J Comput Aided Mol Des (2014) 28:869–884
DOI 10.1007/s10822-014-9772-9

Crystal structure of an antiviral ankyrin targeting the HIV-1 capsid and molecular modeling of the ankyrin-capsid complex

Warachai Praditwongwan · Phimonphan Chuankhayan · Somphot Saoin · Tanchanok Wisitponchai · Vannajan Sanghiran Lee · Sawitree Nangola · Saw See Hong · Philippe Minard · Pierre Boulanger · Chun-Jung Chen · Chatchai Tayapiwatana

Received: 24 December 2013 / Accepted: 24 June 2014 / Published online: 5 July 2014

Abstract Ankyrins are cellular repeat proteins, which can be genetically modified to randomize amino-acid residues located at defined positions in each repeat unit, and thus create a potential binding surface adaptable to macromolecular ligands. From a phage-display library of artificial ankyrins, we have isolated Ank^{GAG}1D4, a trimodular ankyrin which binds to the HIV-1 capsid protein N-terminal domain (NTD^{CA}) and has an antiviral effect at the late steps of the virus life cycle. In this study, the determinants of the Ank^{GAG}1D4-NTD^{CA} interaction were analyzed using peptide scanning in competition ELISA, capsid mutagenesis, ankyrin crystallography and molecular modeling. We determined the Ank^{GAG}1D4 structure at 2.2 Å resolution, and used the crystal structure in molecular docking with a homology model of HIV-1 capsid. Our results indicated that NTD^{CA} alpha-helices H1 and H7 could mediate the formation of the capsid-Ank^{GAG}1D4 binary complex, but the interaction involving H7 was predicted to be more stable than with H1. Arginine-18 (R18) in H1, and R132 and R143 in H7 were found to be the key players of the Ank^{GAG}1D4-NTD^{CA} interaction. This was confirmed by R-to-A mutagenesis of NTD^{CA}, and by sequence analysis of trimodular ankyrins negative for capsid binding. In Ank^{GAG}1D4, major interactors common to H1 and H7 were found to be S45, Y56, R89, K122 and K123. Collectively, our ankyrin-capsid binding analysis implied a significant degree of flexibility within the NTD^{CA} domain of the HIV-1 capsid protein, and provided some clues for the design of new antivirals targeting the capsid protein and viral assembly.

Keywords HIV-1 · CA protein · Ankyrins · Molecular docking · Modeling · Viral assembly · Antivirals

Warachai Praditwongwan and Phimonphan Chuankhayan have contributed equally.

Electronic supplementary material The online version of this article (doi:10.1007/s10822-014-9772-9) contains supplementary material, which is available to authorized users.

W. Praditwongwan · S. Saoin · C. Tayapiwatana (✉)
Division of Clinical Immunology, Department of Medical Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai 50200, Thailand
e-mail: asimi002@hotmail.com

W. Praditwongwan · S. Saoin · C. Tayapiwatana
Biomedical Technology Research Unit, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency at the Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai 50200, Thailand

P. Chuankhayan · C.-J. Chen (✉)
Life Science Group, Scientific Research Division, National Synchrotron Radiation Research Center, 101 Hsin-Ann Road, Hsinchu 30076, Taiwan
e-mail: cjchen@nsrrc.org.tw

T. Wisitponchai
Biomedical Engineering Center, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand

V. S. Lee
Computational Simulation Modelling Laboratory (CSML), Department of Chemistry and Center of Excellence for Innovation in Chemistry and Materials Science Research Center, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

V. S. Lee
Department of Chemistry, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia

S. Nangola
Division of Clinical Immunology and Transfusion Sciences, Department of Medical Technology, School of Allied Health Sciences, University of Phayao, Phayao 56000, Thailand

Introduction

HIV-1 assembly is driven by the polymerization of the Gag polyprotein precursor (Pr55Gag) [1–6], which is followed by the budding of membrane-enveloped immature virions at the cell surface and ends by their release into the extracellular milieu [7]. The initial steps of HIV-1 assembly require the anchoring of Pr55Gag at the inner face of the cell plasma membrane via its N-terminal myristoyl group and a charged surface of basic residues in the N-terminal region of the matrix (MA) domain [8, 9]. To become infectious, the immature virion must undergo maturation through a sequential proteolytic processing, in which Pr55Gag is cleaved into the major structural proteins MA, capsid (CA) and nucleocapsid (NC) [10]. CA and NC constitute the inner core of the mature virion and play a major structural role in the packaging and organization of the viral genome [1–6, 11–15].

Within the immature HIV-1 particle, the Pr55Gag lattice is stabilized by three different CA–CA intermolecular interfaces involving both N-terminal (NTD^{CA}) and C-terminal (CTD^{CA}) domains, i.e. NTD^{CA}–NTD^{CA}, NTD^{CA}–CTD^{CA} and CTD^{CA}–CTD^{CA}, respectively [14]. Thus, interaction with one or the other CA domain might negatively interfere with the viral assembly and budding processes, causing an overall decrease in the virus infectivity. HIV-1 CA therefore represents a promising target for antiretrovirals, not only for its major role in viral assembly, but also its multiple crucial functions at early steps of the viral life cycle. These include virus uncoating, interaction with host restriction factors and cytoplasmic transit and nuclear import of the viral preintegration complex [13].

