CHEMPHYSCHEM

Supporting Information

Quantitative Characterization of Configurational Space Sampled by HIV-1 Nucleocapsid Using Solution NMR, X-ray Scattering and Protein Engineering

Lalit Deshmukh,[a] Charles D. Schwieters,[b] Alexander Grishaev,[c] and G. Marius Clore*[a]

cphc_201600212_sm_miscellaneous_information.pdf

EXPERIMENTAL

Protein expression and purification. Full-length nucleocapsid (residues 378-432, plasmid HXB2) was sub-cloned in a pET-11a vector. To increase expression levels, the truncated nucleocapsid (residues 390-428, plasmid HXB2) was sub-cloned in a modified pET-11a vector containing a N-terminal GB1 tag (B1 immunoglobulin binding domain of streptococcal protein G),[1] and a Factor Xa cleavage site. Both constructs were expressed in BL21-CodonPlus (DE3)-RIPL competent cells (Agilent Technologies). Uniformly $^{2}H/^{15}N/^{13}C$-labeled proteins were expressed at 37°C using our earlier protocol.[2] Briefly, cells were grown in 1 L minimal M9 medium containing 0.3 g/L $^{2}H/^{15}N/^{13}C$ Isogro (Sigma-Aldrich), $^{2}H_2O$, 1g/L $^{15}NH_4Cl$ and 2g/L $^{2}H_7$,$^{13}C_6$-D-glucose (Cambridge Isotope Laboratories, Inc). Cells, induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at an optical density of A_{600} ~0.8, were harvested ~8 hours later.

For both proteins, cells were re-suspended in a lysis buffer containing 100 mM Tris, pH 8.0, 0.1 mM $ZnCl_2$, 5 mM β-mercaptoethanol (BME), and 1 cOmplete Protease Inhibitor Cocktail tablet (Roche Applied Science). Nucleic acids were precipitated by adding 4% (w/v) polyethyleneimine, pH 8.0 (Sigma-Aldrich) to a final concentration of 0.4% (w/v). Proteins were purified by combination of ion exchange and size exclusion chromatography. The cell lysates were loaded onto a HiPrep 16/10 Q FF column (GE Healthcare) with a 0 to 1 M NaCl gradient in buffer containing 100 mM Tris, pH 8.0, 0.1 mM $ZnCl_2$ and 5 mM BME. For the full-length nucleocapsid, relevant flow-through fractions were loaded onto a HiPrep 16/10 SP FF column (GE Healthcare) with a 0 to 1 M NaCl gradient in buffer containing 100 mM Tris, pH 8.0, 0.1 mM $ZnCl_2$. For the truncated nucleocapsid, eluted protein was concentrated (Amicon ultra-15, 3 kDa cut-off) and loaded onto a HiLoad 26/60 Superdex 75 column (GE Healthcare) pre-equilibrated with 20 mM Tris, pH 8, 0.1 mM $ZnCl_2$, 2 mM $CaCl_2$, and 100 mM NaCl. The eluted protein was concentrated and treated with Factor-Xa protease (New England BioLabs Inc.) overnight at room temperature. Both proteins were further purified using a Mono S™ 10/100 GL column (GE Healthcare) with a 0 to 1 M NaCl gradient in buffer containing 100 mM Tris, pH 8, 0.1 mM $ZnCl_2$.

All constructs were verified by DNA sequencing and mass spectrometry (using an Agilent 1100 LC/MS system equipped with an Agilent Zorbax 300SB-C3 column coupled to a quadrupole mass analyzer).

NMR Sample Preparation. All heteronuclear NMR experiments were performed on uniformly $^{15}N/^{13}C/^{2}H$ labeled samples prepared in buffer containing 20 mM sodium phosphate, pH 6.5, 50 mM NaCl, 0.1 mM $ZnCl_2$, 93% H_2O/7% D_2O and 1 mM dithiothreitol (DTT). Aligned samples were prepared using 1,2 dimyristoyl-sn-glycero--3-phosphocholine (DMPC)/06:0 diether phosphatidyl choline (PC) bicelles (q=3) (Avanti polar lipids) doped with 0.1% PEG-2000-PE (Avanti polar lipids) to improve bicelle stability.[3] Backbone $^{1}D_{NH}$, $^{1}D_{NC'}$, and $^{1}D_{C\alpha C'}$ coupling data were measured on samples containing 0.1- 0.35mM protein. Heteronuclear ^{15}N-$\{^{1}H\}$ NOE measurements were measured on samples containing ~ 0.2 mM full length nucleocapsid.

