JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

HIV-1 Nucleocapsid Proteins as Molecular Chaperones for Tetramolecular Antiparallel G-Quadruplex Formation

Arivazhagan Rajendran,[†,⊥] Masayuki Endo,*[,§,¶] Kumi Hidaka,[†] Phong Lan Thao Tran,[∥,▽] Jean-Louis Mergny,[∥] Robert J. Gorelick,[#] and Hiroshi Sugiyama*[,†,§,¶]

[†]Department of Chemistry, Graduate School of Science, Kyoto University, Kitashirakawa-oiwakecho, Sakyo-ku, Kyoto 606-8502, Japan

[§]Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Yoshida-ushinomiyacho, Sakyo-ku, Kyoto 606-8501, Japan

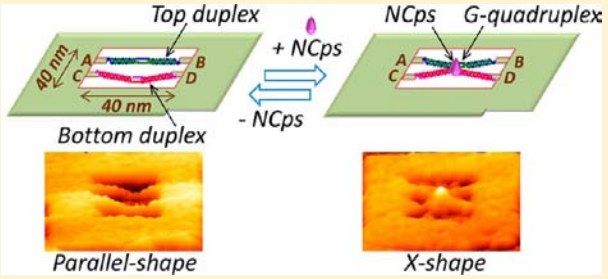
[¶]CREST, Japan Science and Technology Corporation (JST), Sanbancho, Chiyoda-ku, Tokyo 102-0075, Japan

[∥]INSERM, U869, ARNA Laboratory, Université de Bordeaux, 2 rue Robert Escarpit, Pessac F-33607, France

[#]Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, P.O. Box B, Frederick, Maryland 21702-1201, United States

Supporting Information

ABSTRACT: HIV-1 nucleocapsid proteins (NCps) facilitate remodeling of nucleic acids to fold thermodynamically stable conformations, and thus called nucleic acid chaperones. To date only little is known on the stoichiometry, NCp−NCp interactions, chaperone activity on G-quadruplex formation, and so on. We report here the direct and real-time analysis on such properties of proteolytic intermediate NCp15 and mature NCp7 using DNA origami. The protein particles were found to predominantly exist in monomeric form, while dimeric and multimeric forms were also observed both in free solution and bound to the quadruplex structure. The formation and the dissociation events of the G-quadruplexes were well documented in real-time and the intermediate-like states were also visualized. We anticipate that this pioneering study will strengthen our understanding on the chaperone activity of HIV-1 proteins which in turn will be helpful for the drug design based on G-quadruplex and also for the development of drugs against AIDS.



INTRODUCTION

Among the noncanonical secondary structures of DNA,[1,2] G-quadruplexes formed by the G-rich sequences are one of the exciting therapeutic targets since it was initially identified to be formed by the TTAGGG tandem repeats of telomeric DNA sequences. Apart from telomeres, the quadruplex sequences are also found in the promoter regions of several proto-oncogenes,[3] and it has been estimated that more than 376,000 G-quadruplexes are present in the human genome.[3] These sequences are also found in several other organisms.[4] Although concrete evidence on the formation of G-quadruplexes inside a cell is still a matter of debate,[5] these structures have been assumed to be involved in various cellular processes including chromosomal alignment, replication, transcription, genome recombination, and control over cancer cell proliferation.[6] For instance, it has been reported that more than 40% of human protein-coding genes contain at least one G-quartet motif in their promoter regions which indicates a possible role in mediating the activation or inhibition of transcription.[7] Also, 37% of recombination hotspots in the human genome contain at least one G-quartet motif, compared with 13.8% of the coldspots.[8] Further, the potential significance of G-rich sequences in various biological processes is indirectly supported by the identification of G-quadruplex-binding proteins. The β-subunit of telomere-binding protein from *Oxytricha*,[9] repressor-activator protein 1 (RAP1),[10] unimolecular quadruplex telomere-binding protein 25 (uqTBP25, a rat liver nuclear protein),[11] yeast meiosis-specific protein (Hop1),[12] topoisomerase I,[13] and Ku autoantigen[14] have been shown to facilitate the formation of G-quadruplexes and thus called G-quadruplex structure chaperones. Similarly, the G-quadruplex unwinding proteins, the viral SV40 large T-antigen helicase,[15] the *Saccharomyces cerevisiae* Sgs1,[16] Pif1 helicase,[17] human Werner's and Bloom's syndrome[18] helicases were also identified. These findings further support the hypothesis that the assembly and disassembly of G-quadruplex structures may play vital roles inside cells.

In addition to the biological significance and therapeutic target for apoptosis, G-rich sequences are shown to be the inhibitors of human immunodeficiency virus type 1 (HIV-1) replication in culture.[19] Cell culture experiments revealed that

Received: September 8, 2013
Published: November 13, 2013

   dx.doi.org/10.1021/ja409085j | J. Am. Chem. Soc. 2013, 135, 18575−18585

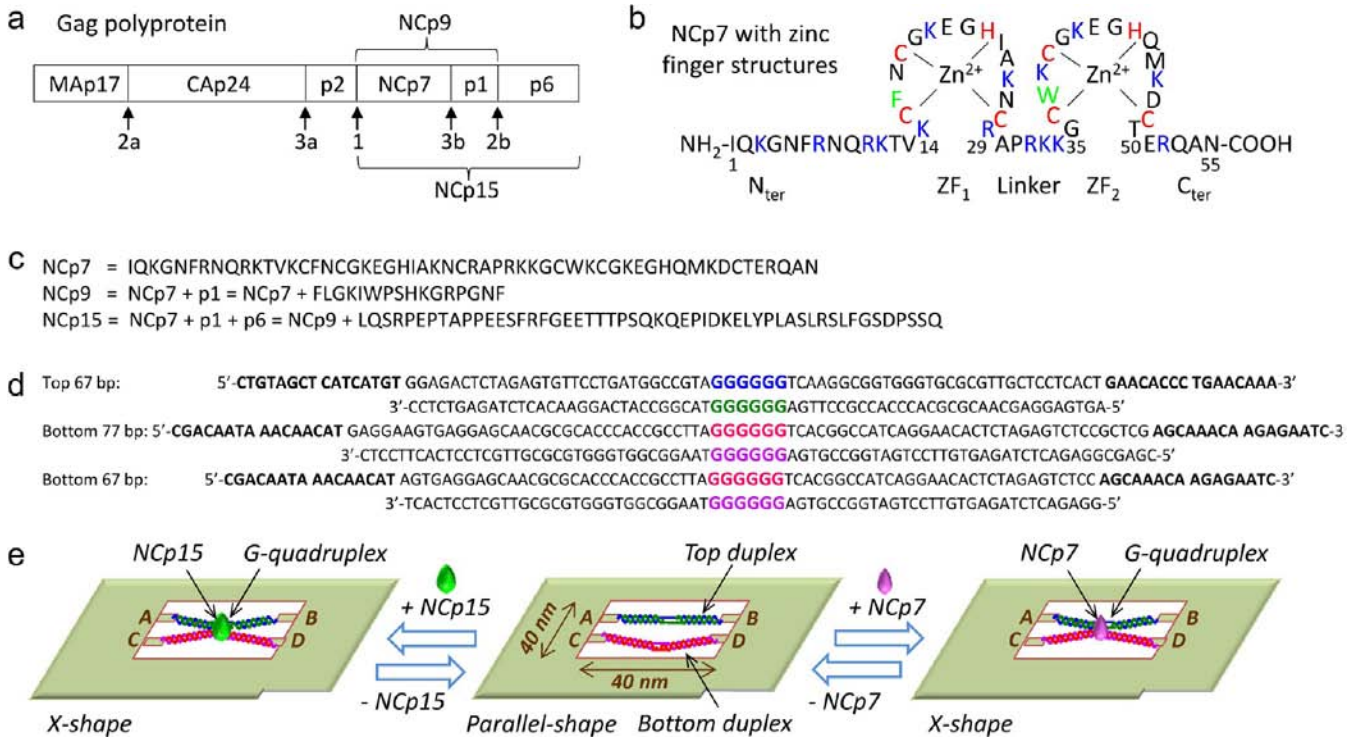


Figure 1. (a) Schematic representation of the cleavage steps that release the three forms of NCps from HIV-1 Gag during viral maturation. Primary, secondary and tertiary cleavage sites are indicated by the numbered arrows. (b) Amino acid sequence and the CCHC zinc finger domains of NCp7. The zinc-coordinating residues, 15 basic residues [5 arginines (R) and 10 lysines (K)], and 2 aromatic residues [phenylalanine (F) and tryptophan (W)] are shown in red, blue, and green, respectively. The sequence shown is for the pNL4-3 isolate. The GenBank accession no. AF324493. (c) The sequences of all three forms of NCps. (d) Six G-repeats containing sequences used in this study. Bold font regions represent the sticky ends that can be attached to the connectors in the origami frame. Colored regions in the middle represent the G-repeats. The numbers of G-repeats tested were three, four, and six. The length of the top duplex was 67 bp, while the bottom duplex used was either 67 bp (short-duplex) or 77 bp (long-duplex). The distance between two connectors (for example A and B) in the origami was designed to be 64 bp. (e) Schematic explanation of the DNA origami frame with G-G mismatch repeats containing duplexes, and protein-mediated conformational changes in the incorporated strands.

the G-quadruplex blocks the binding of HIV virions to cells and virus-mediated cell fusion. A serious of quadruplex-forming oligonucleotides and their modified counterparts have been shown to inhibit the activity of HIV-1 integrase, an enzyme which is responsible for the integration of viral DNA into host genome.[20] Also, the locked and intercalating nucleic acids capable of forming G-quadruplex structure have been reported to have enhanced anti-HIV-1 activity. Further, the studies on the interaction of HIV-1 proteins with G-quadruplex structures are of ongoing interest.