Ankyrins and their artificial derivatives DARPins (*d*esigned *a*nkyrin-*r*epeat *prot*e*ins*) are attractive scaffolds to generate protein-based agents exerting different specific functions in a large variety of cells of prokaryotic and eukaryotic origins, and a variety of compartments, intracellular as well as extracellular [16–24]. Thus, artificial ankyrins might be designed as antiretroviral agents [25], capable of inhibiting the HIV-1 replication by blocking the binding of viral Gp120 to its cell surface receptor CD4 [22], or by altering certain functions of the Gag polyprotein, as in the case of Ank^{GAG}1D4 [26]. Ank^{GAG}1D4 is a DARPin which was isolated by screening a phage-displayed library of ankyrin repeat proteins on a viral target consisting of the carboxy-terminally truncated form of HIV-1 Gag polyprotein corresponding to the MA and CA domains (MA–CA). HIV-1 infection of SupT1 cells stably expressing N-myristoylated or non-N-myristoylated version of Ank^{GAG}1D4 resulted in a significant decrease in the viral progeny yield [26]. Experimental data indicated that Ank^{GAG}1D4 exerted its intracellular antiviral activity at the late phase of the HIV-1 life cycle, by negatively interfering with the Gag protein assembly and budding machinery [26].

The aim of this study was to identify the binding sites of Ank^{GAG}1D4 within the NTD^{CA} sequence by competition ELISA, using a panel of NTD^{CA}-based overlapping peptides as competitors of the Ank^{GAG}1D4-CA interaction. In addition, a rational model of Ank^{GAG}1D4-CA complex was constructed from data of molecular docking analysis, using X-ray structure of the Ank^{GAG}1D4 monomer and a homology model of CA. Our study brings further evidence that ankyrin repeat proteins represent a novel class of protein-based HIV-1 blockers which can act extracellularly or intracellularly, as viral restriction factors, and represent a promising alternative to antibodies for potential clinical applications. In addition, the identification of the target(s) of Ank^{GAG}1D4 in the CA structural protein at the amino acid level provided some information on the molecular mechanisms of the HIV-1 particle assembly. It should contribute to the development of new therapeutic strategies directed towards Gag oligomerization and viral assembly.

Materials and methods

Bacterial strain, plasmids and synthetic peptides

The pQE-30 expression plasmid (Qiagen), was used for production of 6xHis(H_6)-tagged recombinant proteins in M15[pREP4] bacterial cells (Qiagen). Plasmid pNL4-3,

S. S. Hong · P. Boulanger (✉)
Laboratory of Retrovirus and Comparative Pathology, UMR-754, University Lyon 1, 50, Avenue Tony Garnier, 69366 Lyon Cedex 07, France
e-mail: Pierre.Boulanger@univ-lyon1.fr

S. S. Hong · P. Boulanger
UMR-754, Retrovirus and Comparative Pathology, INRA, 50, Avenue Tony Garnier, 69366 Lyon Cedex 07, France

S. S. Hong
Institut National de la Santé et de la Recherche Médicale (INSERM), 101, Rue de Tolbiac, 75013 Paris, France

P. Minard
Institut de Biochimie et de Biophysique Moléculaire et Cellulaire (IBBMC) UMR-8619, Université de Paris-Sud et CNRS, 91405 Orsay Cedex, France

C.-J. Chen
Department of Physics, National Tsing Hua University, Hsinchu 30013, Taiwan

C.-J. Chen
Institute of Biotechnology, National Cheng Kung University, Tainan 701, Taiwan

obtained from the NIH AIDS Research and Reference Reagent Program (Division of AIDS, NIAID, NIH), was used as the template for isolation of the DNA fragments encoding H_6-CA (wild-type recombinant protein producing in baculovirus system), H_6-CA-WT and H_6-CA-mutant (wild-type and mutant recombinant proteins generating in bacterial system, respectively). The pTriEx expression plasmid (Novagen), was used for production of H_6-CA-WT and H_6-CA-mutant protein in BL21 (DE3) bacterial cells (Stratagene). The pBlueBac4.5 transfer vector was purchased from Invitrogen. The pBlueBac4.5-His vector derived from pBlueBac4.5 by an in-phase insertion of a H_6-encoding cassette at the 3′ end of the multiple cloning site. Synthetic pentadecapeptides covering the CAp24 protein of HIV-1 clade CRF01_AE [27] were kindly provided by Prof. Dr. Kiat Ruxrungtham (Chulalongkorn University, Faculty of Medicine, Bangkok, Thailand). They are presented in the Online Resource 1.