NMR Spectroscopy. All heteronuclear NMR experiments were carried out at 35°C on Bruker 500 and 800 MHz spectrometers equipped with z-gradient triple resonance cryoprobes. Spectra were processed using NMRPipe[4] and analyzed using the CCPN software suite.[5] Backbone amide ($^{1}D_{NH}$) RDCs were measured on $^{2}H/^{13}C/^{15}N$-labeled proteins using the TROSY-based ARTSY technique.[6] Backbone $^{1}D_{NC'}$ and $^{1}D_{C\alpha C'}$ RDCs were measured on $^{2}H/^{13}C/^{15}N$-labeled proteins using quantitative $J_{NC'}$ and $J_{C\alpha C'}$ pulse sequences with TROSY readout[7] and analyzed using Xplor-NIH.[8] The ^{15}N-$\{^{1}H\}$ NOE and reference spectra were recorded with a 10 second saturation time for the NOE measurement and equivalent

recovery time for the reference measurement in an interleaved manner, each preceded by an additional 1 sec recovery time.

SAXS Data Collection and Analysis. All SAXS data were collected at 25°C at Beam Line 12-IDB, Advanced Photon Source (Argonne National Laboratory, Argonne, IL). The sample buffer was the same as that employed in the NMR experiments except for the use of H_2O instead of 93% H_2O/7% D_2O. SAXS data for the truncated and full-length nucleocapsid variants were collected during two separate sessions with the truncated variant data recoded with the incident beam intensity lower by a factor of ~25 relative to the session where the full- length variant data were collected, leading to a much higher uncertainty for the truncated data set compared to that of the full length protein. For the full-length nucleocapsid, X-ray scattering data were acquired at protein concentrations of 5 mg/mL (~0.8 mM) and 2.5 mg/mL (~0.4 mM) using a Pilatus 2M detector positioned 2.0 m from the sample capillary in a highly offset geometry with 14 keV incident radiation resulting in an observable q-range of 0.01−0.92 $Å^{-1}$. Scattered radiation was detected subject to a 13 keV low-energy cutoff. For the truncated nucleocapsid, X-ray scattering data were acquired at protein concentrations of 2.3 mg/mL (~0.5 mM), 1.2 mg/mL (~0.25 mM) and 0.6 mg/mL (~0.125 mM) with an observable q-range of 0.01−0.92 $Å^{-1}$. Q-axis mapping was done using a silver behenate standard sample. A total of 20 sequential data frames with exposure times of 2 s were recorded for the full length variant data and 10 s for the truncated variant data, with the samples kept at 25°C throughout the measurements. To prevent radiation damage, 100 μL volumes of samples and buffers were oscillated during data collection. Individual data frames were masked, corrected for the detector sensitivity, radially integrated, and normalized by the corresponding incident beam intensities and sample transmissions. The final 1D scattering profiles and their uncertainties were calculated as means and mean uncertainties over the 20 individual frames. The buffer data were then subtracted from the samples. Triplicate measurements were performed for all recorded concentration points. Radii of gyration extracted from the lowest-q data *via* Guinier fits coincided within their uncertainties with the values extracted as the second moments of the P(r) functions obtained via regularized Fourier transforms of the data. The highest concentration data were used for the structure refinement as no concentration-dependent effects were identified from comparative data analysis. Due to the presence of buffer subtraction uncertainty, especially pronounced with lower concentrations and lower incident beam intensities, Q ranges of the data used for the structure calculation were restricted to the regions where the r.m.s. deviations between the triplicate measurements were consistent with the photon count–based data uncertainties (specifically, q values below 0.087 and 0.057 $Å^{-1}$ were excluded for the truncated and full-length nucleocapsid calculations; for small proteins such as the ~6 kDa nucleocapsid this is not a problem; see below).

