BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

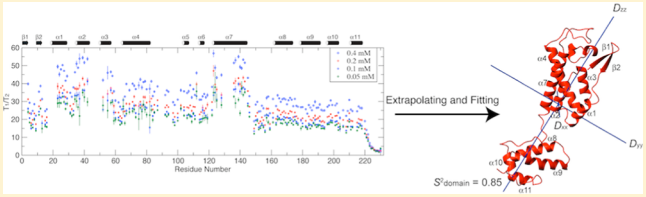
Determining Interdomain Structure and Dynamics of a Retroviral Capsid Protein in the Presence of Oligomerization: Implication for Structural Transition in Capsid Assembly

Kang Chen and Nico Tjandra*

Laboratory of Molecular Biophysics, Biochemistry and Biophysics Center, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, United States

Supporting Information

ABSTRACT: Capsid (CA) proteins from all retroviruses, including HIV-1, are structurally homologous dual-domain helical proteins. They form a capsid lattice composed of unitary symmetric CA hexamers. X-ray crystallography has shown that within each hexamer a monomeric CA adopts a single conformation, where most helices are parallel to the symmetry axis. In solution, large differences in averaged NMR spin relaxation rates for the two domains were observed, suggesting they are dynamically independent. One relevant question for the capsid assembly remains: whether the interdomain conformer within a hexamer unit needs to be induced or pre-exists within the conformational space of a monomeric CA. The latter seems more consistent with the relaxation data. However, possible CA protein oligomerization and the structure of each domain will affect relaxation measurements and data interpretation. This study, using CA proteins from equine infectious anemia virus (EIAV) as an example, demonstrates a linear extrapolation approach to obtain backbone ^{15}N spin relaxation time ratios T_1/T_2 for a monomeric EIAV-CA in the presence of oligomerization equilibrium. The interdomain motion turns out to be limited. The large difference in the domain averaged $\langle T_1/T_2 \rangle$ for a CA monomer is a consequence of the orthogonal distributions of helices in the two domains. The new monomeric interdomain conformation in solution is significantly different from that in CA hexamer. Therefore, if capsid assembly follows a nucleation−propagation process, the interdomain conformational change might be a key step during the nucleation, as the configuration in hexagonal assembly is never formed by diffusion of its two domains in solution.



Large proteins (>20 kDa) may consist of multiple domains. Individual domains can function as a scaffold or catalytic unit so that the full-length protein is more efficient in serving as a structural component or performing an enzymatic reaction, or both.[1] A retroviral capsid (CA) protein, consisting of two helical domains, is a typical example of the utilization of both domains to assemble a dense core to protect the retroviral genomic RNA in complex with the nucleocapsid (NC) protein. CA and NC proteins, in addition to matrix (MA) protein, are the three major proteolytic products of the precursor Gag poly protein formed during viral maturation. CA proteins from all species of retroviruses are 26−27 kDa in size and structurally homologous.[2] X-ray crystallography and cryo-electron-microscopy (EM) structures of CA proteins from N-tropic murine leukemia virus (N-MLV),[3] HIV-1,[4−6] and Rous sarcoma virus (RSV)[7] have revealed that the outer layer of the capsid core is composed of a hexagonal and pentagonal assembly of the seven-helix CA N-terminal domain (CA^N) and the inner layer is formed through dimerization of neighboring hexamers by the four-helix C-terminal domain (CA^C). The interdomain conformation within all known CA assemblies is unique, and most helices, helices α1−3 and α7 of the CA^N and helices α8−11 of the CA^C, are parallel to the hexamer symmetric axis. Although the interdomain structure of CA is available for the assembled lattice, the assembly process of CA itself is dynamic and less understood.[8] The proteolysis cleavage of Gag poly protein, a process that has been targeted by HIV-1 drugs, regulates the release of CA, but may or may not be directly related to the CA assembly.[9,10] Structure and dynamics studies of CA in solution can perhaps address a relevant question of whether the assembly-ready conformer has to be induced from or is already pre-existing in its monomer state in solution. Analogous to helix−coil transition[11] or protein−ligand conformational switch,[12] if a retroviral capsid assembly has to be induced, other factors, for example, pH or ligand− or protein−protein interaction, must be involved to create an activated conformer of CA from its equilibrium state, corresponding to the initial process of nucleation−propagation; on the contrary, pre-existing would suggest the assembly starts from a subpopulation of assembly-ready interdomain conformer within an equilibrated ensemble of CA structures in solution.

Generally, solution NMR can provide interdomain structural information, that is, interdomain NOEs, residual dipolar

Received: May 10, 2013
Revised: July 16, 2013
Published: July 19, 2013

dx.doi.org/10.1021/bi400592d | *Biochemistry* 2013, 52, 5365−5371

Table 1. Extended Model Free Fitting for ^{15}N T_1/T_2 of EIAV-CA

concn (mM)	domain	τ_m[a] (ns)	$2D_{zz}/(D_{yy} + D_{xx})$	D_{yy}/D_{xx}	S^2_{domain}	τ_{domain} (ns)	α (deg)	β (deg)	γ (deg)	χ^2/N
0	CA^N	13.6	1.84	1.69	1.00	n/a	155	112	86.9	15.6
	CA^C				0.854	3.86	74.9	18.7	39.6	
0.4	CA^N	18.6	1.70	1.51	1.00	n/a	158	112	82.2	60.9
	CA^C				0.900	1.87	61.4	18.2	50.5	