The main structural proteins of HIV-1 are expressed as a 55 kDa single polyprotein Gag which is essential for retroviral replication, virion assembly and genome packaging. Gag alone is sufficient for the formation of virus-like particles in a mammalian cell.[21] It is composed of a matrix (MAp17), capsid (CAp24), spacer peptide (p2), nucleocapsid (NCp7), spacer peptide (p1), and p6 from N- to C-terminal (Figure 1a). HIV-1 encodes three enzymes called protease (PR), reverse transcriptase (RT), and integrase (IN). During or shortly after virus budding from the infected cell, the PR is activated and results in the cleavage of Gag to the individual mature structural proteins. In the cleavage reaction, NCp15 (123 amino acids, a proteolytic intermediate of nucleocapsid protein (NCp)) was first produced and subsequently cleaved to NCp9 (71 amino acids) and further into the mature NCp7 (55 amino acids) through the consecutive removal of p6 and p1 (Figure 1c). NCp7 is a basic protein with two zinc finger domains that are CCHC motifs, which are separated by proline-rich linker as show in Figure 1b. Like other retroviral NCps, HIV-1 NCp is a multifunctional protein. It binds nucleic acids through electrostatic interactions of the basic residues (especially those in the N-terminal sequence and the linker) with the phosphodiester backbone of nucleic acids. Although NCp binds throughout strands in a nonspecific manner, it exhibits sequence-specific binding to runs of Gs, UGs or TGs through interactions that involve the zinc fingers.[22] Further, it facilitates remodeling of nucleic acids to fold thermodynamically stable conformations, and thus called nucleic acid chaperone.[22] HIV-1 NCp has been reported to promote and stabilize G-quadruplex structures[23] by an unknown mechanism, though a destabilizing effect was found for thrombin binding aptamer.[24]

To date only little is known on the structure, stoichiometry, NCp−NCp interactions, and chaperone activity as it relates to G-quadruplex structures, the searching mechanism for the target sequence, and so on. Here we aimed to unravel these phenomena by the direct and real-time analysis on the stoichiometry of HIV-1 NCps on bare mica surface, and the NCps-induced formation of a tetramolecular G-quadruplex structure[25] by the combined use of DNA origami (Figure 1e)[26] and high-speed atomic force microscopy (HS-AFM).[27] Our single-molecule analysis provided the unprecedented direct evidence for the oligomerization of NCps under nucleic acid free environment. Further, to the best of our knowledge, this is the first report on the HIV-1 NCps-induced G-quadruplex formation investigated by using AFM. Moreover, we report

dx.doi.org/10.1021/ja409085j | J. Am. Chem. Soc. 2013, 135, 18575−18585

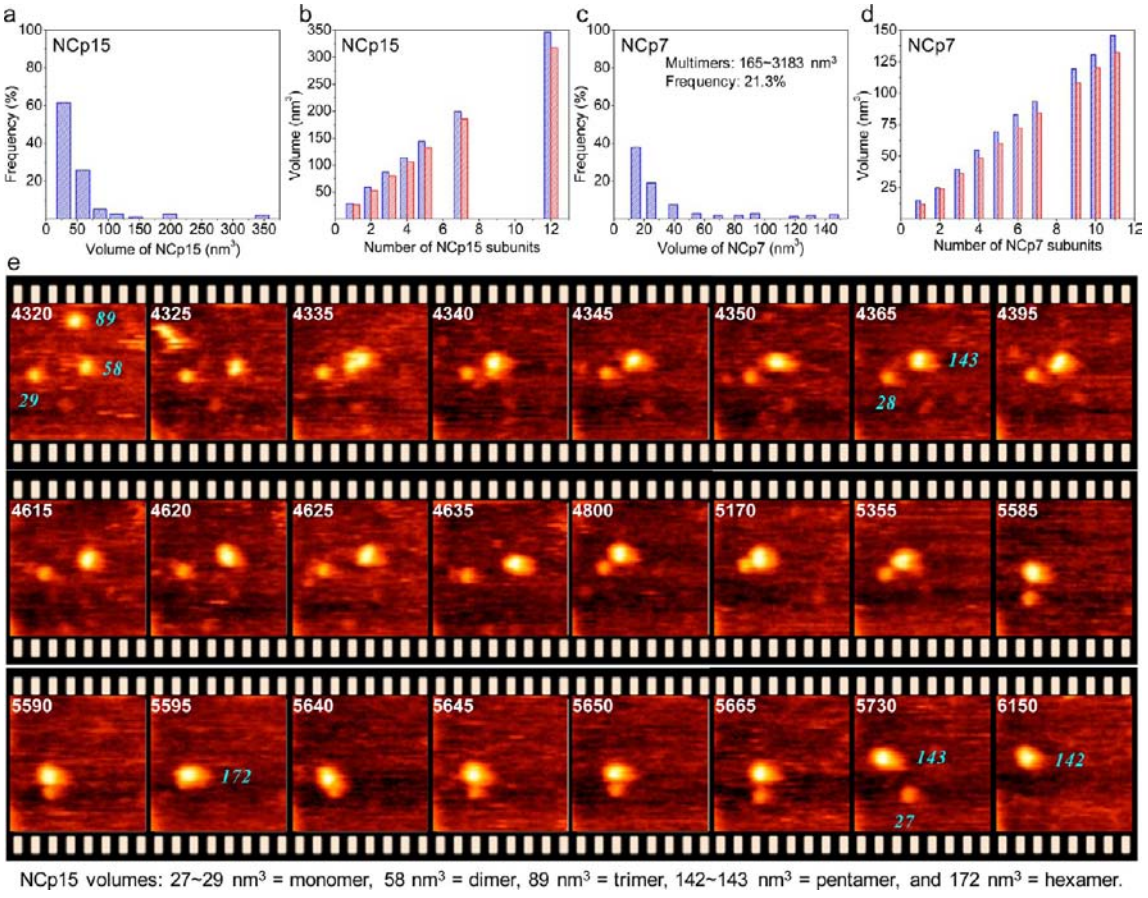


Figure 2. Frequency distribution of volumes of NCp15 (a) and NCp7 (c) on bare mica surface. For simplicity, higher order oligomers with volumes ranging from 165 to 3183 nm^3 are not included in the plot given in (c). Comparison of the experimental (blue bars) and theoretical (red bars) volumes of NCp15 (b) and NCp7 (d). (e) Snapshots of the real-time HS-AFM imaging of the NCp15 assembly. The numbers at the top left corner and near the protein particles represent the imaging time in seconds and particle volume in nm^3, respectively. Image size, 100 × 100 nm; scan speed: 0.2 frame/s; [Tris-HCl] = 20 mM, pH 7.6; [$MgCl_2$] = 10 mM; [Protein] = 0.4 μM; [KCl] = 0 mM.

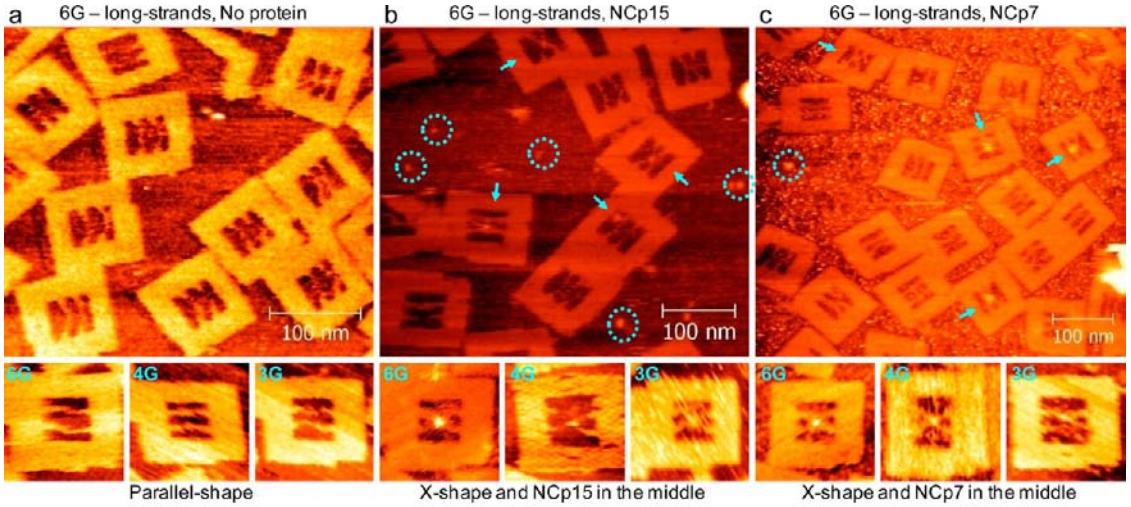
here the very first direct and real-time analysis on protein-induced G-quadruplex formation at the single-molecule level.