Construction of recombinant H_6-CA baculovirus expression vector

The DNA fragment encoding the H_6-CA was recovered from the pNL4-3 plasmid using the following pair of primers in a standard PCR method: Fwd_p24 *Nhe*I, 5′-GAGGAGGAGGTGCTATAGTGCAGAACCTCCAG-3′ and Rev_p24 *Kpn*I, 5′-GAGGAGGAGCTGGTACCTTACAAAACTCTTGCTTTATGGCC-3′. PCR was carried out using the Accuprime™ Pfx polymerase (Invitrogen), and the PCR product further purified using a GeneJet™PCR purification kit (Fermentas). The H_6-CA-encoding fragment was treated with *Nhe* I and *Kpn* I and subsequently ligated to the intermediate plasmid vector pBlueBac4.5-His. The ligated product pBlueBac4.5-H_6-CA was transferred to XL1-blue competent cells (Stratagene), and the correctness of the construct was verified using standard procedures.

Production and purification of recombinant protein H_6-CA from Sf9 cells

Recombinant H_6-CA was produced in the baculovirus (BV) expression system as described in detail in previous studies [28, 29]. In brief, Sf9 cells were cotransfected with pBlueBac4.5-H_6-CA (10 μg) and Bac-N-Blue™ DNA (a modified genome of *Autographa californica* multiple nucleopolyhedrosis virus; AcMNPV), using Cellfectin® II reagent, according to the Bac-N-Blue™ transfection and expression manual (Invitrogen). Recombinant virus BV-H_6-CA was isolated using the blue plaque selection method as previously described [26, 28, 30]. Sf9 cells were infected with BV-H_6-CA at a multiplicity of infection of 2–5 plaque forming units (pfu) per cell. At 48 h post infection, cells were harvested and lysed by the freeze-thawing method. The cell lysate was clarified by centrifugation at 15,000*g*, for 30 min at 4 °C, and H_6-CA protein detected by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot analysis. After electric transfer to nitrocellulose membrane in a semi-dry apparatus, blots were blocked with 5 % skimmed milk in Tris-buffered saline (TBS) at room temperature (RT) for 1 h, and reacted with mouse monoclonal anti-His tag antibody (at 1:3,000 dilution in the blocking solution) at RT for 1 h with slow agitation. After washing with TBST (0.05 % Tween-20 in TBS), HRP-conjugated goat anti-mouse IgG (1:3,000 dilution in 5 % skimmed milk in TBS) was added and further incubated at RT for 1 h. After an extra washing step, the specific band of H_6-CA protein was detected using the TMB membrane peroxidase substrate (KPL). Clarified Sf9 cell lysates was the starting material for the production and purification of H_6-CA protein. H_6-CA protein was purified by affinity chromatography on HisTrap column, using ÄKTA prime™ plus (GE Healthcare Bio-Sciences). The degree of purity of the H_6-CA protein was evaluated by SDS-PAGE analysis in 15 % polyacrylamide gel under reducing conditions. The gels were fixed and stained with Coomassie blue G250. The H_6-CA concentration was determined by the Bradford protein assay (Thermo Fisher Scientific).

Expression and purification of Ank^{GAG}1D4 from bacterial cells

H_6-tagged Ank^{GAG}1D4 protein was produced as previously described [26]. In brief, plasmid-harboring M15 [pREP4] cells were grown in 500 mL Terrific broth™ (Sigma-Aldrich) containing ampicillin (100 μg/mL), kanamycin (25 μg/mL), and 1 % (w/v) glucose at 37 °C with shaking, until $OD_{600\ nm}$ reached 0.8. Induction of recombinant protein expression was performed by addition of 1 mM IPTG, and the culture maintained for another 4 h at 30 °C with shaking. Cells were collected by centrifugation at 3,000 rpm for 30 min at 4 °C. The cell pellets were resuspended in lysis buffer (TBS pH 7.4, containing 1 μg/mL lysozyme, 1 mM EDTA and a cocktail of protease inhibitors), and subjected to three cycles of freeze-thawing. After centrifugation at 15,000*g* for 30 min at 4 °C, the Ank^{GAG}1D4 recovered in soluble form in the supernatant of ankyrin^{GAG}1D4 was purified by affinity chromatography on HisTrap column as above, followed by gel filtration chromatography on Sephadex G-75 (GE Healthcare Bio-Sciences). The purity of the Ank^{GAG}1D4 protein was determined by SDS-PAGE analysis and Coomassie blue staining. Purified Ank^{GAG}1D4 was chemically biotinylated, using the EZ-Link Sulfo-NHS-LC Biotin kit and following the instruction manual (Thermo Fisher Scientific).