[a] τ_m was calculated from $1/(2D_{zz} + 2D_{yy} + 2D_{xx})$.

coupling (RDC),[13−16] and relaxation time ratios T_1/T_2.[17−22] NOEs are measurable only if the interdomain contacts involving protons are within a short distance and long-lived. Even when these requirements are fulfilled, the NOEs may not be measurable in large proteins, because deuteration may be required to obtain sharper resonance lines. RDCs are very useful in positioning domains relative to each other. If a multidomain protein is within oligomerization equilibrium, RDCs are averaged results of all possible oligomers exchanging up to a millisecond time scale and are strongly biased toward larger oligomers. RDC data interpretations quickly become difficult if the oligomerization is coupled to interdomain structural change. The ratio of backbone ^{15}N spin relaxation times T_1/T_2 is insensitive to fast local dynamics of the amide H−N bond[23] and reflects the overall rotational diffusion of the domain.[24−27] The average T_1/T_2 values as well as their variation can be used to estimate a domain rotational correlation time and its geometrical shape, respectively. The average $\langle T_1/T_2 \rangle$ of each domain can be directly used to indicate the presence of interdomain motion.[28,29]

The previous solution NMR studies on RSV-CA[28] have suggested that the motion of the two retroviral CA domains are independent of each other based on the lack of interdomain NOEs and the difference in ^{15}N T_1/T_2.[8] This observation is consistent with the pre-existing mechanism for CA assembly. However, CA proteins are likely to be in oligomerization equilibrium through domains CA^N or CA^C, and the measured T_1/T_2 might contain contributions from oligomers that can bias the relaxation values for each domain differently. For instance, the CA protein of equine infectious anemia virus (EIAV), belonging to the same lentivirus subfamily as HIV-1, weakly oligomerizes and dimerizes through its CA^N and CA^C domains, respectively, and these two equilibria appear to be independent processes as a result of the different motional time scales.[30] The assembled EIAV capsid is in the same conical shape as HIV-1 capsid.[31] The study here uses NMR ^{15}N relaxation and extrapolation data analysis to overcome the oligomerization problem and at the same time to quantitatively reveal if the assembly-ready interdomain conformer could be present in the conformational space of a monomeric EIAV-CA. Our results showed that the interdomain structure of EIAV-CA monomer is relatively rigid and different from the assembled structure, and the CA assembly should be induced by other factors.

MATERIALS AND METHODS

Backbone ^{15}N T_1 and T_2 Measurements. The expression and purification of $^{15}N/^2H$-labeled EIAV-CA protein were described.[30] Four samples at protein concentrations of 0.426, 0.213, 0.106, and 0.053 mM were prepared in a buffer of 20 mM potassium phosphate at pH 6.7. All spin relaxation measurements were performed at 27 °C on a Bruker Avance 600 MHz spectrometer equipped with a cryogenic probe. Pulse sequences with TROSY detection scheme and minimal water saturation were applied.[32] For T_1 measurements eight data points with relaxation delays of 80, 240, 420, 620, 840, 1140, 1460, and 1940 ms were collected. For $T_{1\rho}$ the ^{15}N spin−lock power was set at 2000 Hz, and eight data points with relaxation delays of 2, 7.2, 12.8, 20, 28.8, 39.2, 50.4, and 67.2 ms were collected. Both T_1 and $T_{1\rho}$ experiments were collected in an interleaved manner. NMRPipe was used to process all spectra.[33] Relaxation time constants and errors were obtained from exponential fitting on peak heights using Sparky (T. D. Goddard and D. G. Kneller, University of California, San Francisco). The calculation of T_2 using T_1 and $T_{1\rho}$ was previously described.[32]

Extrapolation and Extended Model Free Fitting of ^{15}N T_1/T_2. The relaxation time ratios T_1/T_2 at four different protein concentrations were linearly extrapolated. The extrapolated value at the zero concentration was termed T_1/T_2 at 0 mM. The error of the extrapolated T_1/T_2 was obtained from Monte Carlo steps. The extended model free (EMF) fitting of T_1/T_2 was carried out as described previously.[34] A fully anisotropic tensor was adopted as the overall rotational diffusion tensor. The domain coordinates were obtained from the crystal structure[35] with the refined β-hairpin motif.[30] The EMF fitting parameters included three principal components of the overall rotational diffusion tensor (D_{zz}, D_{yy}, and D_{xx}), two sets of Euler angles (α, β, and γ) defining the orientation of the overall diffusion tensor within each of the domain coordinates, and the domain motion order parameters S^2_{domain} and correlation time τ_{domain} that describe the amplitude and time scale of the domain motion within the overall diffusion tensor frame (Table 1). The χ^2 minimization was against the difference between the measured T_1/T_2 and the calculated values, that is, eq 1, where N (N = 151) is the total number of residues used for fitting and $\delta(T_1/T_2)$ is the experimental error. A simplex searching algorithm written in MATLAB (The Mathworks, Natick, MA, USA) was used to locate the global minimal.

$$\chi^2 = \left(\frac{1}{N}\right)\sum_{i=1}^{N}\left[\frac{(T_{1i}/T_{2i})_{calcd} - (T_{1i}/T_{2i})_{measd}}{\delta(T_{1i}/T_{2i})_{measd}}\right]^2 \quad (1)$$