Structure Calculations and Computational Methodology. Structure calculations for the truncated nucleocapsid were performed with ensemble size, N_e, ranging from 1 to 8. The starting coordinates were taken from PDB 2M3Z.[9] These calculation employed the SARDC potential term[10] which calculates the alignment tensor directly from molecular shape – applicable for experiments performed using purely steric alignment media such as neutral bicelles. In the time-scale regime considered here, each ensemble member has its own effective alignment tensor. Note that if one were to let the alignment tensors float to optimize fits to experimental variables, there would be far too many parameters for the given datasets. The SAXS curves were used in the range q = 0.087 to 0.54 $Å^{-1}$ and 0.057 to 0.65 $Å^{-1}$ for the truncated

and full-length nucleocapsid constructs, respectively, linearly interpolated at 30 points. SAXS errors were scaled down by a factor of $\sqrt{20}$ to account for the fact that they were obtained by averaging 20 curves. Corrections for globbing, numerical integration over solid angle[11] and bound-solvent contributions were recomputed at each temperature of simulated annealing. For computation of the final SAXS curves and χ^2 value, 100 points were used and a cubic spline was used to evaluate *I(q)* at each experimental value of q.

It should be noted that the q_{min} values of 0.087 and 0.057 $Å^{-1}$ used for the truncated and full-length nucleocapsid SAXS data implies that the SAXS data do not provide information on distances exceeding ~40 Å for the truncated construct and ~70 Å for the full-length construct. The maximum dimension (d_{max}) of the truncated construct could then theoretically be ill-defined in the absence of any other structural information or additional restraints, since the largest inter-atomic distance with the linker fully extended could reach up to ~50 Å. In general, however, d_{max} extracted from SAXS data is a "soft" parameter difficult to obtain accurately, even more so for a conformationally flexible protein such as the nucleocapsid studied here. Therefore we do not claim that we are able to determine d_{max} directly from the SAXS data. Indeed, if we were basing our SAXS data analysis on interpretation of the fitted $P(r)$ function we would have a problem as we would not be able to extend d_{max} beyond ~40 Å for the truncated construct. However, minimization of the difference between observed and calculated SAXS curves in simulated annealing calculations does not promote either more compact or more extended states since removal of the lower q-data would simply leave d_{max} under-restrained by the SAXS data. Moreover, even for the fitted SAXS data for the truncated construct sufficient number of points are available to determine the size of the system as a Guinier fit can be performed reliably over ~30 data points in the angular range corresponding to $q.R_{gyr}$ from 0.7 to 1.0, and the uncertainty of the radius of gyration, R_{gyr}, for these data is dominated by their low signal-to-noise rather than the resolution range. More importantly, in our case the SAXS data are not the only source of shape and size information available, since the RDC alignment tensor calculated from the molecular coordinates is very sensitive to shape and size, and use of both the structures of the Zn knuckles and the experimental RDC data therefore impose additional restraints leading to linker geometries that are far from extended, further decreasing the impact of information on large distance separations in the SAXS data of the truncated construct.

The truncated nucleocapsid ensemble structure was calculated in two steps which employed the same potential energy terms and force constants. Torsion angles were restrained by the torsionDBPot multidimensional torsion angle database potential of mean force[12] applied to all active torsion angles. A separate torsionDBPot term (denoted torsionDBPot linker in Table S1) with increased force constant was applied to the dihedral angles in the linker region (see Table S1 footnote [a] for rationale). The residueAffPot term[13] was used to help characterize any potential interactions between the N- and C-terminal zinc knuckles. Standard terms to enforce covalent geometry (bonds, angles and improper torsions) were applied, in addition to a quartic repulsive non-bonded term[14] to prevent atomic overlap. During initial structure calculations**,** the atoms of the C-Zn knuckle were kept fixed and the atoms of the N-Zn knuckle moved as rigid bodies, while atoms in the linker region (residues 407-412) were given all degrees of freedom. The position and orientation of the rigid-body N-Zn knuckle were randomized in space, with a different configuration for each ensemble member. This randomization was followed by patching the linker covalent geometry and then excluding ensembles with one or more members having significant atomic overlap. Initial high-temperature dynamics for the smaller of 500 steps or 1 ps were run