RESULTS AND DISCUSSION

Volume Analysis. A previous study on NCp7 concentration dependent analysis by gel electrophoresis indicated that different number of protein particles are binding to the quadruplex structure with increased number of particles by increasing the concentration.[23] Even though there is no detectable cooperativity, binding of multiple NCp7 molecules was reported to occur under saturating conditions with HIV-1 Psi-associated RNA hairpins.[28] These observations motivated us to analyze the molecular volume of NCps on bare mica surface, to understand whether protein assembly is possible even in the absence of nucleic acid or only in the presence of genetic material. The initial cleavage product NCp15 and the mature NCp7 were used for this purpose. The estimated volume vs the frequency of occurrence of different sized NCp15 particles on bare mica surface is plotted as given in Figure 2a (for AFM images, see Figure S1). From the plot it is clear that NCp15 is present as a monomer in majority of the cases (62% out of 117 particles considered) and dimer as a secondary form with 26% yield. It also forms the oligomers containing 3, 4, 5, 7, and 12 monomer units with the individual yield of 5% or less. Though we have not observed the particles of other sizes, they are likely to form depending on the experimental conditions such as protein concentration. The experimentally measured and theoretically calculated volumes of NCp15 particles are in good agreement, as shown in Figure 2b. For instance, the estimated average volumes of NCp15 monomer and dimer on bare mica surface are respectively 28 ± 5 and 59 ± 6 nm^3, in good agreement with theoretical volumes of 26 and 53 nm^3, respectively. Similarly, the volume analysis of NCp7 indicated that 38% exist in the monomer form out of 174 particles considered, 19% as a dimer, and several other particles with larger volumes were also found (Figure 2c; for AFM images, see Figure S1). The maximum volume of an NCp7 particle that we have detected was 3183 nm^3, indicating that it forms oligomers consisting of hundreds of monomer subunits. As shown in Figure 2d, the experimental and theoretical volumes of NCp7 correlate well. These observations clearly indicated that a significant amount of NCps assembly takes place prior their binding to the nucleic acids.

Real-Time Protein Assembly. Once we established the concept of protein−protein interactions, we then were interested in monitoring the NCps assembly in real-time. As far as we know, NCp−NCp interactions in the absence of nucleic acid has not been reported so far, and monitoring such an interaction in real-time will be quite extraordinary. The NCp15 oligomerization event was monitored with an image acquisition speed of 0.2 frame/s, and the snapshots are given in Figure 2e. Three protein particles with different sizes were seen

dx.doi.org/10.1021/ja409085j | J. Am. Chem. Soc. 2013, 135, 18575−18585



d

Sequences	No protein[a]		NCp15		NCp15-KCl		NCp7		NCp7-KCl	
	Long	Short	Long	Short	Long	Short	Long	Short	Long	Short
6G	23	14	60 (13)	47 (25)	64 (34)	63 (34)	60 (19)	49 (14)	69 (39)	70 (42)
Counted	374	333	146	193	173	201	318	248	182	163
4G	21	8	39 (2)	27 (4)	57 (7)	39 (8)	41 (3)	29 (4)	62 (10)	37 (6)
Counted	333	381	215	252	182	248	245	246	158	167
3G	13	6	30 (11)	28 (15)	39 (7)	36 (9)	33 (8)	29 (19)	36 (7)	36 (8)
Counted	132	371	154	248	203	226	188	253	204	231

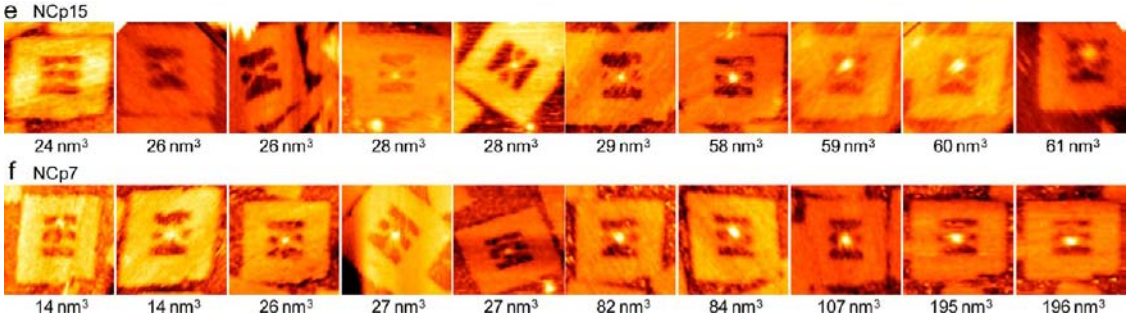


Figure 3. (a) Zoom-out AFM image of the DNA origami frames with incorporated duplexes recorded in the absence of NCps and G4-promoting salts. The G-repeat number in this case is six. (b) AFM image for the conditions as in the case of (a) but after addition of NCp15. (c) Image recorded for the conditions as in (a) but after addition of NCp7. Arrows indicate the X-shape. Representative zoom-in images for the sequences that contained six (6G), four (4G), and three (3G) G-repeats are also given in each case. Some free proteins on the mica surface are marked with dotted circles. (d) The % yield of X-shapes in the absence and presence of NCps calculated from the AFM images for all possible combinations of the strands. "Long" and "Short" represent the long-strand and short-strand. The numbers in the parentheses indicate the yield of X-shapes with protein oligomers. "Counted" means the number of origami tiles counted in each case. [a]The "No protein" data were taken from our parallel study, ref 36, obtained under identical conditions. AFM images of different sized NCp15 (e) and NCp7 (f) bound to the parallel or X-shape. The G-repeat number in this case is six. The estimated volume in each case is also given. All the images given in this figure were recorded for the long-strands. Zoom-in image size, 125 × 125 nm; [Tris-HCl] = 20 mM, pH 7.6; [$MgCl_2$] = 10 mM; [EDTA] = 1 mM; [Protein] = 0.4 μM; [KCl] = 0 or 100 mM.

at 4320 s. Volume analysis indicated that these particles were in monomeric (29 nm^3), dimeric (58 nm^3), and trimeric (89 nm^3) forms. Though the particles were immobilized/weakly attached on mica surface, they still retain their dynamic behavior as can be seen at 4325 s in which the trimer moved from its original position. Subsequently, the trimer made contact with the dimer between 4335 and 4345 s, leading to their association at 4350 s where only two particles appeared in the image. The larger particle was characterized to be a pentamer with a volume of 143 nm^3, and the smaller particle was a monomer which had not changed from its original size. The larger particle seems to adopt different conformations and the interactions between the monomer subunits within the pentamer may be weak, as it adopts different shapes in the AFM image between 4350 and 5355 s. For example, it adopts a spherical structure at 4350 and 4365 s, elliptical shapes tilted to the right-hand side at 4395 and 4625 s and to the left-hand side at 4620 s, and a triangular shape with three particles at each edge of the triangle at 4635 and 5355 s. Further, the monomer−pentamer assembly was also observed between 4800 and 5595 s leading to the formation of a hexamer NCp15 (172 nm^3) at 5595 s. Finally, a monomer unit was found to dissociate from the hexamer

dx.doi.org/10.1021/ja409085j | J. Am. Chem. Soc. 2013, 135, 18575−18585

between 5640 and 5730 s, and the dissociated monomer detached from the surface at 6150 s. This monomer dissociation from the hexamer further indicates that the NCp−NCp interactions may be weak and are reversible.

G-Quadruplex Observation System. Next, we investigated the chaperone activity of NCps on the formation of a tetramolecular G-quadruplex. For the quadruplex-forming sequences, we adopted two Watson−Crick duplexes with 3−6 contiguous G−G mismatches in the middle (Figure 1d). To structurally control the stoichiometry and orientation of the sequences of interest, we performed our analyses within a DNA origami frame (Figure 1e)[29,30] prepared by the "scaffolded DNA origami method".[26,31,32] To bring the duplexes closer, particularly the G−G-mismatch regions, and to promote quadruplex formation, we imposed structural flexibility in the strands by increasing the length of the strands. As a result, the length of the top duplex was kept constant at 67 bp (Figure 1d), the bottom duplex was varied to either 67 bp (short-strand) or 77 bp (long-strand), and the length between two connecting sites in the origami frame was constantly 64 bp. HIV-1 NCps were shown to exhibit the molecular chaperone activity on these type of four-stranded intermolecular G-quadruplexes,[23] while a destabilizing effect was observed for the TBA sequence[24] which forms a single-stranded intramolecular G-quadruplex. Thus, to study the chaperone activity of these proteins at single-molecule level, we have chosen these DNA sequences. The conformational switching was observed using HS-AFM[33−35] by monitoring the topological changes of the strands.