RESULTS

Simulation and Extrapolation on Concentration-Dependent T_1/T_2. The observation of ^{15}N T_1/T_2 ratios as a function of protein concentration originates from weak protein oligomerization.[36] When a protein A is in a monomer to oligomer (A_n) equilibrium as indicated by eq2, two basic equations, eqs 3 and 4 exist, where c is the total protein concentration, K_d is the dissociation constant, and n is the number of monomers within one oligomer. Solving eqs 3 and 4 for $[A]$ and $[A_n]$ can be performed using Mathematica (Wolfram, IL, USA) for various oligomeric states. For examples, the analytical solutions for $[A]$ when n = 2 and 3 are given in eqs 5 and 6, respectively. With the knowledge of

dx.doi.org/10.1021/bi400592d | Biochemistry 2013, 52, 5365−5371

$[A]$, one can readily obtain the population fraction of monomer f_A in the equilibrium using eq 7 and calculate the averaged value $(T_1/T_2)_{av}$ according to eq 8, where $(T_1/T_2)_{mono}$ and $(T_1/T_2)_{oligo}$ are relaxation times ratios for the monomer and oligomer, respectively.

$$A_n \overset{K_d}{\leftrightarrow} nA \tag{2}$$

$$c = [A] + n[A_n] \tag{3}$$

$$K_d = \frac{[A]^n}{[A_n]} \tag{4}$$

$$[A] = \sqrt{K_d(c/2 + K_d/16)} - K_d/4 \tag{5}$$

$$n = 2$$

$$[A] = K_d/(27B) - 3B$$

$$B = \{c^3/18 - c(c^2/3 + K_d/27)/6 + [K_d^3/531441 + (K_d c/162)^2]^{1/2}\}^{1/3} \tag{6}$$

$$n = 3$$

$$f_A = \frac{[A]}{c} \tag{7}$$

$$\left(\frac{T_1}{T_2}\right)_{av} = f_A\left(\frac{T_1}{T_2}\right)_{mono} + (1 - f_A)\left(\frac{T_1}{T_2}\right)_{oligo} \tag{8}$$

Now we consider simulating a system of monomer−dimer (n = 2) equilibrium with a weak dissociation constant K_d of 5 mM and $(T_1/T_2)_{mono}$ and $(T_1/T_2)_{dimer}$ of 10 and 20, respectively. Using eqs 5, 7, and 8, we calculated and plotted $(T_1/T_2)_{av}$ at each concentration of 0.05, 0.1, 0.2, and 0.4 mM (Figure 1A), which were chosen to closely match the current experimental conditions. The linear and quadratic fittings yielded $(T_1/T_2)_{av}$-intercept values of 10.07 and 10.01, respectively. Such intercept values are taken as extrapolated data points, both of which are within 5%, roughly the experimental error range, of the target $(T_1/T_2)_{mono}$ value of 10. We then fit another system of monomer−trimer (n = 3) equilibrium with a weak dissociation constant K_d of 0.125 mM^2 and $(T_1/T_2)_{mono}$ and $(T_1/T_2)_{trimer}$ of 10 and 30, respectively. Here we have assumed that there is one monomer−monomer interface within a dimer and three interfaces within a trimer; therefore, the dissociation constant for the trimer is chosen to correspond to 3 times the free energy change of the dissociation of the dimer ($-RT$ ln K_d). Using eqs 6, 7, and 8 we again calculated and plotted $(T_1/T_2)_{av}$ at each concentration of 0.05, 0.1, 0.2, and 0.4 mM (Figure 1B). The linear and quadratic fittings yielded $(T_1/T_2)_{av}$-intercept values of 10.39 and 8.800, respectively. Only the linear fitting intercept value of 10.39 is within 5% of the targeted $(T_1/T_2)_{mono}$ value of 10.

Of course, our simulation is rudimentary and has a number assumptions. However, it does show that simpler linear fitting could be a more reliable method to obtain monomeric relaxation values when only data from a small number of concentrations are available. This constitutes one reason we adopted the linear extrapolation for the study.

Concentration-Dependent T_1/T_2 of EIAV-CA and Its Extrapolation. The ^{15}N T_1/T_2 ratios were measured at four EIAV-CA concentrations, 0.4, 0.2, 0.1, and 0.05 mM, showing strong concentration dependence for residues in structured

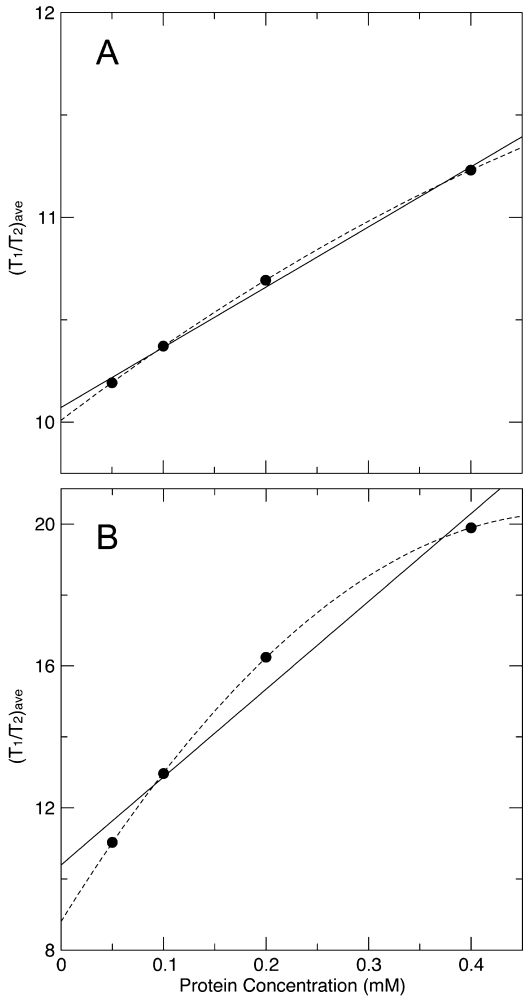


Figure 1. Linear (solid lines) and quadratic (dashed lines) fittings on simulated $(T_1/T_2)_{av}$ data points (solid dots) under weak dimerization (A) and trimerization (B) equilibria. The target $(T_1/T_2)_{av}$-intercept value is 10.