while increasing force constants from initial to final values in 25 steps. At this point, a longer high-temperature dynamics run was performed for the smaller of 20000 steps or 2000 ps. All high-temperature dynamics runs employed a linear target function for the SARDC term with a force constant of 0.1 kcal.mol^{-1}, a 40 kcal.mol^{-1} reduced force constant for the SAXS term, and excluded the linker region in the computation of non-bonded interactions. A quadratic SARDC potential, force constants specified in Table S1 and nonbonded interactions between all atoms were employed in subsequent simulated annealing and gradient minimization calculations. Simulated annealing was performed from 3000 K to 25 K in 25 K increments where dynamics for the smaller of 1000 steps or 1 ps were run at each temperature. Final gradient minimization was performed after simulated annealing, first allowing only linker degrees of freedom, and then additionally freeing torsion angles for all side-chains except those coordinating the zinc atoms. During a second refinement calculation, ensemble weights were optimized using the previously described approach.[2, 10b] In this stage, the backbone atoms, zinc atom and zinc-coordinating side chains of the C-Zn knuckle were fixed in space; the backbone atoms, zinc and zinc-coordinating side chains of the N-Zn knuckle were allowed to move as a rigid body; and the remaining atoms were given Cartesian degrees of freedom. The high-temperature dynamics and simulated annealing schedule were identical to those of the initial structure calculation described above. A total of 100 structures were calculated for each ensemble size (N_e = 1, 2, 3, 4, 6 and 8), and the 10 lowest energy structures for each ensemble size were analyzed.

To calculate the full-length nucleocapsid ensemble, we started with the 10 lowest energy refined structures of the truncated nucleocapsid N_e = 3 ensemble. The protocol included fixing the linker and the two zinc knuckles, adding five tails to each ensemble member and giving them torsion degrees of freedom such that experimental data was fit, while the conformational sampling of the tails seemed complete. During these calculations, sum populations of each of the three sub-ensembles corresponding to the truncated nucleocapsid ensemble members used to seed the calculation were fixed, while the population of each of the five submembers was allowed to vary. Note that the SAXS data are sensitive to tail conformations due to their dependence on overall shape, while the RDC data for the knuckles and linker are sensitive to the tail conformations through the alignment tensor's dependence on shape, as the relative orientation of these associated bond vectors was fixed throughout. In addition to RDCs for the zinc knuckles and linker, RDCs for the tail regions were also used during refinement. In this protocol, molecular dynamics was run at 3000 K for the shorter of 2000 steps or 20000 ps, whichever came first. While the initial coordinates of each copy of the tail were identical, atomic velocities were randomized such that a different course of dynamics was taken for each. Simulated annealing was then performed from 3000 K to 25 K in 25 K increments during which force constants were geometrically scaled, as specified in Table S1. At each temperature, dynamics was performed for the shorter of 1000 steps or 1 ps. Final gradient minimization was performed after simulated annealing, first allowing only tail degrees of freedom, and then additionally freeing side-chain torsion angles. 10 structures were calculated from each set of starting coordinates, and the lowest energy structure selected for analysis, resulting in a final set of 10 unique N_e = 3x5 ensembles for the full-length nucleocapsid.

Table S1. Force constants used during simulated annealing calculations of truncated nucleocapsid

Term	Force Constant	units
SARDCPot	0.05 to 0.5	$kcal.mol^{-1}$
solnScatPot	400	$kcal.mol^{-1}$
residueAffPot	1	
nbond	0.5 to 4	$kcal.mol^{-1}.Å^{-4}$
atomic radius	0.9 to 0.8	
torsionDBPot[a]	0.001 to 2	
torsionDBPot linker[a]	2.001 to 4	
bond	1000	$kcal.mol^{-1}.Å^{-2}$
angle	200 to 500	$kcal.mol^{-1}.radian^{-2}$
improper	50 to 500	$kcal.mol^{-1}.radian^{-2}$
E_{weight}	$500N_e$ to $5N_e$	$kcal.mol^{-1}$