The origami frame with incorporated duplexes was prepared, purified, immobilized on the mica surface, and imaged by AFM. We first characterized the topology of the long-strands with six G-repeats.[36] In the absence of the proteins and G4-promoting salts, the duplexes may not come into contact as the conditions are unfavorable for G-quadruplex formation. Therefore, the AFM topography of the incorporated duplexes within the origami frame structure was expected to be parallel. The AFM image recorded under this condition is given in Figure 3a. As expected, we could see the parallel shape of the incorporated duplexes in most cases (77%), indicating that the G-quadruplex formation is not favorable under these conditions. The minor but significant amount of X-shapes observed under this condition could possibly because of the Mg^{2+} ions present in the buffer which can induce small amount of quadruplex structure.[36,37] The parallel shape was also obtained for all other sequences with different number of G-repeats such as four and three guanines (representative zoom-in images in Figure 3a) and different length of the strands such as 67 and 77 bp.

NCps-Induced G-Quadruplex Formation. Next, we performed our analysis in the presence of proteins. Formation of a quadruplex structure should bring the duplexes into contact in the middle, which in turn should result in a topographical change in the AFM image from the parallel to an X-shape. Interestingly, the X-shape was seen clearly in several instances (60%, Figure 3b, indicated by arrows) when the image was recorded for the long-strands with six G-repeats in the presence of NCp15. This X-shape was always observed in the middle of the incorporated strands, and no interaction was found on the duplex regions. Further, the bright spot with high pixel intensity seen in the middle of the X-shape is characterized to be the protein particle. Similar changes were also observed in the presence of NCp7, as shown in Figure 3c. Such a topographical change and the bright spot indicated the formation of the NCps-induced G-quadruplex structure. For both proteins used, the shorter G-tracts such as four and three were also found to undergo similar conformational changes (Figure 3b,c, and Figures S2 and S3). Same trend was also observed for the short-strands (Figures S4 and S5).

We have also tested the effect of NCps on control sequences in that the top and bottom duplexes contained six contiguous G-T and T-T mismatches, respectively. Both NCps are unable to induce X-shape to these mutated strands (0/243 and 0/295 origami counted respectively for NCp15 and NCp7, Figure S6) even in the presence of 3 times higher concentration of proteins. These observations validated our conclusion that the formed X-shape in case of G-G mismatch sequences is due to the formation of the G-quadruplex structure. Regarding the strand polarity of the G-quadruplex structure, we believe that the antiparallel structure is formed, because we fixed both the duplexes inside the origami frame and the strands within a duplex are oriented antiparallel. In addition, there may be a possibility for the formation of a mixed G-quadruplex conformation in which two strands orient in one direction and other two in the opposite direction, particularly if a duplex rotates. However, irrespective of strand polarity, the formed G-quadruplex adopts an exclusive stoichiometry of four strands.

Yield Calculations. The estimated quantitative yields of G-quadruplex structures formed in each case are tabulated in Figure 3d. Both the NCps are capable of inducing G-quadruplex formation at G-G mismatched region with G-tracts of as low as three. Concurrent addition of either NCps and KCl (a G4-promoting salt) slightly increased the X-shapes to 64 and 69% observed for NCp15 and NCp7, respectively. Note, the chaperone activity is independent of both ATP hydrolysis and G4-promoting salts, while previous studies involving the HIV-1 NCps binding to G-quadruplex was carried out in the presence of quadruplex-favorable ionic environments.[23,38] Moreover, only slight increase in the yield of G-quadruplex was observed in the simultaneous presence of NCp and KCl, indicating that the protein itself is sufficient for the chaperone activity. In addition, the yields of the G-quadruplexes induced by 0.4 μM NCps (for example, 60% for long-strands with six G's) are slightly lower than the yields observed in our parallel studies in the presence of 100 mM KCl (76%)[36] and slightly higher when compared to the presence of 1 μM G-quadruplex-binding ligand (PDC-biotin, 58%).[39] Increasing the protein concentration is likely to increase the amount of G-quadruplex structure. However, in the control experiments with 3 times higher concentration of proteins (1.2 μM), nonspecific binding of proteins to the origami structure and incorporated duplexes were often observed (Figure S6). This nonspecific binding was particularly higher in case of NCp7. This could be a major reason for the observed aggregation in the bulk studies with NCp7 at concentrations over 1 μM.[23]

Volume of Quadruplex-Bound NCps. We then performed the volume analysis of protein particles that are bound to the parallel and X-shapes within the origami frame. As shown in Figure 3e for NCp15 and Figure 3f for NCp7, we could see different sized particles bound to the parallel and X-shapes. The calculated volumes also indicated that the bound particles are monomers and dimers in case of NCp15, and range from monomers to oligomers consisting of several subunits in case of NCp7. Even though the larger particles were not observed in case of NCp15, they are likely to form with low yield, because in most cases NCp15 exists in the monomeric and dimeric forms in solution. These results are consistent with the previous

dx.doi.org/10.1021/ja409085j | J. Am. Chem. Soc. 2013, 135, 18575−18585

report that multiple number of protein monomers interact with the G-quadruplex;[23] however, the mechanism of protein assembly is significantly different than proposed previously. Our single-molecule analysis indicated that a significant amount of protein assembly takes place prior to its binding to the quadruplex structure. However, previous bulk analysis by gel electrophoresis hypothesized that the initial monomer binding nucleates an ordered arrangement of consecutive NCps (precisely NCp7) along the four-single-stranded tails.[23] Moreover, the existence of the oligomers in free solution and on the X-shape clearly indicates that the preformed oligomers are themselves capable of interacting with the target sequences and inducing the G-quadruplex structure. A solid piece of evidence to support this conclusion is derived by the real-time analysis on the multiple molecular events of NCp15, as described later in this article. It is worth mentioning here that the concentrations of the proteins used in the present study (0.4 μM) and previous study (0.25−0.4 μM) are comparable. A possible reason for the previous assumption is that the bulk analysis was based on the protein concentration-dependent study. The oligomerization might be enhanced at higher concentrations of the protein, leading to the occurrence of larger particles and their binding to the quadruplex structure, which in turn might have reflected like the seeding/nucleation effect as assumed.

Mg^{2+} was found to increase the higher molecular weight protein-quadruplex complex.[23] As AFM analysis requires Mg^{2+} ions to immobilize the DNA origami on mica surface, all experiments were carried out in the presence of Mg^{2+}. Thus, the higher molecular weight protein−quadruplex complex obtained in both previous and present study can be interpreted in terms of preassembly of proteins and its binding to the quadruplex structure. Note, the concentration of Mg^{2+} used in both studies is 10 mM. Also, the preliminary kinetic analysis indicated that the formation of higher molecular weight complexes of NCp7-quadruplex was obtained after the shortest incubation time tested of 1 min, and longer incubation failed to increase the amount of the higher order complexes.[23] This may be considered to be indirect evidence that the protein assembly takes place before its binding to the DNA.

The exact stoichiometry of NCp7 within the assembled protein−quadruplex complex was not estimated previously. Our results indicated that, irrespective of the proteins (NCp15 or NCp7), monomer to oligomers consisting of several protein subunits were found to interact with the quadruplex structure. Among the NCps tested, NCp7 seems to have a higher tendency to form oligomers than NCp15. The partially processed NCp9 was shown to be inefficient in promoting the protein oligomer binding to quadruplex DNA, while the present study indicated that the unprocessed NCp15 is capable of promoting the protein oligomer−quadruplex interaction, though the oligomer binding (particularly higher order oligomers) is more favorable in case of fully matured NCp7 (see Figure 3e,f). The height, width, and breath of both the protein oligomers were increased when compared to the respective monomers, indicated that the oligomerization takes place spherically and not through lateral association as suggested previously.[23] Besides, the protein−protein interaction seems to be weaker and the subunits in a protein oligomer were also visualized by AFM (Figure S7).

Structural Requirements for NCps Binding. To confirm that the zinc finger structures and the aromatic residues−G-quartet interactions are important for the chaperone activity of NCps, we have carried out control experiments in which the proteins were pretreated with EDTA. Preincubation with EDTA will remove the zinc ions from the binding sites and the zinc fingers will be unstructured. This may inhibit the protein function that induces the G-quadruplex structure. The EDTA treated protein particles on bare mica surface indicated that the oligomerization of NCp15 was not affected by the EDTA treatment (Figure S8), while it was diminished in case of NCp7 (Figure S9). This indicated that zinc fingers have different effect on the oligomerization of unprocessed and processed NCps. Interestingly, addition of either of the EDTA treated protein failed to induce G-quadruplex structure (only 4/323 and 5/352 X-shapes were obtained for NCp15 and NCp7, respectively), supporting that the zinc finger structures and the stacking interactions by the aromatic residues ^{16}Phe and ^{37}Trp present in the zinc finger (see Figure 1b) are mandatory for the chaperone activity of these proteins. However, the proteins with disrupted zinc fingers are still capable of binding the duplex DNA (Figures S8 and S9), indicating that the duplex DNA binding ability is due to the nonspecific electrostatic interactions and zinc finger structures are not necessary for such an interaction.

