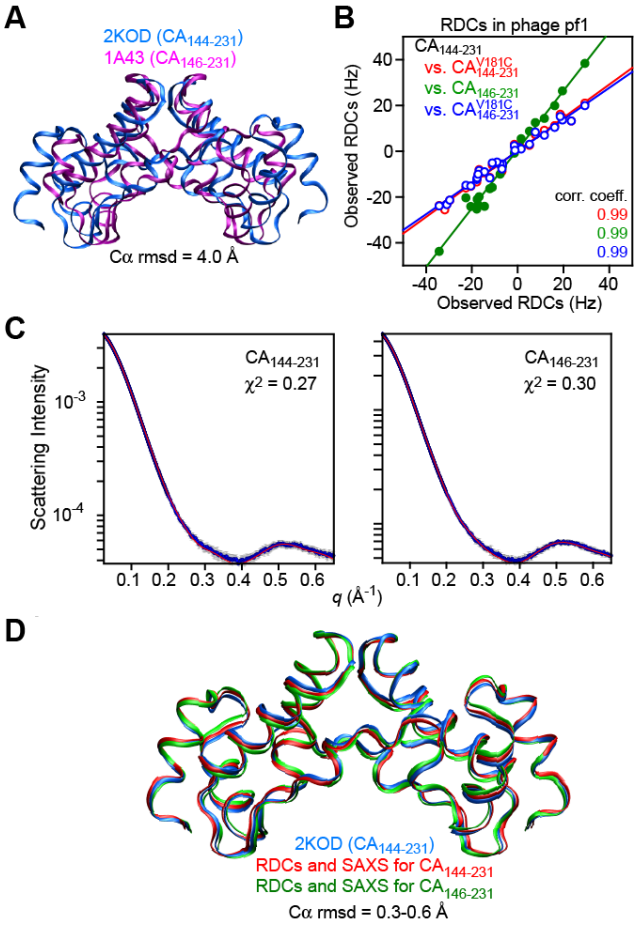

Figure S2. Joint SAXS/WAXS and RDC rigid body simulated annealing refinement of the $CA_{144-231}$ and $CA_{146-231}$ C-terminal domain constructs of HIV-1 capsid. (*A*) Superposition of the NMR structure of $CA_{144-231}$ (2KOD)[S2] (blue) and the X-ray structure of $CA_{146-231}$ (1A43)[S3,S4] (purple). (*B*) Correlation between RDCs measured in phage pf1 for the $CA_{144-231}$ construct versus those measured for the shorter $CA_{146-231}$ (green) construct and the obligate dimer $CA^{V181C}_{144-231}$ (red) and $CA^{V181C}_{146-231}$ (blue) constructs. The high correlation in phage pf1, as well as in bicelles (cf. Fig. 1B of main paper), indicate that the orientation of the two subunits in the dimer is identical in all four C-terminal domain constructs (cf. Figure 2 of main text) in solution. (*C*) Agreement between observed (blue with black error bars equal to 1 s.d.) and calculated (red line) SAXS/WAXS scattering curves for the $CA_{144-231}$ and $CA_{146-231}$ C-terminal domain constructs following rigid body refinement using the X-ray coordinates (PDB id 3NTE)[S1] for the C-terminal domain subunits. The starting coordinates were obtained by superimposing the coordinates for the C-terminal domain from the 3NTE crystal structure[S1] onto the coordinates of the NMR structure of the $CA_{144-231}$ dimer (2KOD)[S2] and the X-ray structure of the $CA_{146-231}$ dimer (1A43).[S3,S4] (D) Superposition of the 2KOD[S2] NMR coordinates (blue) with those obtained after joint SAXS/WAXS and RDC refinement for the $CA_{144-231}$ (red) and $CA_{146-231}$ (green) constructs.