Construction and expression of recombinant H_6-CA-mutant

Site specific mutagenesis of H_6-CA-mutant

The key amino acid residues of H_6-CA were mutated to alanine (A) using a QuickChange® Lightning Site-Directed Mutagenesis Kit (Stratagene) with pairs of specific primers shown in the Online Resource 2. In brief, the pNL4-3 was hybridized to the desired pair of primers and replicated by *PfuUltra*® high-fidelity (HF) DNA polymerase. Subsequently, the parental template was treated with *Dpn* I endonuclease before transformation of *Escherichia coli* strain, XL-1 Blue. The presence of the mutated H_6-CA (H_6-CA mutant) was analyzed by standard sequencing method (Online Resource 3).

Construction of recombinant H_6-CA mutant expression vector

The DNA fragment encoding the mutated H_6-CA was amplified from the pNL4-3_H_6-CA mutant plasmid as described above, using the following pair of primers in standard PCR method: Fwd_p24 *Xcm*I, 5′-GAGGAG CCATCACTCTTCTGGCCTATAGTGCAGAACCTC-3′ and Rev_p24 *Kpn*I, 5′-GAGGAGGAGCTGGTACCT TACAAAACTCTTGCTTTATGGCC-3′. PCR was carried out using KOD Hot Start DNA polymerase (Novagen) and further purified using a GeneJet™PCR purification kit (Fermentas). The purified fragments were next treated with *Xcm*I and *Kpn*I and ligated into pTriEx plasmid which had been cut with the same enzymes as previously. Finally, the ligation product was introduced into *E. coli*, XL1-Blue and the corrected clones were identified by standard sequencing method resulting in the new recombinant construct named pTriEx-H_6-CA mutant.

Production of recombinant H_6-CA mutant protein from bacterial cells

H_6-tagged CA mutant protein was produced following the described protocol elsewhere [31]. In brief, the bacterial BL21 (DE3) harboring pTriEx-H_6-CA mutant plasmid were cultured in 100 mL of terrific broth (1.2 % (w/v) tryptone, 2.4 % (w/v) yeast extract, 0.4 % (w/v) glycerol, 17 mM KH_2PO_4, 72 mM K_2HPO_4) containing ampicillin (100 μg/mL), and incubated at 37 °C with shaking until OD_{600} reached to 0.8. To induce protein expression, the culture was added with IPTG at a final concentration of 0.1 mM, and continuously shaken at 30 °C for 16–18 h. After induction, the cells were centrifuged at 5,000 rpm at 4 °C for 30 min. The pellet was washed with TBS (pH 7.4) and centrifuged again. The bacterial cells were resuspended in TBS and subsequently lysed by ultrasonication. The cell lysate was centrifuged at 10,000 rpm at 4 °C for 30 min, and the soluble protein was collected. The total protein concentration was quantified using a BCA Protein Assay (Pierce) whereas the concentration of H_6-CA protein was measured using Genscreen ULTRA HIV Ag-Ab kit (BIORAD). The presence of soluble H_6-CA was confirmed by SDS-PAGE and immunoblotting using monoclonal anti-CA as the detector.

Crystallization and data collection of Ank^{GAG}1D4

Crystallization was performed using the hanging-drop vapor-diffusion method at 18 °C. The initial condition was obtained from the Cryo II screen kit (Emerald BioSystems) containing polyethylene glycol (PEG) 400 (*w/v*, 30 %), PEG 3000 (*w/v*, 5 %), glycerol (*v/v*, 10 %) in HEPES buffer (100 mM, pH 7.5). This condition was optimized to improve the crystal quality by adjusting the concentration of PEG 400. Equal volumes (1 μL) of protein solution and crystallization reagent (100 mM HEPES, pH 7.5; PEG 400, 25 %; PEG 3000, 5 %; glycerol, 10 %) were mixed and equilibrated against the reservoir (110 μL). Ankyrin crystals grew within 2 days. The crystals were cryo-protected with glycerol (20 %) and frozen in liquid nitrogen before data collection. X-ray diffraction data were collected at the beamline 13B1 with a CCD detector (Q315, ADSC) at the National Synchrotron Radiation Research Center (NSRRC), Taiwan, and at the beamline BL12B2 with a CCD detector (Q210r, ADSC; Taiwan contracted beamline) at SPring-8, Japan. The data were processed with the *HKL-2000* program [32].

Structural determination and refinement of Ank^{GAG}1D4

The initial structure of $Ankyrin^{GAG}$1D4 was obtained using the molecular replacement method with the previously determined structure of DARPin (PDB code: 3NOC; [33]) as the search model using *MOLREP* in the *CCP4* suite [34], and further model building was performed with *Coot* [35]. All refinement was performed with the *refmac* program in *CCP4*. The correctness of the stereochemistry of the models was verified using *PROCHECK* [36].

Homology modeling

The degree of homology between the CA proteins of different HIV-1 isolates was determined using the *PSI-BLAST* program [37]. Sequence alignment of the CA N-terminal domains (Online Resource 4) showed a 91 % sequence identity between the CA protein used for our ankyrinotope mapping and the monomeric mutant CA_W184/M185A (PDB code: 2LF4; [38]). The homology models of CA

were therefore built using CA_W184/M185A as the template and the automated protein homology-modeling server SWISS-MODEL [39]. The CA_W184/M185A models had 20 different lowest-energy structures, due to the high flexibility of the 145-YSPTS-149 linker between the N- and C-terminal domains of the CA protein in solution. These different structures were clustered, on the basis of the various orientations of the linker, and candidates from each group were used as templates. The *PROCHECK* program [36] was used to check the stereochemical quality, the Ramachadran plots and G-factors of our homology modeling.