[a] The torsionDBPot potential term biases backbone and side chain dihedral angles towards values seen in high resolution crystal structures in the Protein Databank.[12] In the protocols used for structure calculation here, the standard procedure of ramping the scale factor (or force constant) from a value of 0.001 at high simulated annealing temperatures to a larger value of 2 at the end of simulated annealing was augmented for backbone atoms, comprising the linker degrees of freedom in the truncated form and the backbone degrees of freedom for the refinement calculations on the full-length construct. For these backbone degrees of freedom, we found it to be advantageous to use effective initial and final scale factors of 2.001 and 4, respectively, so that backbone torsion angles would better occupy allowed regions of Ramachandran space. Unlike the mostly freely moving side chain degrees of freedom on the knuckles which otherwise move as rigid bodies, these backbone degrees of freedom directly compete with the orientational effects of the SARDC term. Thus the increased scale factor on portions of the torsionDBPot term must be balanced carefully against the SARDC term, and indirectly against the SAXS term.

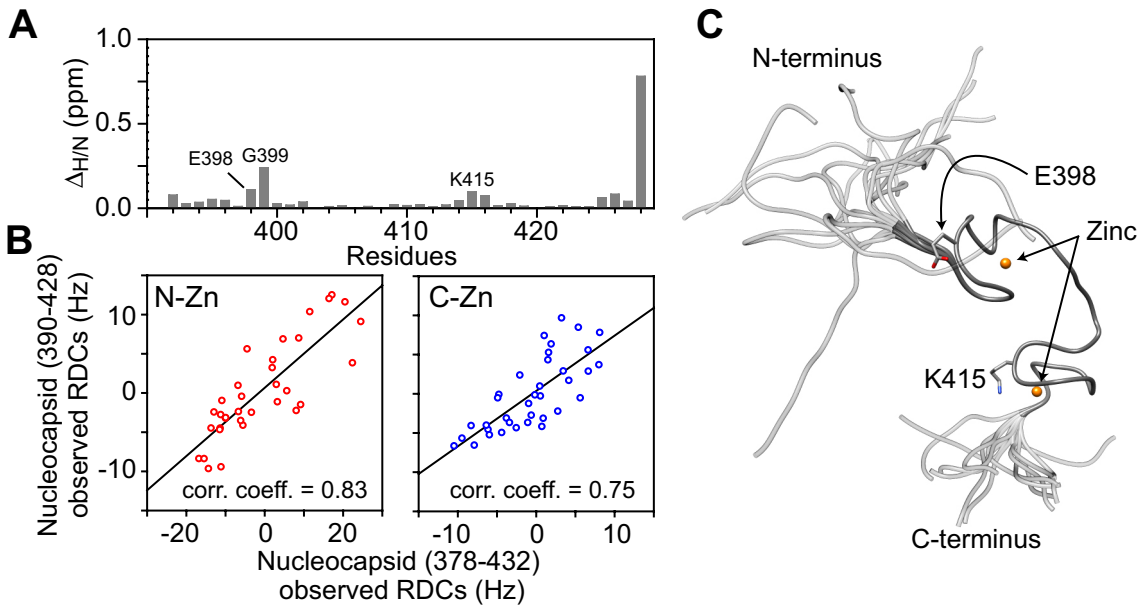

Figure S1. NMR of full-length (residues 378-432) and truncated (residues 390-428) HIV-1 nucleocapsid constructs. (A) Comparison of $^{1}H_N/^{15}N$ chemical shifts of full-length and truncated nucleocapsid constructs. Both spectra were acquired with identical acquisition parameters at 500 MHz (35°C) in 20 mM sodium phosphate, pH 6.5, 50 mM NaCl, 1 mM DTT and 0.1 mM $ZnCl_2$. Residues located in the N-Zn and C-Zn knuckles that exhibit the largest chemical shift differences are labeled. These chemical shift differences between the two constructs arise from weak interactions between the disordered termini and surface exposed regions of the zinc knuckles. The large chemical shift difference for Glu428 is simply due to the fact that this residue is at the C-terminus of the truncated construct. (B) Correlation between backbone RDCs of full-length (*x* axis) and truncated (*y* axis) constructs measured in a neutral alignment medium of bicelles. Left panel, N-Zn in red; right panel, C-Zn in blue. While a correlation is present, it is poor owing to the fact that the presence of tails at the N- and C-termini of the full-length construct alters the overall shape of the conformational ensemble. (C) Ribbon representation of HIV-1 nucleocapsid. Residues of truncated nucleocapsid (residues 390-428) are shown in solid grey ribbons while the flexible N- and C-termini are shown as semi-transparent gray ribbons. Side-chains of residues in the zinc knuckles that exhibit the largest chemical shift differences between the full-length and truncated construct are depicted in stick representation. Zinc atoms are shown as orange spheres.