Regarding the structural requirements in the DNA sequences, it has been shown previously that NCps require 5-nt-long single-stranded tails flanking the quartets.[23] However, the present study utilized the double-stranded tails and the chaperone activity of NCps is not inhibited. By considering the previous results, present observation and the searching mechanisms described below, we anticipate that the single or double-stranded tails may be useful for the initial binding of the proteins through the nonspecific electrostatic interactions and their searching for the target sequence, while G-quadruplex is stabilized by the zinc finger−quartet interactions.

Searching Mechanism. The G-quadruplex structure adopted in our method is different from other structures described before,[23,24] as the quadruplex structure is formed within the duplex entity. Also, the quadruplex structure is formed from the duplex precursors. This attracted our attention to investigate the searching mechanism of the NCp for its target sequence. In general, the duplex DNA-binding proteins initially bind to a nonspecific region, and search for their target sequence to execute the functions such as catalytic reactions. The searching mechanism includes translocation of the proteins by one-dimensional (1D) sliding, 3D hopping from one part of the duplex to the other, and intersegmental transfer via loop formation in which the protein interacts with more than one part of the duplex DNA.

NCp15 was used for this investigation as a representative example. As can be seen in Figure 4a, the sliding event was successfully captured. The NCp15 in its dimer form is bound to the bottom duplex. During the successive scanning, the position of the protein is changed from left (430, 825, and 855 s) to middle (440, 450, 850, 945, and 955 s) and further to right (950 s) side. In addition, the protein moved forward and backward during the sliding. The translocation seems to be discrete, and the estimated 1D diffusion coefficient ($D = \langle \Delta l^2 \rangle / 2t$, where Δl is the distance translocated by the protein on the DNA during a time of t) is in the order of 10^{-13} −10^{-14} cm^2/s. The hopping event was also documented as part of the searching mechanism. As shown in Figure 4b, the NCp15 dimer was bound to the bottom duplex at 15 s. Between 15 and 25 s, it was sliding on the bottom duplex. At 30 s the protein was found to jump to the top duplex, and after briefly

dx.doi.org/10.1021/ja409085j | J. Am. Chem. Soc. 2013, 135, 18575−18585

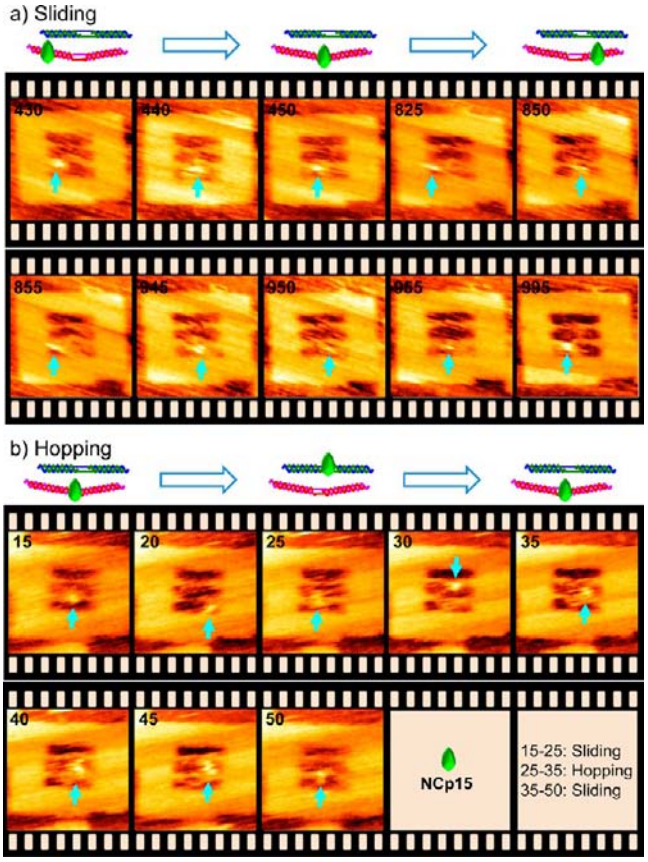


Figure 4. Snapshots of the real-time HS-AFM imaging of the searching mechanism: (a) 1D sliding and (b) 3D hopping. Long-strand with six G-repeats was used in these studies. Image size: 125 × 125 nm; scan speed: 0.2 frame/s; [Tris-HCl] = 20 mM, pH 7.6; [$MgCl_2$] = 10 mM; [EDTA] = 1 mM; [NCp15] = 0.4 μM; [KCl] = 0 mM.

interacting with the top duplex it came back to the bottom duplex at 35 s by reverse hopping. After the hopping events, the protein particle was found to slide again between 35 and 50 s. The formation of a loop within a single duplex was not observed in our experiments. This could be due to the short length of the incorporated duplexes and also the fact that they were strongly anchored within the frame structure. However, a loop-like structure in which the protein particle was found to interact with both duplexes was visualized in our real-time analysis, described later in this article. Regarding the binding, NCp is known to interact with nucleic acids through electrostatic interactions. As the experiments were carried out under low salt concentrations, these electrostatic interactions could be stronger. Besides, this is the first investigation on the searching mechanism of NCps for the G-rich target sequence constituted within the duplex DNAs.

Chaperone Activity of NCp15 Monomer. We have continued our real-time analysis and successfully monitored the NCp15 monomer-induced G-quadruplex formation, as detailed in Figure S10 and Movie S1). The kinetics of NCp15-induced G-quadruplex formation is comparable with the salt-induced formation of (3+1)-type[40,41] and four-stranded G-quadruplex[36] structures as the X-shape was formed at 40 s in the former case and 85 s in the latter case, while faster than the G4-binding ligand (required 510 s in the presence of PDC−biotin).[39]

Dissociation Event. The real-time analysis on the unfolding process of quadruplexes could widen our understanding on their folding pathways and intermediates involved. We have carried out such an analysis using six contiguous Gs containing long-strands. The NCp15 incubated origami assembly was adsorbed on mica surface and imaging was performed under a protein-free environment with an image acquisition speed of 0.2 frame/s. Two instances of such dissociation events were documented; snapshots of first instance are given in Figure 5, and movies of both instances are given in the Supporting Information (Movies S2 and S3). The X-shape of the strands and NCp15 in the middle of the X-shape were seen at 0 s and was maintained until 70 s. Within next 5 s, the X-shape was dissociated into parallel-shape (75s) and the protein was bound to the bottom duplex. Between 80 and 85 s, the strands seemed to be weakly interacting. Interestingly, the reverse conformational switching of parallel to X-shape was observed at 90 s. This type of change is uncommon for the salt-induced G-quadruplex structures,[36,40,41] while similar changes were observed in our parallel study on PDC−biotin ligand-induced quadruplex formation.[39] This X-to-parallel and reverse conformational switching were continued for a long time of up to 600 s. At 605 s, the X-shape was broken into the parallel-shape, and the protein was bound to the bottom duplex. After that, the reverse conformational switching was not observed until the end of the scanning of up to 1185 s. In this event, once the protein is assembled into a dimer, the disassembly into its subunits was not observed throughout the imaging.

The unfolding event can be characterized into three distinct steps: (i) the initial G-quadruplex structure with an X-shape topography (0−70 s); (ii) intermediate-like states between the fully folded G-quadruplex to fully unfolded G−G mismatches (75−600 s); and (iii) the final completely unfolded state (605−1185 s). It is noteworthy that, during the rapid forward and backward conformational switching between 75 and 600 s, the strands were neither completely separated nor resulted in a strong X-shape (see 85, 100, 105, 145, 255, 285, 360, 370, 465, 475, and 540 s). Rather, the strands seemed to be partly in contact with each other. Thus, we believe that these states might reflect intermediate states such as hairpin and triplex,[2,39] or the duplex-bound NCp15 may produce the rapid conformational changes. The dissociation kinetics of NCp15-quadruplex seems to be slow and takes about 605 s, while fast kinetics involved in case of salt or ligand-induced quadruplexes.[36,39−41] Also, it seems that NCp15-quadruplex/duplex DNA interaction is strong, and the protein-free buffer alone is not strong enough to separate them. In the second instance (Movie S3), though the X-to-parallel and reverse conformational changes were observed, the final unfolding into a parallel-shape was not observed for nearly same scanning time.