Molecular docking

Docking simulations of Ank^{GAG}1D4-CA complexes were carried out using the rigid-body docking algorithm ZDOCK in Discovery Studio (DS) 2.5 [40–42]. Initially, both Ank^{GAG}1D4 and NTD^{CA} proteins were minimized with Smart Mimimizer, without incorporating solvent and salt concentrations, followed by a 5,000-step energy minimization and 0.1 root-mean-square (RMS) gradient. Distance cutoff for scoring function was defined as 10 Å. The desolvation and electrostatic energy were also included in the ZDOCK scoring [40–42]. The clustering poses, RMSD cutoff, interface cutoff, and maximum number of clusters were defined as 6.0, 9.0, and 100 Å, respectively. To compare the interaction between helix 1 and helix 7 of NTD^{CA} with Ank^{GAG}1D4 protein, as suggested by the data from competition ELISA, the oligopeptides 14-PLSPRT LNAWVKVIEEKGF-32 corresponding to NTD^{CA} helix 1, and 121-NPPIPVGDIYKRWIILGLNKIVRMYQP-147, corresponding to helix 7 and the mini beta-hairpin between helices 6 and 7, were selected as potential ankyrinotopes for the analysis of helix 1-Ank^{GAG}1D4 and helix 7-Ank^{GAG}1D4 complexes, respectively. Moreover, the non-conserved residues of the binding site, based on the ankyrin consensus sequence previously reported [17], and blocked residues of Ank^{GAG}1D4 were defined. Docked poses which were candidates for each category (cluster center) were typed with CHARMm polar H force field, and refined with the RDOCK protocol [43]. To generate all plausible Ank^{GAG}1D4-CA complexes, the analysis of candidates of each group was performed using the same parameters.

Identification of key amino acid residues

The key residues in the interaction between the CA protein and Ank^{GAG}1D4, all docking poses obtained after running RDOCK were identified by collecting interacting atomic pairs with distance threshold of 5.0 Å in the complex, as described previously [44]. Initially, the interaction pairs were gained from DS 2.5 in string character, showing atom and residue pairs per one interaction pair. Numerous data of interaction pairs were rapidly and precisely analyzed by the programming language. The string form was converted to numeric vector, after that the same vector was counted in the same any position. Finally, the data represented the frequency of interaction pairs in any residue. The highest frequency interaction pair in any residue position in the CA was defined as first-key amino acid residue of the binding of CA to Ank^{GAG}1D4. Any residue position that had frequency interaction less than that of the first-key residue but more than others was identified as second-key amino acid residue. Likewise, the first-key residue of the binding of Ank^{GAG}1D4 to CA was identified in all candidate-poses of the m1H1 and m20H7 models, using the same method. The complex containing the first-key residues in both CA and Ank^{GAG}1D4 proteins was selected for further analysis.

Ankyrinotope mapping by competition ELISA

Microtiter plate conditioning

Microtiter plates were pretreated according to our previously described procedure [45]. In brief, 100 µL of 10 mM BCML (N, N-bis (carboxymethyl) lysine hydrate) diluted in 0.1 M $NaPO_4$ was added to the plate, which was then incubated at 4 °C overnight. The wells were washed three times with 200 µL washing buffer (0.05 % Tween 20 in water) for 5 min at 25 °C with shaking, and blocked with 200 µL of 0.05 % BSA in TBST. After 1 h-blocking, the wells were rinsed with the washing buffer (50 mM Tris–HCl, pH 7.5, containing 500 mM imidazole and 0.05 % Tween-20), followed by 100 mM Na_2 EDTA, pH 8.0. The plate was incubated with 10 mM $NiSO_4$ for 1 h at RT, and then sequentially washed with washing buffer and 50 mM Tris–HCl, pH 7.5, containing 500 mM NaCl. The plate was then ready to use for the capture of H_6-tagged proteins.

H_6-CA capture on Ni^+-coated plate

Purified H_6-CA was diluted in TBS-0.5 % BSA to a final protein concentration of 5 µg/mL, and the solution applied to the pre-treated plate and incubated overnight at 4 °C. Unbound protein was removed by washing four times with 20 mM Tris–HCl buffer, pH 7.5, containing 0.5 M NaCl and 6 M urea, with increasing concentrations of imidazole (20, 40, 60, and 80 mM). The wells were then washed once with TBS-0.1 % BSA containing 0.05 % Tween-20. A solution of biotinylated-Ank^{GAG}1D4 at 5 µg/mL was mixed with an equal volume of 60 µM solution of pentadecapeptide. The reaction mixture was added to the coated wells and the plate was incubated for 1 h at RT. The wells were then washed with a high-stringency washing buffer (50 mM Tris–HCl, pH 7.5, 50 mM NaCl, 0.1 % BSA, and

0.05 % Tween-20) followed by 0.1 % BSA-0.05 % Tween-20 in TBS. HRP-conjugated streptavidin (1:12,000 dilution in TBS-0.5 % BSA) was added, and plates incubated for 1 h at RT. After washing, TMB substrate was added to the wells (100 µL/well). The reaction was stopped with 1 N HCl, and the $OD_{450\ nm}$ was measured using a microplate reader (Biodirect).