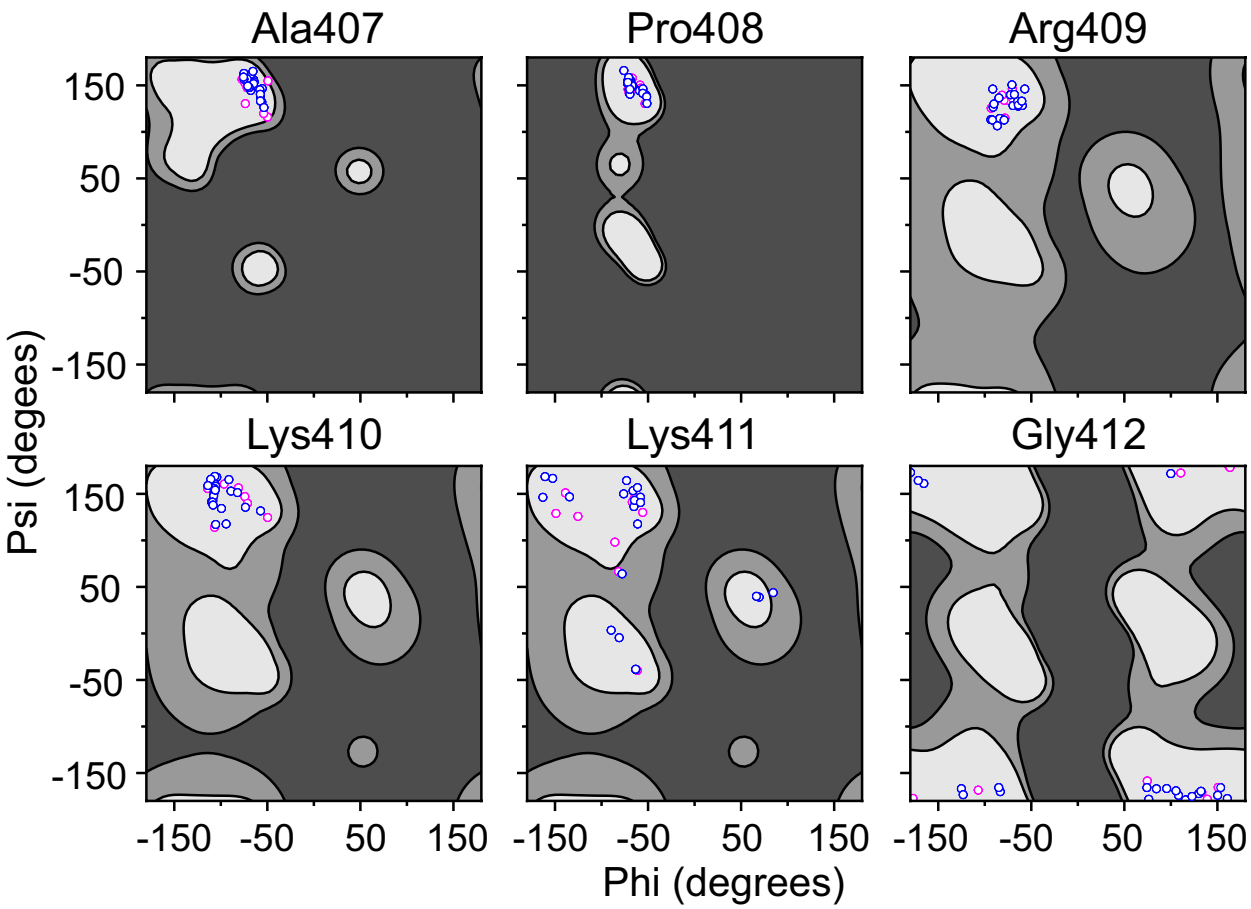

Figure S2. Distribution of ϕ and ψ backbone torsion angles of the 10 lowest energy truncated nucleocapsid ensembles for residues in the linker region (residue 407-412) that connect the N-Zn and C-Zn knuckles. The plot was generated using the ramaStrip script of Xplor-NIH.[8] The circles in magenta represent the three ensembles out of 10 where significant violations are observed for the N-C' and/or Cα-C' RDCs of Arg409 in the context of the full-length nucleocapsid in which the zinc knuckles and linker are fixed and only the tails (represented by a five-membered sub-ensemble for each ensemble member) are allowed to move during the course of RDC and SAXS-driven simulated annealing.

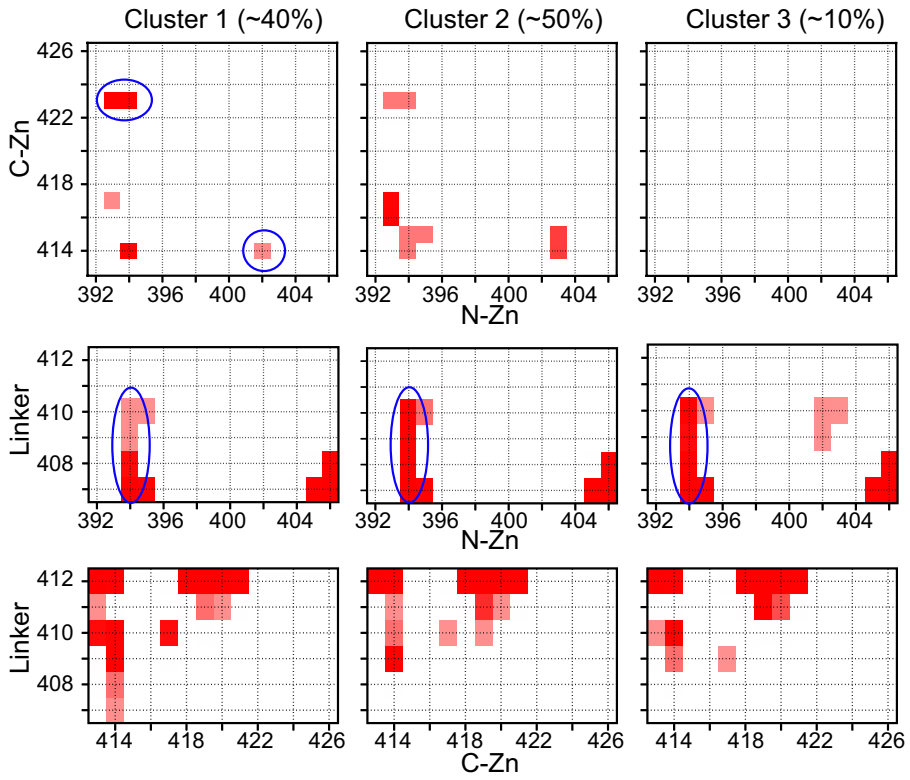

Figure S2. Summary of transient residue-residue contacts ≤4 Å found in the calculated nucleocapsid ensemble. The frequency of occurrence of a given contact within a particular cluster is shown as red scale with dark, medium and light red representing 70-100, 30-70 and 10-30% occurrence. The blue circles indicate weak inter-knuckle and knuckle-linker NOEs previously observed in 2D and 3D NOE spectra [15]. These include inter-knuckle contacts between Phe393-Met423 and Ala402-Trp414, and contacts between Asn394 of the N-Zn knuckle and residues within the linker. Note a few weak experimental inter-knuckle NOEs were also observed between Phe393-Trp414[15c,d] and these are represented in Cluster 1 at a contact cutoff ≤ 5 Å. Contact maps were generated using the contactMap helper script of Xplor-NIH.[8] The majority of residues involved in inter-knuckle and knuckle-linker contacts are highly conserved throughout all HIV-1 strains: ≥99% in the N-Zn knuckle except for Arg 403 (73%); ≥97% in the C-Zn knuckle except for Lys 418 (64%); and ≥ 98% in the linker except for Lys410 (93%) and Lys411 (89%).[16]