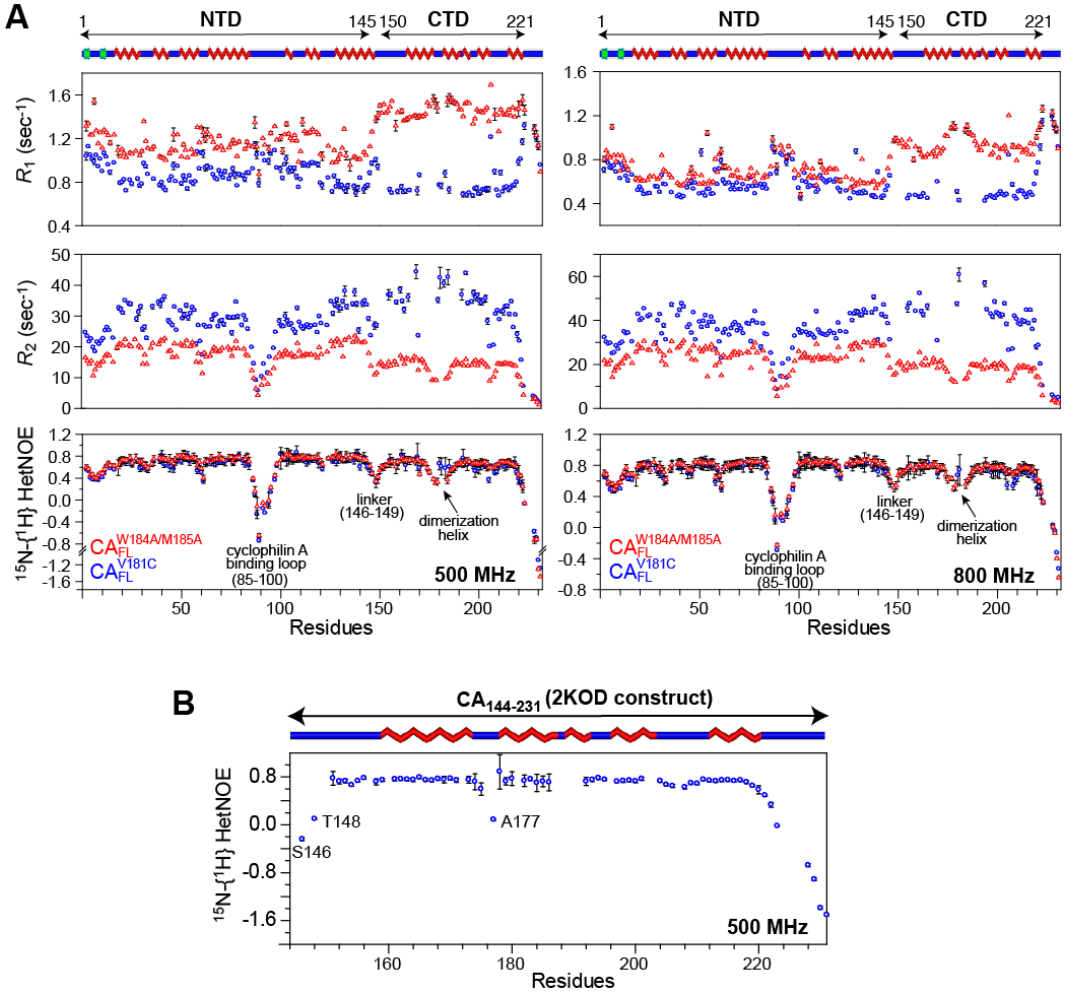

Figure S3. TROSY-based ^{15}N R_1, R_2 and ^{15}N-{^{1}H} heteronuclear NOE relaxation analysis of HIV-1 capsid constructs. (*A*) ^{2}H/^{15}N-labeled full-length CA_{FL}^{V181C} dimer (blue) and $CA_{FL}^{W184A/M185A}$ monomer (red) mutant constructs (0.5 mM in subunits, 35°C, pH 6.5). All spectra were recorded in an interleaved manner. At 500 MHz, the following relaxation delays were used. For $CA_{FL}^{W184A/M185A}$: R_1 (0, 240, 400, 560, 720, 800, 960, 1120 ms), $R_{1\rho}$ (2, 22, 42, 62, 72, 82, 102, 122 ms); for CA_{FL}^{V181C}: R_1 (0, 400, 800, 1000, 1200, 1320, 1400, 1600 ms), $R_{1\rho}$ (2, 20, 30, 35, 40, 50, 60, 70 ms). At 800 MHz the relaxation delays were as follows. For $CA_{FL}^{W184A/M185A}$: R_1 (0, 400, 800, 1000, 1200, 1320, 1400, 1600 ms), $R_{1\rho}$ (2, 20, 35, 50, 55, 60, 70, 80 ms); for CA_{FL}^{V181C} construct: R_1 (0, 400, 1000, 1400, 1600, 1720, 1800, 2000 ms), $R_{1\rho}$ (2, 10, 20, 35, 40, 45, 50, 60 ms). R_2 is given by $[R_{1\rho} - R_1\cos^2(\theta)/\sin^2(\theta)]$ where θ is the angle between the effective spin-lock field and the external magnetic field (and a value of 90° represents a resonance exactly on-resonance with the spin-lock field). The strengths of the radiofrequency spin-lock fields (ν_{RF}) employed in the $R_{1\rho}$ experiments were 1.5 and 2 kHz for $CA_{FL}^{W184A/M185A}$ and CA_{FL}^{V181C}, respectively. The delineations for the N- (residues 1-145) and C- (residues 150-221) terminal domains, as well as the location of the secondary structure elements, are indicated at the top of the panels. A few key motifs (such as the cyclophilin-A binding loop, the linker connecting the N- and C-terminal domains, and the dimerization helix of the C-terminal domain) are labeled. The low {^{1}H}-^{15}N heteronuclear NOE values (<0.5) for residues 184-188 in the $CA_{FL}^{W184A/M185A}$ mutant monomer construct