Multiple Molecular Events of NCp15. Next, we have documented the multiple molecular events of NCp15 (Figure 6). The origami frame with incorporated duplexes was prepared and immobilized on mica surface under protein-free conditions, while images were recorded using a protein containing buffer. During the course of the imaging, NCp15 monomers on mica surface that are closely located to an origami frame were seen at 1800 s (indicated by arrow). These monomers are highly dynamic and moving from their original position to the top of the origami structure at 1810 s. After briefly placing themselves on the origami structure, they entered into the inner vacant space in the origami frame at 1815 s. At 1830 s, both the monomers appeared clearly in the images as indicated by the arrows. Interestingly, the protein assembly was observed at

dx.doi.org/10.1021/ja409085j | J. Am. Chem. Soc. 2013, 135, 18575−18585

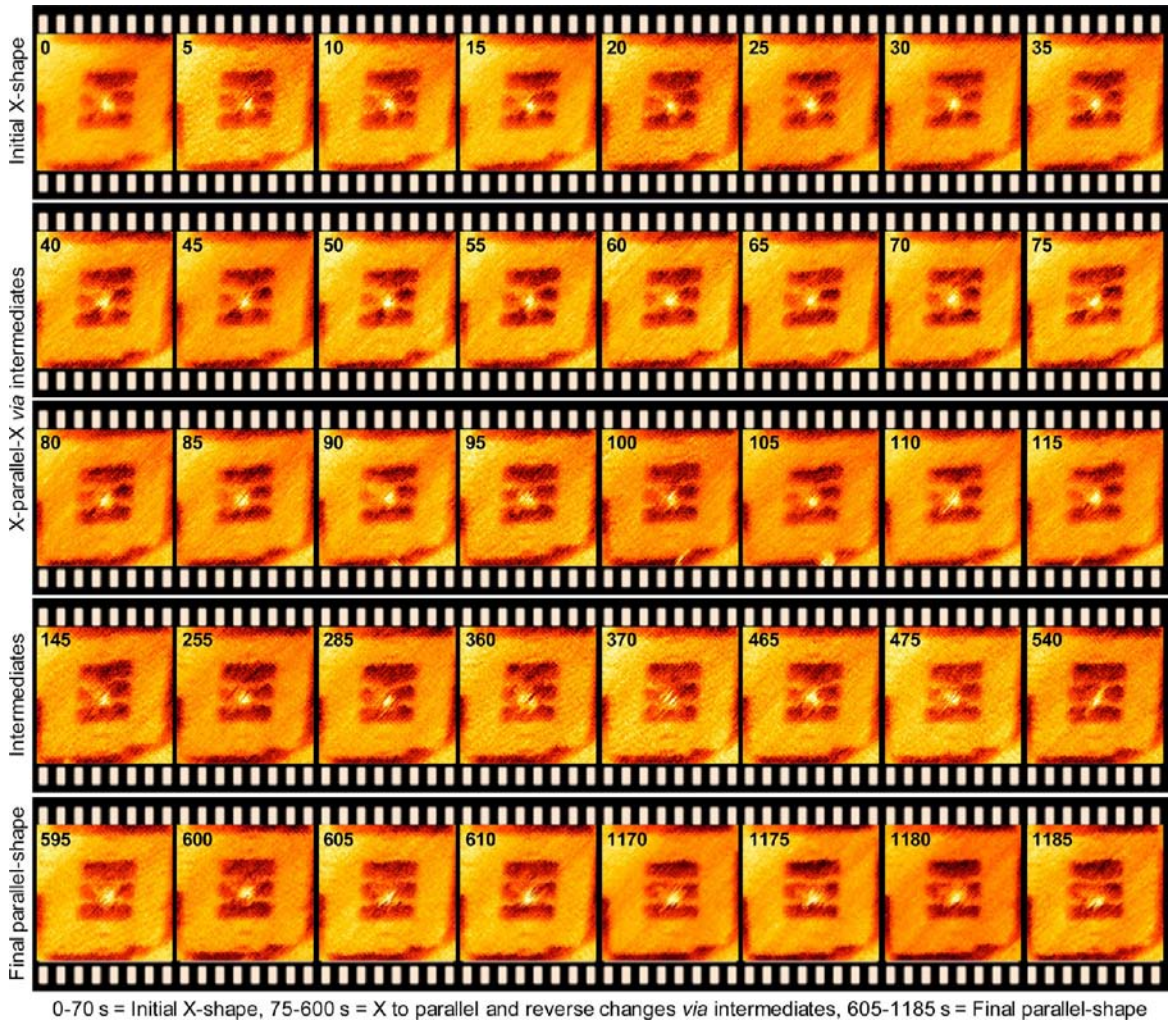


Figure 5. Snapshots of the real-time HS-AFM imaging of G-quadruplex dissociation. Long-strand with six G-repeats was used in this study. Image size: 125 × 125 nm; scan speed: 0.2 frame/s; [Tris-HCl] = 20 mM, pH 7.6; [$MgCl_2$] = 10 mM; [EDTA] = 1 mM; [NCp15] = 0.4 μM; [KCl] = 0 mM.

1915 s with the assembled particle volume corresponding to a dimer. This strongly supports our prediction that the protein assembly takes place before they induce/binding to a G-quadruplex structure. Though the proteins are placed along with the duplex DNAs and DNA origami, the assembly takes place without any assistance from the nucleic acids. After protein assembly, the dimer NCp15 moved close to the top duplex at 2655 s and bound to the top duplex at 2675 s. Sliding was observed between 2675 to 2770 s. Until 2770 s the dimeric NCp15 was found to interact with only one duplex, and at 2770 s the protein was observed nearly in the center of the top duplex where the G-G mismatches were placed. It seems like the target sequence was recognized by the protein at this time. Thereafter, a loop-like structure (or the DNA synapsis that appear in intersegmental transfer) in which the protein interacts with both the duplexes was observed from 2775 to 2785 s, supporting our assumption that between 2730 and 2770 s the target sequence was recognized by the protein. The formation of X-shape was observed at around 2790 s, and the X-shape was retained until 3000 s. Deformation of X-shape was seen at 3005 s, where the protein dimer moved to the bottom duplex. Finally, the dissociation of the NCp15 dimer was observed between 3005 and 3090 s. In this case, the dissociation of the protein in its dimeric form was observed. The dissociated protein may search for the next target. We could not get any information on the "successive model" in which the proteins remain bound with substrate DNA after the chaperone action, then continue to scan for additional recognition sites, as the target DNAs are fixed within a origami frame and only one target site is present inside the frame structure.

Updated Mechanism of Action. Finally, the role of HS-AFM in updating the mechanism of action involved in the chaperone activity of NCps can be summarized as follows: (i) NCp−NCp interactions take place prior to the protein binding to the G-quadruplex structure, and the process of protein assembly is independent of nucleic acids. (ii) The nucleation-like mechanism for the NCps−G-quadruplex assembly seems to be not valid. (iii) It seems that the protein−protein interactions are not strong, and the monomer subunits are weakly bound in their oligomeric form. Further, the oligomerization of both NCp15 and NCp7 is reversible and in free solution these proteins can behave like both a monomer and oligomers containing different subunits. (iv) The protein oligomerization takes place spherically, though a lateral association in the context of DNA was suggested previously.[23] (v) Among the NCps tested, NCp7 seems to have higher tendency to form the oligomers than NCp15. (vi) Irrespective of the size of the protein particles, the unprocessed NCp15 and

dx.doi.org/10.1021/ja409085j | J. Am. Chem. Soc. 2013, 135, 18575−18585

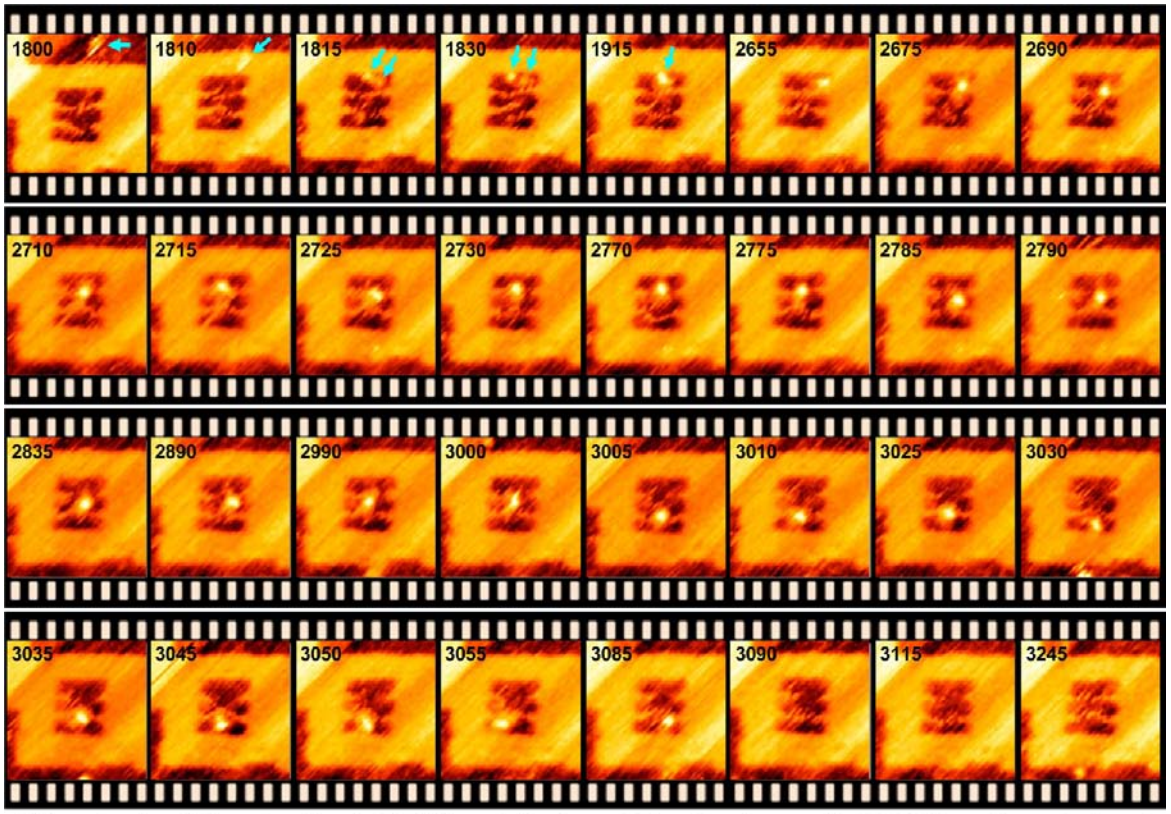


Figure 6. Snapshots of the real-time HS-AFM imaging of the multiple molecular events of NCp15. Long-strand with six G-repeats was used in this study. Image size: 125 × 125 nm; scan speed: 0.2 frame/s; [Tris-HCl] = 20 mM, pH 7.6; [$MgCl_2$] = 10 mM; [EDTA] = 1 mM; [NCp15] = 0.4 μM; [KCl] = 0 mM.