regions of the protein (Figures 2 and 3; Table S1 of the Supporting Information). The full-length EIAV-CA protein has a total of 212 non-proline residues. Backbone ^{1}H−^{15}N resonances for 186 residues were assigned, and missing residues are I2-M3, R18-T22, I43-E51, M53, N54, G65, Q127, and Y129−I134, which are located along the domain CA^N oligomeric interfaces. The previous R_2-CPMG dispersion results demonstrated residues with significant microsecond−millisecond exchange contributions are T16, D82, L87, and M102, the last three being in the cyclophilin A (CypA) binding loop.[30] Those residues undergoing chemical exchange or missing have been removed from the analysis. Furthermore, residues with significantly low signal-to-noise ratios, yielding large errors in their relaxation times, and the C-terminal flexible residues, displaying no concentration dependence, were removed. Finally, 90 and 61 total residues for domains CA^N and CA^C, respectively, were left for T_1/T_2 analysis, which are tabulated in the Supporting Information (Table S1). To obtain proper ^{15}N T_1/T_2 for the monomeric EIAV-CA, we applied the linear as well as the quadratic extrapolations using all four sets of T_1/T_2. For all 151 data points the quadratic fitting did not yield statistically better fitting, judged by the F test (data not shown); therefore, only the linear extrapolation was applied

dx.doi.org/10.1021/bi400592d | Biochemistry 2013, 52, 5365−5371

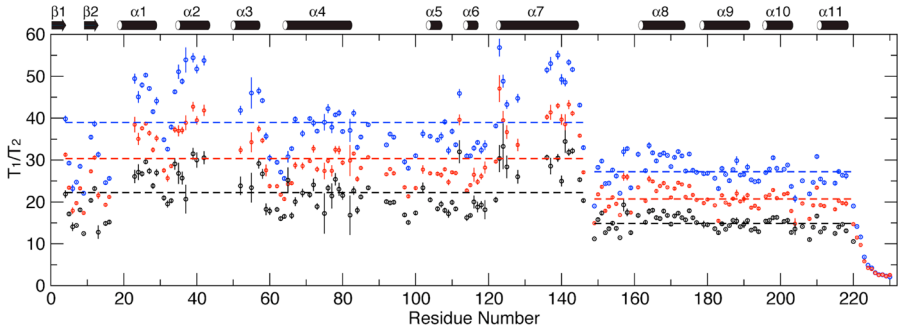


Figure 2. Measured (for clarity, only the 0.4 and 0.2 mM data sets are shown in blue and red, respectively) and extrapolated (0 mM, black) backbone ^{15}N spin relaxation time ratios T_1/T_2. The dashed lines mark the average $\langle T_1/T_2 \rangle$ for each domain in each measurement and end at domain boundaries.

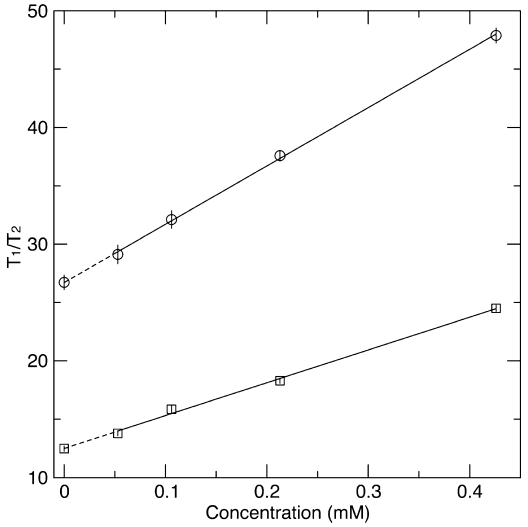


Figure 3. Examples of linear extrapolation. The T_1/T_2 ratios are plotted as a function of protein concentration for N25 (circle) and M215 (square).

(examples of fitting are shown in Figure 3). The obtained new relaxation time ratios from the linear extrapolation were termed T_1/T_2 at 0 mM and assumed to be the corresponding value in the monomeric state (Figure 2).

Empirical Estimation on $\langle T_1/T_2 \rangle$ Yields Large Scale of Interdomain Motion. Similar to the previous observations on RSV-CA,[28] the domain averaged $\langle T_1/T_2 \rangle$ values differ significantly for all measured and extrapolated data sets (Figure 2). The size ratio of the domain EIAV-CA^N (residues 1−146) and EIAV-CA^C (residues 149−220) is 2:1. If the two domains are completely independent (S^2_{domain} = 0), under the assumptions of isotropic diffusion and isotropic distribution of H−N vectors, the $\langle T_1/T_2 \rangle$ value of domain CA^C should be 50% of the $\langle T_1/T_2 \rangle$ value of domain CA^N, roughly scaled with the relative domain sizes. For the 0 mM data set, the $\langle T_1/T_2 \rangle$ value of domain CA^C is 14.86, 66% of the $\langle T_1/T_2 \rangle$ value of domain CA^N of 22.26 (Figure 2). The large difference in domain $\langle T_1/T_2 \rangle$ suggests the presence of a large-scale interdomain motion based on the assumption of the isotropic diffusion in each domain.