are indicative of a high degree of internal mobility consistent with disruption/local unfolding within the dimerization helix (residues 179-192) as a consequence of the two mutations. (*B*) ^{2}H/^{15}N/^{13}C-labeled $CA_{144-231}$ construct (corresponding to that used in the 2KOD[S2] NMR structure) (~1 mM in subunits, 35°C, pH 6.5). The ^{15}N-{^{1}H} heteronuclear NOEs were measured using a (10 + 1 s) saturation/recovery delay.

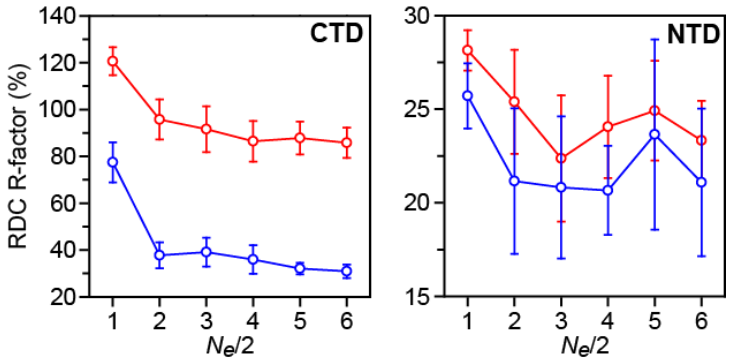

Figure S4. Cross-validation of the wild-type dimer and monomer CA_{FL} ensembles against the experimental bicelle RDCs for the mutant CA_{FL}^{V181C} dimer (blue) and $CA_{FL}^{W184A/M185A}$ monomer (red), respectively. The wild-type ensembles are those calculated based on the experimental SAXS/WAXS and bicelle RDC data at all concentrations (cf. Fig. 4 of main text). The left- and right-hand panels displays the RDC R-factors for the C- and N-terminal domains, respectively, as a function of ensemble size $N_e/2$. In the case of the C-terminal domain, the wild-type dimer ensemble cross-validates well against the CA_{FL}^{V181C} mutant dimer data with the RDC R-factor decreasing from ~80% at N_e =1 to ~30% at $N_e/2$= 5. Although, the RDC R-factor for the C-terminal domain of the wild-type monomer ensemble versus the $CA_{FL}^{W184A/M185A}$ monomer mutant data also decreases as the ensemble size increases (from ~120% at $N_e/2$ = 1 to ~90% at $N_e/2$ = 5), the fit remains very poor, indicating that the distribution of the N-terminal domain relative to the C-terminal domain for the wild-type and $CA_{FL}^{W184A/M185A}$ mutant monomers are different. This was also evident from the results of SVD analysis presented in Table S1 (footnote e), and was confirmed by the results of ensemble simulated annealing calculations driven by the bicelle RDC and SAXS/WAXS data for $CA_{FL}^{W184A/M185A}$ mutant monomer (see Figure 7, main text).