fully matured NCp7 are capable to induce G-quadruplex structure and is independent of both ATP hydrolysis and G4-promoting salts. (vii) The exact stoichiometry of the NCps within the assembled protein−G-quadruplex complex is estimated in the present study. (viii) It is not clearly understood to what extent the complete processing of Gag to its final form of NCp7 is required for the enhancement of specific rearrangements of nucleic acid secondary structure, and contradictory results were obtained with different cases. For instance, it was reported that only processed NCp is capable of facilitating minus-strand transfer.[42,43] However, a recent study explained that while both NCp7 and Gag are effective at annealing $tRNA_3^{Lys}$ to the primer-binding site on HIV-1 genome, Gag is more effective than NCp.[44] Moreover, the partially processed NCp9 was shown to be inefficient in promoting the protein oligomer binding to the quadruplex DNA, indicated the importance of C-terminal processing.[23] While the present study indicated that the unprocessed NCp15 is capable of promoting the protein oligomer−quadruplex interaction, the oligomer binding (particularly higher order oligomers) is more favored in the case of fully processed NCp7. Further, the C-terminal processing of NCp may not be necessary for the molecular chaperone activity of NCp for the formation of tetramolecular G-quadruplex structure, as even the unprocessed NCp15 induces the quadruplex formation, particularly with the same yield as that of the fully processed NCp7. (ix) If the quadruplex is formed in the entity of duplex DNA, then the protein involves several searching mechanisms including 1D sliding, 3D hopping, and translocation by transient loop formation. (x) The G-quadruplex is stabilized by the zinc finger−quartet interactions. (xi) We hypothesize that the single- or double-stranded tails may be useful for the initial binding of the proteins and their searching for the target sequence. Finally, (xii) once the chaperone activity is over, the protein either remains bound to the target site or dissociates without disintegrating into its monomer subunits. The dissociated protein may search for another target.

■ CONCLUSION

In conclusion, we present here the very first report on the direct and real-time investigations on the molecular chaperone activity of HIV-1 NCps for the formation of a tetramolecular G-quadruplex structure at single-molecule level. The quadruplex structure was found to be induced by both unprocessed NCp15 and matured NCp7. Volume analysis of protein particles on bare mica surface indicated that both the proteins exist in monomer as a major fraction and dimer form as a secondary fraction, while multimers consisting of several protein subunits were also observed. The molecular mechanism of protein assembly on the G-quadruplex structure is well documented, and the proteins with different sizes were found to interact with the incorporated duplexes and induce the quadruplex structure. Determination of the stoichiometry and affinity of binding between NCps and the G-quadruplex would provide valuable information about the possible sequence of events during genome recognition, dimerization, and packaging. Further, based on our results on NCp−NCp interactions, we postulate that such interactions may also be possible in the viral capsid

structure. This finding may serve as a model to understand the formation of the capsid structure from two copies of the genomic RNA and approximately 2000 NCps that generate a highly compact nucleoprotein core which is internalized within the mature capsid. As part of the searching mechanism, the translocation events such as 1D sliding, 3D hopping, and loop formation were captured. Moreover, the real-time formation and dissociation events of the G-quadruplex were imaged, and the intermediate-like states were also visualized. Finally, the complete trajectory of the protein particles from the initial appearance on mica surface to the dissociation from the target site after chaperone activity was traced out. Our HS-AFM analysis provided both traditional and novel information that helped us to update the NCps-quadruplex binding model and the molecular chaperone activity of NCps. We anticipate that this pioneering mechanistic investigation at single-molecule level will strengthen the understanding on the chaperone activity of HIV-1 nucleocapsid proteins which in turn will be helpful for the development of drugs both targeting the G-quadruplex structures and also against AIDS.

EXPERIMENTAL SECTION

Preparation of Proteins. Recombinant HIV-1 NCp7 and NCp15 were prepared and purified as described previously.[45] Briefly, HPLC fractions containing the purified, expressed protein (characterized by MALDI-MS and N-terminal sequencing) were collected, pooled, and lyophilized. The proteins were resuspended in the presence of 1 equiv of Zn^{2+} per finger, and their concentration was calculated on the basis of amino acid analysis measurements. Solutions were aliquoted into sterile polypropylene vials, relyophilized, and stored at −80 °C until used. The molecular masses of the proteins were determined by MALDI-TOF mass spectrometry and are 6351.37 and 13963.75 for NCp7 and NCp15, respectively. These values are consistent with the sequences listed in Figure 1c.

Studies in the Presence of Proteins. Design of DNA origami frame,[29,30] duplex DNAs with G-G mismatches, and preparation of the origami frame with incorporated duplexes[36,39] are described elsewhere, and a brief outline is also given in the Supporting Information. For the static analysis in the presence of proteins, the protein (final concentration of 0.4 μM, origami:protein = 1:40) was added after the purification of the origami. However, the observation buffer contained no protein. For the analysis in the absence of K^+, all the experimental steps such as preparation of origami and duplex strands, sephacryl column, surface immobilization, and observation buffer contained no KCl, while all these steps contained 100 mM KCl for the experiments in the presence of K^+.

Real-Time Observations. For the real-time analysis of the searching mechanism, formation of a quadruplex structure by NCp15 monomer and the multiple molecular events of NCp15, the origami assembly was fabricated, purified and adsorbed on mica surface in the absence of protein, while imaging was carried out by using 1x Tris-HCl buffer (20 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$) containing 0.4 μM protein. In case of deformation event, NCp15 (final concentration of 0.4 μM) was added to the purified origami assembly and was immobilized on mica surface. To increase the amount of protein-bound X-shapes, the origami assembly on mica surface was further incubated with 2 μL of 0.4 μM NCp15 for 5 min. Imaging was performed by using the protein-free 1x Tris-HCl buffer after gently washing the surface 3−5 times.

AFM Imaging. AFM images were recorded using a fast-scanning AFM system[33−35] (Nano Live Vision, RIBM Co. Ltd., Tsukuba, Japan) with a silicon nitride cantilever (resonant frequency 1.0−2.0 MHz, spring constant 0.1−0.3 N/m, EBDTip radius <15 nm, Olympus BL-AC10EGS-A2). The sample (2 μL) was adsorbed onto a freshly cleaved mica plate (ϕ 1.5 mm, RIBM Co. Ltd., Tsukuba, Japan) for 5 min at room temperature, and then the surface was gently washed 3−5 times using 1x Tris-HCl buffer solution with the same concentration of salt in which origami was prepared. Scanning was performed using the tapping mode in the same buffer solution. AFM images were recorded either using the observation buffer that contained KCl or KCl-free buffer depending on the requirements. All images reported here were recorded with an image acquisition speed of 0.2 frame/s. The yield calculations of the parallel and X-shapes were carried out by counting the shapes in the AFM images.

Molecular Volumes of Proteins. Theoretical molecular volumes of proteins based on molecular mass were calculated using the equation reported before:[46,47]

$$V_c = (M_0/N_0)(V_1 + dV_2)$$

where, M_0 is the molecular weight of protein (13963.75 and 6351.37 Da for NCp15 and NCp7), N_0 is Avogadro's number, V_1 and V_2 are respectively the partial specific volumes of protein and water, and d is the extent of protein hydration. The V_1 is considered to be 0.74 cm^3/g,[48,49] the d has been estimated to be 0.4 mol of water/mol of protein[50] and the V_2 is 1 cm^3/g. The calculated molecular volumes for NCp7 and NCp15 in monomeric form are 12 and 26 nm^3, respectively.

ASSOCIATED CONTENT

Supporting Information

Additional experimental details, AFM images, additional discussions, and real-time movies. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

endo@kuchem.kyoto-u.ac.jp
hs@kuchem.kyoto-u.ac.jp

Present Addresses

[⊥]Faculty of Medicine and Life Science Center of TARA, University of Tsukuba, 1-1-1 Tennodai, Tsukuba-shi, Ibaraki-ken 305-8577, Japan
[▽]Department of Molecular Biology, Princeton University, Princeton, NJ 08544

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Sebastien Lyonnais and Gilles Mirambeau for their kind assistance with the proteins used in this study. We express our sincere thanks for the CREST grant from the Japan Science and Technology Corporation (JST), grants from the WPI program (WPI-iCeMS, Kyoto University), and JSPS KAKEN-HI (Grant numbers 24310097, 24225005 and 24104002). Financial supports from The Mitsubishi Foundation and The Asahi Glass Foundation to M.E. are also acknowledged. A.R. expresses sincere thanks to the Japan Society for the Promotion of Science (JSPS) for the Postdoctoral fellowship. P.L.T.T. and J.-L.M. sincerely thank the Aquitaine Regional Council and the ANR program G4-Toolbox, Quarpdiem and Oligoswitch for financial support. For R.J.G., this project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN261200800001E with Leidos Biomedical Research, Inc.