Numeric Fitting on T_1/T_2 Results in Limited Interdomain Motion. We fit the 0 mM T_1/T_2 data set using the EMF approach described previously (Table 1).[34] The resulting diffusion anisotropy ($2D_{zz}/(D_{xx} + D_{yy})$) and asymmetry ($D_{yy}/D_{xx}$) are large, 1.84 and 1.69, respectively, indicating the 0 mM T_1/T_2 values are very sensitive to structure coordinates. The S^2_{domain} values for both domains are high, 1 and 0.854 for domains CA^N and CA^C, respectively, meaning interdomain motion is limited and the empirically estimated large scale of domain motion is not correct. The reason for the large difference in domain $\langle T_1/T_2 \rangle$ was obvious once we rotated and positioned the two domains into the overall diffusion tensor (Figure 4). It should be noted the simple domain rotation

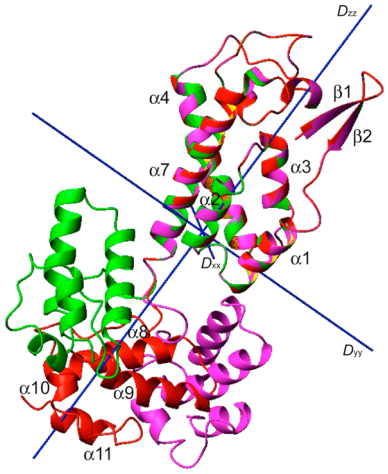


Figure 4. Alignment of domain CA^N within the dual-domain EIAV-CA structure from the current T_1/T_2 analysis (red), monomeric crystal structure (magenta, 2EIA-B), and a monomer within the HIV-1-CA hexamer crystal structure (green, 3GV2).

generated structure shown in Figure 4 represents only a mean interdomain conformation, not a structure ensemble from any simulated annealing calculations. The helices of domain CA^C are all perpendicular to the long axis (D_{zz}) of the diffusion tensor, sampling the fastest components of the overall motion; in contrast, helices α1−3 and α7 of domain CA^N run parallel to the long axis, experiencing slower tumbling from D_{yy} and D_{xx}. One cross-validation of the domain positioning result is to compare the T_1/T_2 values of the N-terminal β-hairpin residues, representing only 4% of the T_1/T_2 data set, to the rest of the data. The N-terminal β-hairpin, although well ordered within the domain CA^N,[37−39] is close in T_1/T_2 values to domain CA^C. This could be true only if the two domains share the same diffusion tensor and the amide bond vectors of the β-hairpin are oriented similar to helices of domain CA^C (Figures 2 and 4).

In parallel, the EMF fit for T_1/T_2 values of the 0.4 mM data set was performed (Table 1). The nearly 4-fold higher χ^2 value

dx.doi.org/10.1021/bi400592d | Biochemistry 2013, 52, 5365−5371

indicates the fitting quality is much worse than the extrapolated 0 mM data set. The overall correlation time of 18.6 ns is significantly higher than the monomer value of 13.6 ns. The lower diffusion anisotropy and asymmetry indicate less sensitivity of the 0.4 mM data set to structural coordinates. The lower anisotropy/asymmetry also suggested a more compact conformer was formed upon oligomerization. The S^2_{domain} of CA^C is higher, indicating even less interdomain motion in the oligomer. The D_{zz} axes of the fit overall diffusion tensors differ by 2.8° and 4.3° for domains CA^N and CA^C, respectively, from those of the monomeric tensor, indicating the oligomer (mostly dimer here) tensor direction is not significantly different from that of the monomer tensor.

Validation of T_1/T_2-Derived Interdomain Structure. The new interdomain conformation obtained using the extrapolated T_1/T_2 is different from that in the crystal structures of monomer EIAV-CA[35] or hexamer structure of HIV-1-CA (Figure 4).[5] The hexamer is the unit for capsid assembly of all retroviruses. This inconsistency in the interdomain conformations prompted us to inspect if the T_1/T_2-derived interdomain conformation in solution is a true representation. To further investigate this question, we performed a small-angle X-ray scattering (SAXS) experiment at concentration of 0.05 mM. The SAXS profile showed the T_1/T_2-derived CA structure provides a better fit than the crystal structures of either monomer or a monomer unit within a hexamer (Figure 5). As could be expected, the fit for the T_1/T_2-derived CA structure is not perfect at the long-range distance region, indicating the presence of a tiny population of oligomer even at this diluted concentration.

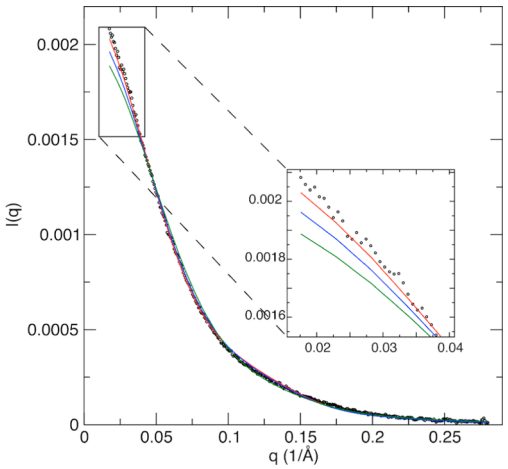


Figure 5. Experimental (dots) and fitting curves (solid) of small-angle X-ray scattering. The structural coordinates of the current T_1/T_2-determined (red), the monomeric EIAV-CA (2EIA, blue), and one monomer in hexameric HIV-1-CA (3GV2, green) were used for fittings. The EIAV-CA sample for SAXS measurement is 0.05 mM in the same buffer as NMR samples. Fitting of the structural coordinates was performed using XPLOR-NIH.[51].