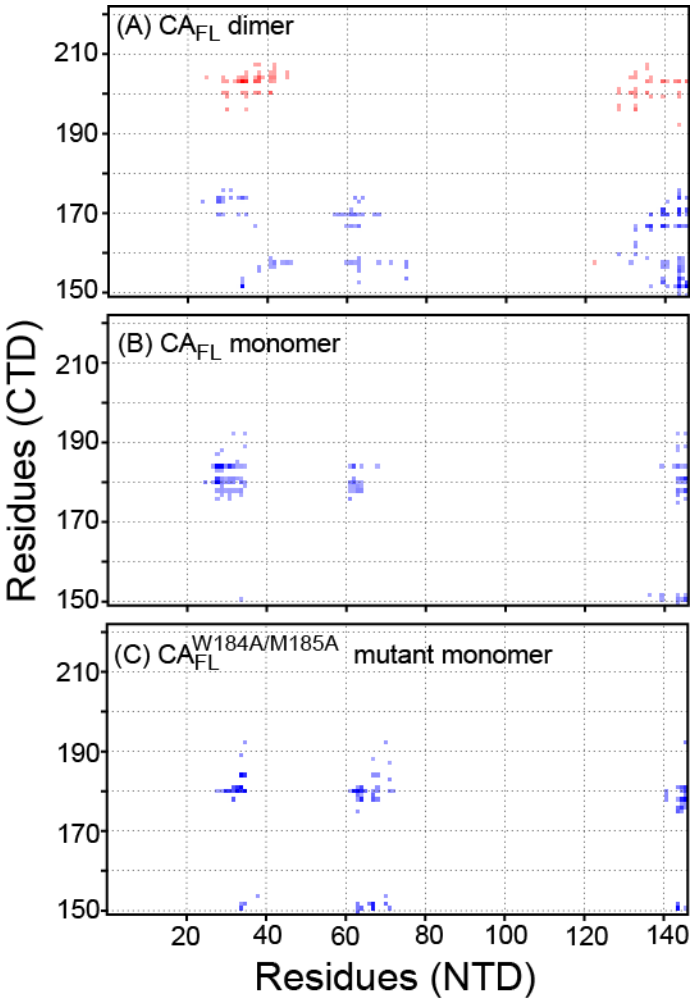

Figure S5. Summary of transient contacts found between the N- and C-terminal domains in the calculated ensembles for (*A*) the wild-type CA_{FL} dimer, (*B*) the wild-type CA_{FL} monomer, and (*C*) the $CA_{FL}^{W184A/M185A}$ mutant monomer. The blue and red dots indicate intra and inter-subunit contacts, respectively. A contact is defined as any atom of the respective residues within 4 Å. Greater color intensity indicates that the contact has a greater weighted probability over all calculated structures.

Supplementary references

(S1) Du, S.; Betts, L.; Yang, R.; Shi, H.; Concel, J.; Ahn, J.; Aiken, C.; Zhang, P.; Yeh, J. I. *J. Mol. Biol.* **2011**, *406*, 371-386.

(S2) Byeon, I. J.; Meng, X.; Jung, J.; Zhao, G.; Yang, R.; Ahn, J.; Shi, J.; Concel, J.; Aiken, C.; Zhang, P.; Gronenborn, A. M. *Cell* **2009**, *139*, 780-790.

(S3) Worthylake, D. K.; Wang, H.; Yoo, S.; Sundquist, W. I.; Hill, C. P. *Acta Cryst. Section D* **1999**, *55*, 85-92.

(S4) Gamble, T. R.; Yoo, S.; Vajdos, F. F.; von Schwedler, U. K.; Worthylake, D. K.; Wang, H.; McCutcheon, J. P.; Sundquist, W. I.; Hill, C. P. *Science* **1997**, *278*, 849-853.