Ankyrin-mediated protein capture and enzyme-linked immunoassay

Protein interaction was evaluated using an ELISA-modified method, based on the ankyrin-mediated capture of a target protein, followed by enzyme-linked immunoassay (AMELIA; Online Resource 5). The microtiter plate was coated with avidin (10 µg/mL) at 4 °C overnight for capturing the biotinylated-Ank^{GAG}1D4 (Biot-Ank^{GAG}1D4). The Biot-Ank^{GAG}1D4 was diluted in TBS-2 % BSA to a final protein concentration of 10 µg/mL and then immobilized on avidin-coated plate by incubation at RT for 1 h. After incubation, unbound proteins were removed by washing three times with TBS-2 % BSA containing 0.05 % Tween-20 (TBS-T). An equal amount of CA protein, which was measured by the Genscreen ULTRA HIV Ag-Ab kit (BIORAD), was added to Ank^{GAG}1D4-captured plate and incubated for 1 h at RT and then washed three times with TBS-T. The Ank^{GAG}1D4-mediated target protein capture was detected by anti-CA monoclonal antibody (a kind gift from Prof. Dr. Watchara Kasinrerk, Faculty of Associated Medical Sciences, Chiang Mai University, Thailand) at dilution 1:1,000 by incubation at RT for 1 h and then washed three times with TBS-T. The HRP-conjugated anti-mouse immunoglobulins (KPL; 1:3,000) were subsequently added, and incubated for 1 h at RT. After washing, TMB substrate was added to the wells. The reactions were finally stopped with 1 N HCl, and measured the optical density at 450 nm using a microplate reader (Biodirect).

Results

Ank^{GAG}1D4 binding sites on the HIV-1 CA: ankyrinotope mapping

In order to identify the binding sites of Ank^{GAG}1D4 on NTD^{CA} sequence (referred to as ankyrinotopes, by analogy with epitopes and phagotopes), we used competition ELISA with oligo-His-tagged CA protein (H_6-CA) as the immobilized ligate, and biotinylated-Ank^{GAG}1D4 (biot-Ank^{GAG}1D4) as the soluble ligand. Ligand competitors consisted of overlapping pentadecapeptides (referred to as F1 to F35; Online Resource 1) with a 11-residue overlap [27], covering the entire NTD^{CA} domain including its N-terminal ß-hairpin and seven alpha-helices (H1 to H7; Fig. 1). The binding activity of biot-Ank^{GAG}1D4 to H_6-CA was reduced in the presence of four groups of peptides, but two of them showed a low-to-moderate competition effect (20–50 %). Assuming that competition levels $\leq$50 % should not be considered as significant since they could be due to some degree of cross reactivity between soluble competitors, only two groups of peptides competed over this threshold: F5 + F6, corresponding to helix H1 (amino acid residues 16–30), and F32-F35, corresponding to helix H7 (residues 126–145). Both H1 and H7 are oriented on the same face of the NTD^{CA}, although with different degrees of accessibility to the solvent [3, 6, 14, 46]. Since ankyrins bind to their targets via a binding surface rather than a cleft, the possibility existed that H7 might contribute to the formation and/or stability of the Ank^{GAG}1D4-CA complex. Crystallographic study and computational simulation were performed to further define the respective role of H1 or H7 in the binding of CA to Ank^{GAG}1D4.

Structure determination and refinement of Ank^{GAG}1D4 protein

The crystal structure of Ank^{GAG}1D4 was determined and subsequently applied to the modeling of its complex with CA, using molecular docking analysis. The parameters of the crystals and the statistics of X-ray diffraction data and structure refinement are presented in Table 1. The coordinates and structure factors of the crystal structure described herein have been deposited in the RCSB Protein Data Bank (http://www.rcsb.org) under the accession code 4HLL. The structural model of Ank^{GAG}1D4 showed the expected protein folding and helix-turn-helix-β-hairpin repeat motifs characteristic of the previously published ARP structure (Fig. 2A), with an amino-acid sequence identity of 77 % [33]. The molecular surface model, presented in Fig. 2B, showed that the variable amino acid residues forming the CA-binding region are all oriented on the same face of Ank^{GAG}1D4. It also showed the grooves which separate its three constitutive alpha-helical repeat motifs.