DISCUSSION

In summary, we have presented a straightforward method to extract the mean interdomain structure and dynamics information from backbone ^{15}N T_1/T_2 data on a protein undergoing oligomerization. On the basis of simulations and statistical tests we suggested the linear extrapolation, as opposed to a combination of linear and quadratic fittings,[36] should be applied to recover the monomer relaxation data at zero concentration. If the contribution from protein oligomerization was not removed, the data fitting quality would be worse, meaning less agreement between structural coordinates and measured relaxation data. Worse fitting will also cause errors in the resulting parameters; that is, the correlation time would be higher, the diffusion tensor anisotropy and asymmetry could be wrong, the interdomain motion S^2_{domain} could be different, perhaps resulting in an inaccurate biological conclusion, and the tensor direction might not be representative of the monomer if the oligomer tensor direction is significantly different. As has been shown previously, diffusion tensor anisotropy and asymmetry are a few long-range restraints for protein shape that can be used in NMR structure calculation.[25,27,40] Although not being tested here, any inaccuracy in the diffusion tensor can bring errors in refining a monomeric protein structure and domain positioning of a multidomain protein. In theory, we might be able to extend the usage of the precisely measured ^{15}N T_1/T_2 to derive protein quaternary structure of oligomers. However, practical methods to take into account the exchange kinetics if it occurs in the nanosecond time scale are not currently available, although theoretical treatments do exist.[41,42]

Although the interdomain dynamics we identified on EIAV-CA using the EMF method is relatively restricted with a higher order parameter S^2_{domain} of 0.85, it should not be deemed as the limit of the EMF method. Previous simulations and measurements suggested the EMF fitting approach could be valid for larger domain motions with S^2_{domain} as low as 0.4−0.5, corresponding to a semicone angle of 45° in the diffusion in the cone model.[43,44]

Our structural data have an implication on the nucleation of a retroviral CA assembly. The CA assembly is not an isolated event, which results from the proteolytic processing of its precursor Gag, and rather complicated. The nucleation step in capsid assembly, although generally accepted,[45] has not been clearly demonstrated by any physical observations. In fact, the *in vitro* assembly of a cone-shaped HIV-1 capsid using recombinant protein was only successful with a combination of its precursor CA-NC protein and RNA, not CA alone.[46,47] Furthermore, a close inspection of the recent 8 Å cryo-EM structure of an immature Mason−Pfizer monkey virus (MPMV) CA-NC assembly (PDB id 4ARG)[47] revealed an interdomain orientation similar to the one in a matured HIV-1-CA hexamer (PDB id 3GV2).[5] The difference lies in a translational movement of the CA^C or the CA^N domain. These results might suggest the capsid assembly proceeds directly from its assembled proteolyitc fragment CA-NC proteins during maturation without major interdomain reorientation. Therefore, the CA assembly probably does not require the diffusion and nucleation−propagation in solution. A contrasting view can be made from the fact that only a small fraction of CA molecules processed from Gag were utilized for assembly,[48] and the observations of the dual-core HIV-1 virion[49] suggest a CA release and reassembly with several nucleation centers in solution have to take place. These observations argue against direct CA condensation from lattices of Gag or CA-NC. Our T_1/T_2-derived interdomain structure suggested that the capsid assembly, if it is occurring in solution, does not start readily from a pre-existing assembly-ready conformer; rather, a nucleation step involving at least significant interdomain structure reorientation appears to be necessary. This finding implies that factors other than the CA protein itself might play

dx.doi.org/10.1021/bi400592d | *Biochemistry* 2013, 52, 5365−5371

a role in promoting capsid assembly. This is in line with earlier observations where neutral pH and high salt concentration helped nucleation,[50] which had been shown to be important for *in vitro* assembly of tubular lattice using recombinant HIV-1-CA proteins.[45] The essential factors capable of inducing an authentic cone-shaped lattice are not known, but are potential for therapeutic targets.

■ ASSOCIATED CONTENT

Supporting Information

Table S1. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*(N.T.) Phone: (301) 402-3029. E-mail: tjandran@nhlbi.nih.gov.

Funding

The project was funded by the Intramural Research Programs of the National Heart, Lung, and Blood Institute of the NIH.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Alex Grishaev for measuring SAXS data and Charles Schwieters for help with XPLOR-NIH.

■ ABBREVIATIONS USED

CA, capsid; EIAV, equine infectious anemia virus; EMF, extended model free; NC, nucleocapsid

■ REFERENCES

(1) Petsko, G. A., and Ringe, D. (2004) *Protein Structure and Function*; New Science Press in association with Sinauer Associates and Blackwell Science, London, UK.

(2) Turner, B. G., and Summers, M. F. (1999) Structural biology of HIV. *J. Mol. Biol. 285*, 1−32.