REFERENCES

(1) Rajendran, A.; Endo, M.; Hidaka, K.; Sugiyama, H. *J. Am. Chem. Soc.* **2013**, *135*, 1117−1123.
(2) Mashimo, T.; Yagi, H.; Sannohe, Y.; Rajendran, A.; Sugiyama, H. *J. Am. Chem. Soc.* **2010**, *132*, 14910−14918.

(3) Phan, A. T.; Modi, Y. S.; Patel, D. J. *J. Am. Chem. Soc.* **2004**, *126*, 8710−8716.
(4) Cheung, I.; Schertzer, M.; Rose, A.; Lansdorp, P. M. *Nat. Genet.* **2002**, *31*, 405−409.
(5) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. *Nat. Chem.* **2013**, *5*, 182−186.
(6) Rezler, E. M.; Bearss, D. J.; Hurley, L. H. *Annu. Rev. Pharmacol.* **2003**, *43*, 359−379.
(7) Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2007**, *35*, 406−413.
(8) Mani, P.; Yadav, V. K.; Das, S. K.; Chowdhury, S. *PLoS One* **2009**, *4*, e4399.
(9) Fang, G.; Cech, T. R. *Cell* **1993**, *74*, 875−885.
(10) Giraldo, R.; Suzuki, M.; Chapman, L.; Rhodes, D. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7658−7662.
(11) Erlitzki, R.; Fry, M. *J. Biol. Chem.* **1997**, *272*, 15881−15890.
(12) Muniyappa, K.; Anuradha, S.; Byers, B. *Mol. Cell. Biol.* **2000**, *20*, 1361−1369.
(13) Arimondo, P. B.; Riou, J.-F.; Mergny, J.-L.; Tazi, J.; Sun, J.-S.; Garestier, T.; Helene, C. *Nucleic Acids Res.* **2000**, *28*, 4832−4838.
(14) Uliel, L.; Weisman-Shomer, P.; Oren-Jazan, H.; Newcomb, T.; Loeb, L. A.; Fry, M. *J. Biol. Chem.* **2000**, *275*, 33134−33141.
(15) Baran, N.; Pucshansky, L.; Marco, Y.; Benjamin, S.; Manor, H. *Nucleic Acids Res.* **1997**, *25*, 297−303.
(16) Sun, H.; Bennett, R. J.; Maizels, N. *Nucleic Acids Res.* **1999**, *27*, 1978−1984.
(17) Paeschke, K.; Capra, J. A.; Zakian, V. A. *Cell* **2011**, *145*, 678−691.
(18) Sun, H.; Karow, J. K.; Hickson, I. D.; Maizels, N. *J. Biol. Chem.* **1998**, *273*, 27587−27592.
(19) Esté, J. A.; Cabrera, C.; Schols, D.; Cherepanov, P.; Gutierrez, A.; Witvrouw, M.; Pannecougue, C.; Debyser, Z.; Rando, R. F.; Clotet, B.; Desmyter, J.; Clercg, E. D. *Mol. Pharmacol.* **1998**, *53*, 340−345.
(20) Rando, R. F.; Ojwang, J.; Elbaggari, A.; Reyes, G. R.; Tinder, R.; McGrath, M. S.; Hogan, M. E. *J. Biol. Chem.* **1995**, *270*, 1754−1760.
(21) Swanstrom, R.; Wills, J. W. In *Retroviruses* Coffin, J. M., Hughes, S. H., Varmus, H., Eds.; Cold Spring Harbor Laboratory Press: New York, 1997; pp 263−334.
(22) Levin, J. G.; Guo, J.; Rouzina, I.; Musier-Forsyth, K. *Prog. Nucleic Acid Res. Mol. Biol.* **2005**, *80*, 217−286.
(23) Lyonnais, S.; Gorelick, R. J.; Mergny, J.-L.; Cam, E. L.; Mirambeau, G. *Nucleic Acids Res.* **2003**, *31*, 5754−5763.
(24) Kankia, B. I.; Barany, G.; Musier-Forsyth, K. *Nucleic Acids Res.* **2005**, *33*, 4395−4403.
(25) Venczel, E. A.; Sen, D. *J. Mol. Biol.* **1996**, *257*, 219−224.
(26) Rothemund, P. W. K. *Nature* **2006**, *440*, 297−302.
(27) Rajendran, A.; Endo, M.; Sugiyama, H. *Chem. Rev.* **2013**, DOI: 10.1021/cr300253x.
(28) Hagan, N.; Fabris, D. *Biochemistry* **2003**, *42*, 10736−10745.
(29) Endo, M.; Katsuda, Y.; Hidaka, K.; Sugiyama, H. *J. Am. Chem. Soc.* **2010**, *132*, 1592−1597.
(30) Endo, M.; Katsuda, Y.; Hidaka, K.; Sugiyama, H. *Angew. Chem., Int. Ed.* **2010**, *49*, 9412−9416.
(31) Rajendran, A.; Endo, M.; Sugiyama, H. *Angew. Chem., Int. Ed.* **2012**, *51*, 874−890.
(32) Rinker, S.; Ke, Y.; Liu, Y.; Chhabra, R.; Yan, H. *Nat. Nanotechnol.* **2008**, *3*, 418−422.
(33) Rajendran, A.; Endo, M.; Sugiyama, H. *Adv. Protein Chem. Struct. Biol.* **2012**, *87*, 5−55.
(34) Rajendran, A.; Endo, M.; Katsuda, Y.; Hidaka, K.; Sugiyama, H. *J. Am. Chem. Soc.* **2011**, *133*, 14488−14491.
(35) Rajendran, A.; Endo, M.; Katsuda, Y.; Hidaka, K.; Sugiyama, H. *ACS Nano* **2011**, *5*, 665−671.
(36) Rajendran, A.; Endo, M.; Hidaka, K.; Tran, P. L. T.; Mergny, J.-L.; Sugiyama, H. *Nucleic Acids Res.* **2013**, *41*, 8738−8747.
(37) Sravani, M.; Nagaveni, V.; Prabhakar, S.; Vairamani, M. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 2095−2098.
(38) Mukundan, V. T.; Do, N. Q.; Phan, A. T. *Nucleic Acids Res.* **2011**, *39*, 8984−8991.
(39) Rajendran, A.; Endo, M.; Hidaka, K.; Tran, P. L. T.; Teulade-Fichou, M.-P.; Mergny, J.-L.; Sugiyama, H. 2013, submitted.
(40) Sannohe, Y.; Endo, M.; Katsuda, Y.; Hidaka, K.; Sugiyama, H. *J. Am. Chem. Soc.* **2010**, *132*, 16311−16313.
(41) Dhakal, S.; Mao, H.; Rajendran, A.; Endo, M.; Sugiyama, H. In *Guanine Quartets: Structure and Application*; Spindler, L., Fritzsche, W., Eds.; The Royal Society of Chemistry: Cambridge, 2013; pp 73−85.
(42) Guo, J.; Wu, T.; Anderson, J.; Kane, B. F.; Johnson, D. G.; Gorelick, R. J.; Henderson, L. E.; Levin, J. G. *J. Virol.* **2000**, *74*, 8980−8988.
(43) Guo, J.; Wu, T.; Kane, B. F.; Johnson, D. G.; Henderson, L. E.; Gorelick, R. J.; Levin, J. G. *J. Virol.* **2002**, *76*, 4370−4378.
(44) Roldan, A.; Warren, O. U.; Russell, R. S.; Liang, C.; Wainberg, M. A. *J. Biol. Chem.* **2005**, *280*, 17488−17496.
(45) Cruceanu, M.; Urbaneja, M. A.; Hixson, C. V.; Johnson, D. G.; Datta, S. A.; Fivash, M. J.; Stephen, A. G.; Fisher, R. J.; Gorelick, R. J.; Casas-Finet, J. R.; Rein, A.; Rouzina, L.; Williams, M. C. *Nucleic Acids Res.* **2006**, *34*, 593−605.
(46) Schneider, S. W.; Lärmer, J.; Henderson, R. M.; Oberleithner, H. *Pflügers Arch.− Eur. J. Physiol.* **1998**, *435*, 362−367.
(47) Mela, I.; Kranaster, R.; Henderson, R. M.; Balasubramanian, S.; Edwardson, J. M. *Biochemistry* **2011**, *51*, 578−585.
(48) Oncley, J. L.; Scatchard, G.; Brown, A. *J. Phys. Colloid Chem.* **1947**, *51*, 184−198.
(49) Durchschlag, H.; Zipper, P. *J. Appl. Crystallogr.* **1997**, *30*, 803−807.
(50) Grant, E. H. *Phys. Med. Biol.* **1957**, *2*, 17−28.