Homology modeling of the HIV-1 CA protein

The structure of the HIV-1 CA protein used as the template in the present study was that of the double CA mutant W184A/M185A (CA_W184/M185A; PDB code: 2LF4; [38]). The rationale for the choice of CA_W184/M185A is the following: (1) CA_W184/M185A is stable as a monomer in solution, and its structure permits to investigate the interaction with various CA ligands; (2) the structure of CA_W184/M185A shows independently folded NTD^{CA} and C-terminal (CTD^{CA}) domains joined by a flexible linker (YSPTS); (3) because both mutations

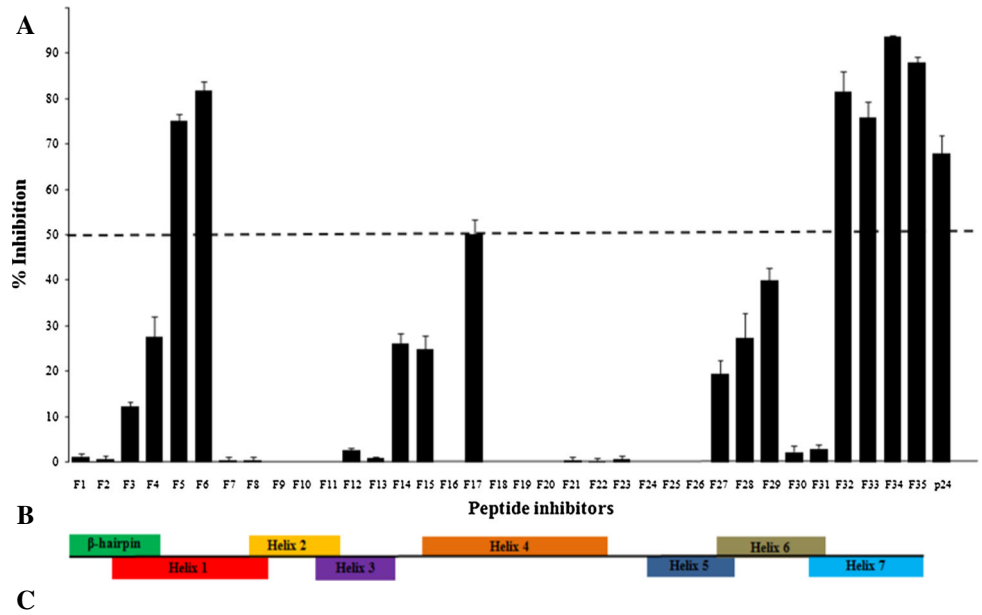

Fig. 1 Ankyrinotope mapping in HIV-1 NTD^{CA}. **A** Data from competition ELISA. Binding of the viral capsid protein CA to Ank^{GAG}1D4 was analyzed in the presence of competing overlapping pentadecapeptides covering the whole N-terminal domain of the CA (NTD^{CA}). Positive control sample for homologous competition consisted of CA protein (rightmost bar, p24) without peptide. The pentadecapeptide sequences are provided in the Online Resource 1. Results were expressed as the percentage of binding inhibition (mean ± SD). **B** Schematic diagram of the position of the overlapping pentadecapeptides, relative to the regions of NTD^{CA} with a secondary structure, i.e. the N-terminal β-hairpin and helices 1–7, represented as boxes of different colours. **C** Amino acid sequence of the NTD^{CA} (HIV-1 clade CRF01_AE). The amino acids residues are colour-coded according to the colours used for the boxes presented in (**B**)

W184A and M185A are localized in the CTD^{CA} of CA_W184/M185A, they should not affect the interaction of the independent NTD^{CA} with Ank^{GAG}1D4; (4) except for the slightly lower infectivity of the virus and lower efficacy of Gag oligomerization, no other functions was altered in CA_W184/M185A, compared to wild-type CA protein [38]; (5) the NTD^{CA} of CA_W184/M185A shows a complete sequence identity with that of the CA protein target used in our previous study with Ank^{GAG}1D4 (pNL4-3 strain) [26]; (6) CA_W184/M185A has a high sequence identity index (91 %) with the pentadecapeptides of HIV-1 clade CRF01_AE [27] used as competitors in our ankyrinotope mapping (Online Resources 1 and 4).

Due to the flexibility of the YSPTS linker forming a hinge between NTD^{CA} and CTD^{CA}, 20 different lowest-energy structural models have been defined for the CA_W184/M185A protein in solution [38]. Based on the different orientations of the YSPTS linker from the structures, H1 was found to be highly accessible in all models, whereas H7 was located deeply within the space between two neighboring helices. According to the orientations of the linker, the structural models were hence classified as group 1 (9 models), group 2 (5 models), and group 3 (6 models). Group 2 lied along the N- to C-terminus axis, whereas groups 1 and 3 turned perpendicular to the main polypeptide chain axis, above and below the axis, respectively (Fig. 3A). Three models, model 1 (m1; Fig. 3B, E), model 5 (m5; Fig. 3C, F) and model 20 (m20; Fig. 3D, G) of the CA template, were selected as representative candidates of groups 1, 2, and 3, respectively. All three candidate-models were used to generate 3D homology structures of CA, using the SWISS-MODEL software in an automated mode.