(3) Mortuza, G. B., Haire, L. F., Stevens, A., Smerdon, S. J., Stoye, J. P., and Taylor, I. A. (2004) High-resolution structure of a retroviral capsid hexameric amino-terminal domain. *Nature 431*, 481−485.

(4) Ganser-Pornillos, B. K., Cheng, A., and Yeager, M. (2007) Structure of full-length HIV-1CA: a model for the mature capsid lattice. *Cell 131*, 70−79.

(5) Pornillos, O., Ganser-Pornillos, B. K., Kelly, B. N., Hua, Y. Z., Whitby, F. G., Stout, C. D., Sundquist, W. I., Hill, C. P., and Yeager, M. (2009) X-ray structures of the hexameric building block of the HIV capsid. *Cell 137*, 1282−1292.

(6) Byeon, I. J. L., Meng, X., Jung, J. W., Zhao, G. P., Yang, R. F., Ahn, J. W., Shi, J., Concel, J., Aiken, C., Zhang, P. J., and Gronenborn, A. M. (2009) Structural convergence between cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell 139*, 780−790.

(7) Cardone, G., Purdy, J. G., Cheng, N. Q., Craven, R. C., and Steven, A. C. (2009) Visualization of a missing link in retrovirus capsid assembly. *Nature 457*, 694−698.

(8) Briggs, J. A. G., and Krausslich, H. G. (2011) The molecular architecture of HIV. *J. Mol. Biol. 410*, 491−500.

(9) Lanman, J., Lam, T. T., Emmett, M. R., Marshall, A. G., Sakalian, M., and Prevelige, P. E. (2004) Key interactions in HIV-1 maturation identified by hydrogen-deuterium exchange. *Nat. Struct. Mol. Biol. 11*, 676−677.

(10) Ganser-Pornillos, B. K., Yeager, M., and Pornillos, O. (2012) Assembly and architecture of HIV. *Adv. Exp. Med. Biol. 726*, 441−465.

(11) Zimm, B. H., and Bragg, J. K. (1959) Theory of the phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys. 31*, 526−535.

(12) Frauenfelder, H., Sligar, S. G., and Wolynes, P. G. (1991) The energy landscapes and motions of proteins. *Science 254*, 1598−1603.

(13) Yuwen, T., Post, C. B., and Skrynnikov, N. R. (2011) Domain cooperativity in multidomain proteins: what can we learn from molecular alignment in anisotropic media? *J. Biomol. NMR 51*, 131−150.

(14) Prestegard, J. H., Al-Hashimi, H. M., and Tolman, J. R. (2000) NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q. Rev. Biophys. 33*, 371−424.

(15) Dosset, P., Hus, J. C., Marion, D., and Blackledge, M. (2001) A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J. Biomol. NMR 20*, 223−231.

(16) Ryabov, Y., and Fushman, D. (2006) Analysis of interdomain dynamics in a two-domain protein using residual dipolar couplings together with N-15 relaxation data. *Magn. Reson. Chem. 44*, S143−S151.

(17) Fushman, D., Xu, R., and Cowburn, D. (1999) Direct determination of changes of interdomain orientation on ligation: use of the orientational dependence of N-15 NMR relaxation in Abl SH(32). *Biochemistry 38*, 10225−10230.

(18) Ryabov, Y. E., and Fushman, D. (2007) A model of interdomain mobility in a multidomain protein. *J. Am. Chem. Soc. 129*, 3315−3327.

(19) Bruschweiler, R., Liao, X. B., and Wright, P. E. (1995) Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. *Science 268*, 886−889.

(20) Ulmer, T. S., Werner, J. M., and Campbell, I. D. (2002) SH3-SH2 domain orientation in Src kinases: NMR studies of Fyn. *Structure 10*, 901−911.

(21) Ryabov, Y., and Fushman, D. (2007) Structural assembly of multidomain proteins and protein complexes guided by the overall rotational diffusion tensor. *J. Am. Chem. Soc. 129*, 7894−7902.

(22) Ryabov, Y., and Fushman, D. (2006) Interdomain mobility in di-ubiquitin revealed by NMR. *Proteins 63*, 787−796.

(23) Kay, L. E., Torchia, D. A., and Bax, A. (1989) Backbone dynamics of proteins as studied by N-15 inverse detected heteronuclear NMR-spectroscopy − application to staphylococcal nuclease. *Biochemistry 28*, 8972−8979.

(24) Palmer, A. G. (2004) NMR characterization of the dynamics of biomacromolecules. *Chem. Rev. 104*, 3623−3640.

(25) Tjandra, N., Garrett, D. S., Gronenborn, A. M., Bax, A., and Clore, G. M. (1997) Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. *Nat. Struct. Biol. 4*, 443−449.

(26) Ryabov, Y., Clore, G. M., and Schwieters, C. D. (2010) Direct use of N-15 relaxation rates as experimental restraints on molecular shape and orientation for docking of protein-protein complexes. *J. Am. Chem. Soc. 132*, 5987−5989.

(27) Ryabov, Y., Suh, J. Y., Grishaev, A., Clore, G. M., and Schwieters, C. D. (2009) Using the experimentally determined components of the overall rotational diffusion tensor to restrain molecular shape and size in NMR structure determination of globular proteins and protein-protein complexes. *J. Am. Chem. Soc. 131*, 9522−9531.

(28) Campos-Olivas, R., Newman, J. L., and Summers, M. F. (2000) Solution structure and dynamics of the Rous sarcoma virus capsid protein and comparison with capsid proteins of other retroviruses. *J. Mol. Biol. 296*, 633−649.