References

[1] J. R. Huth, C. A. Bewley, B. M. Jackson, A. G. Hinnebusch, G. M. Clore, A. M. Gronenborn, *Protein Sci.* **1997**, *6*, 2359-2364.

[2] L. Deshmukh, C. D. Schwieters, A. Grishaev, R. Ghirlando, J. L. Baber, G. M. Clore, *J. Am. Chem. Soc.* **2013**, *135*, 16133-16147.

[3] V. King, M. Parker, K. P. Howard, *Journal of magnetic resonance* **2000**, *142*, 177-182.

[4] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **1995**, *6*, 277-293.

[5] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Proteins* **2005**, *59*, 687-696.

[6] N. C. Fitzkee, A. Bax, *J. Biomol. NMR* **2010**, *48*, 65-70.
[7] a) J. J. Chou, F. Delaglio, A. Bax, *J. Biomol. NMR* **2000**, *18*, 101-105; b) C. P. Jaroniec, T. S. Ulmer, A. Bax, *J. Biomol. NMR* **2004**, *30*, 181-194.
[8] a) C. D. Schwieters, J. J. Kuszewski, N. Tjandra, G. M. Clore, *Journal of magnetic resonance* **2003**, *160*, 65-73; b) C. D. Schwieters, J. Kuszewski, G. M. Clore, *Progr. Nucl. Magn. Reson. Spectrosc* **2006**, *48*, 47-62.
[9] N. Goudreau, O. Hucke, A. M. Faucher, C. Grand-Maitre, O. Lepage, P. R. Bonneau, S. W. Mason, S. Titolo, *J. Mol. Biol.* **2013**, *425*, 1982-1998.
[10] a) J. R. Huang, S. Grzesiek, *J. Am. Chem. Soc.* **2010**, *132*, 694-705; b) V. Venditti, C. D. Schwieters, A. Grishaev, G. M. Clore, *Proc Natl Acad Sci U S A* **2015**, *112*, 11565-11570.
[11] C. D. Schwieters, G. M. Clore, *Prog. Nucl. Magn. Reson. Spectrosc.* **2014**, *80*, 1-11.
[12] G. A. Bermejo, G. M. Clore, C. D. Schwieters, *Protein Sci.* **2012**, *21*, 1824-1836.
[13] Y. Ryabov, J. Y. Suh, A. Grishaev, G. M. Clore, C. D. Schwieters, *J .Am .Chem. Soc.* **2009**, *131*, 9522-9531.
[14] M. Nilges, G. M. Clore, A. M. Gronenborn, *FEBS Lett.* **1988**, *239*, 129-136.
[15] a) N. Morellet, N. Jullian, H. De Rocquigny, B. Maigret, J.L. Darlix, B.P. Roques *EMBO J*, **1992**, *11*, 3059-3065. b) N. Morellet, H. de Rocquigny, Y. Mely, N. Jullian, H. Demene, M. Ottmann, D. Gerard, J.L. Darlix, M.C. Fournie-Zaluski, B.P. Roques, *J. Mol. Biol.* **1994**, *235*, 287-301. c) B.M. Lee, R.N. De Guzman, B.G. Turner, N. Tjandra, M.F. Summers, *J. Mol. Biol.* **1998**, *279*, 633-649. d) L. Zrgarian, C. Tisne, P. Barraud, X. Xu, N. Morellet, B. Rene, Y. Mely, P. Fosse, O Mauffret, *PLoS One* **2014**, *9*, e102150.
[16] HIV mutation browser. http://hivmut.org