(29) de Alba, E. (2009) Structure and interdomain dynamics of apoptosis-associated Speck-like protein containing a CARD (ASC). *J. Biol. Chem. 284*, 32932−32941.

(30) Chen, K., Piszczek, G., Carter, C., and Tjandra, N. (2013) The maturational refolding of the beta-hairpin motif of equine infectious anemia virus capsid protein extends its helix alpha 1 at capsid assembly locus. *J. Biol. Chem. 288*, 1511−1520.

(31) Roberts, M. M., and Oroszlan, S. (1989) The preparation and biochemical-characterization of intact capsids of equine infectious-anemia virus. *Biochem. Biophys. Res. Commun. 160*, 486−494.
(32) Chen, K., and Tjandra, N. (2011) Water proton spin saturation affects measured protein backbone N-15 spin relaxation rates. *J. Magn. Reson. 213*, 151−157.
(33) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRpipe − a multidimensional spectral processing system based on Unix pipes. *J. Biomol. NMR 6*, 277−293.
(34) Baber, J. L., Szabo, A., and Tjandra, N. (2001) Analysis of slow interdomain motion of macromolecules using NMR relaxation data. *J. Am. Chem. Soc. 123*, 3953−3959.
(35) Jin, Z. M., Jin, L., Peterson, D. L., and Lawson, C. L. (1999) Model for lentivirus capsid core assembly based on crystal dimers of EIAV p26. *J. Mol. Biol. 286*, 83−93.
(36) Jensen, M. R., Kristensen, S. M., Keeler, C., Christensen, H. E. M., Hodsdon, M. E., and Led, J. J. (2008) Weak self-association of human growth hormone investigated by nitrogen-15 NMR relaxation. *Proteins 73*, 161−172.
(37) Gitti, R. K., Lee, B. M., Walker, J., Summers, M. F., Yoo, S., and Sundquist, W. I. (1996) Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science 273*, 231−235.
(38) Macek, P., Chmelik, J., Krizova, I., Kaderavek, P., Padrta, P., Zidek, L., Wildova, M., Hadravova, R., Chaloupkova, R., Pichova, I., Ruml, T., Rumlova, M., and Sklenar, V. (2009) NMR structure of the N-terminal domain of capsid protein from the Mason-Pfizer monkey virus. *J. Mol. Biol. 392*, 100−114.
(39) Cornilescu, C. C., Bouamr, F., Carter, C., and Tjandra, N. (2003) Backbone N-15 relaxation analysis of the N-terminal domain of the HTLV-I capsid protein and comparison with the capsid protein of HIV-1. *Protein Sci. 12*, 973−981.
(40) Ryabov, Y. E., Geraghty, C., Varshney, A., and Fushman, D. (2006) An efficient computational method for predicting rotational diffusion tensors of globular proteins using an ellipsoid representation. *J. Am. Chem. Soc. 128*, 15432−15444.
(41) Ryabov, Y.; Clore, G. M.; Schwieters, C. D. Coupling between internal dynamics and rotational diffusion in the presence of exchange between discrete molecular conformations. *J. Chem. Phys.* 2012, *136*, 034108.
(42) Wong, V., Case, D. A., and Szabo, A. (2009) Influence of the coupling of interdomain and overall motions on NMR relaxation. *Proc. Natl. Acad. Sci. U.S.A. 106*, 11016−11021.
(43) Chang, S. L., Szabo, A., and Tjandra, N. (2003) Temperature dependence of domain motions of calmodulin probed by NMR relaxation at multiple fields. *J. Am. Chem. Soc. 125*, 11379−11384.
(44) Chen, K., and Tjandra, N. (2008) Extended model free approach to analyze correlation functions of multidomain proteins in the presence of motional coupling. *J. Am. Chem. Soc. 130*, 12745−12751.
(45) Barklis, E., Alfadhli, A., McQuaw, C., Yalamuri, S., Still, A., Barklis, R. L., Kukull, B., and Lopez, C. S. (2009) Characterization of the in vitro HIV-1 capsid assembly pathway. *J. Mol. Biol. 387*, 376−389.
(46) Ganser, B. K., Li, S., Klishko, V. Y., Finch, J. T., and Sundquist, W. I. (1999) Assembly and analysis of conical models for the HIV-1 core. *Science 283*, 80−83.
(47) Bharat, T. A. M., Davey, N. E., Ulbrich, P., Riches, J. D., de Marco, A., Rumlova, M., Sachse, C., Ruml, T., and Briggs, J. A. G. (2012) Structure of the immature retroviral capsid at 8 angstrom resolution by cryo-electron microscopy. *Nature 487*, 385−389.
(48) Briggs, J. A. G., Simon, M. N., Gross, I., Krausslich, H. G., Fuller, S. D., Vogt, V. M., and Johnson, M. C. (2004) The stoichiometry of Gag protein in HIV-1. *Nat. Struct. Mol. Biol. 11*, 672−675.
(49) Briggs, J. A. G., Wilk, T., Welker, R., Krausslich, H. G., and Fuller, S. D. (2003) Structural organization of authentic, mature HIV-1 virions and cores. *EMBO J. 22*, 1707−1715.
(50) Ehrlich, L. S., Agresta, B. E., and Carter, C. A. (1992) Assembly of recombinant human-immunodeficiency-virus type-1 capsid protein in vitro. *J. Virol. 66*, 4874−4883.
(51) Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson. 160*, 65−73.