The qualities of the models were assessed with the Ramachandran plot and G factor. The CA homology structures were found to contain a large number of residues in the core region (84.3–85.9 %), with no residue other than glycine in disallowed regions (Online Resources 6 and 7). There were two amino-acid residues (1.0 %) in the generously allowed region of m1 (A185 and Q180) and m20 (A185 and N34). However, A185 and Q180 are located in the CTD^{CA}, which does not belong to the Ank^{GAG}1D4

binding region, and N34 was included in the turn between helices 1 and 2. Moreover, these structures had dihedral G-factor values superior to −0.5 (Online Resource 7), which indicated a normal stereochemical geometry of structures. The homology models m1, m5 and m20 of the CA protein were therefore considered as acceptable for protein–protein docking simulation with Ank^{GAG}1D4.

Table 1 Statistics of X-ray diffraction data and structure refinement of Ank^{GAG}1D4

Data collection	
Space group	*C2*
Cell dimension (Å)	$a = 108.24$, $b = 29.02$, $c = 50.04$, $\beta = 98.06$
Resolution range (Å)	26.8–2.2
Observed reflections	
Total	31,543
Unique	8,035
Completeness (%)	99.3 (99.4)
Redundancy	3.4
R_{sym} (%)	4.9 (21.8)
Refinement	
Resolution range (Å)	26.8–2.2
R_{work}/R_{free} (%)	21.1/23.3
No. of water molecules	46
RMS deviation from ideal geometry	
Bond lengths (Å)	0.02
Bond angles (°)	1.95
Average *B* factor (Å^2)	37.57

Numbers in parenthesis refer to the highest resolution shell

Molecular docking simulation

In the low-energy structural models of the CA protein, based on the structure of CA_W184/M185A, helices H1 and H7 had relative orientations that did not, theoretically, allow simultaneous binding of H1 and H7 to the binding surface of Ank^{GAG}1D4. Experimentally, however, data from the peptide interaction scan suggested a possible interaction with one or the other helix, or both (refer to Fig. 1). To solve this ambiguity, we investigated the plausible structures of the hypothetical complexes between Ank^{GAG}1D4 and CA in its low-energy conformations.

The three CA homology models, i.e. m1, m5 and m20, were analyzed in docking simulation, giving six possible types of Ank^{GAG}1D4-CA putative complexes, involving m1H1, m1H7, m5H1, m5H7, m20H1 and m20H7. Plausible complexes (poses) were generated using the ZDOCK protocol in the Discovery Studio 2.5 software [40], which provided a large number of poses with various scores such as the shape complementarity score and electrostatics and desolvation energy terms. ZDOCK poses were classified into clusters based on different orientations of the complexes, and complex candidates from each cluster were refined using RDOCK [43]. Among the possible poses in all six types, three forms of Ank^{GAG}1D4, which bound to H1 or H7, were reported in ZDOCK protocol (Table 2). The highest frequencies of putative complexes involving Ank^{GAG}1D4 and H1 were observed in models 1 (323 poses) and 5 (314 poses). The simulation of these

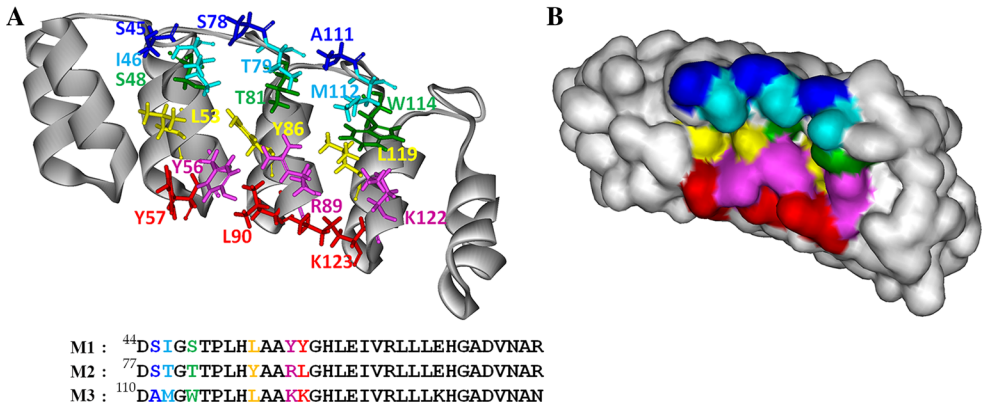

Fig. 2 Crystal structure of the ankyrin Ank^{GAG}1D4. **A** Ribbon model of Ank^{GAG}1D4, and amino acids involved in its CA-binding surface. *Top* the backbone of constant residues is shown by the *grey ribbon trace*, whereas the side chains of variable residues are presented as *coloured sticks*. *Bottom* amino acid sequence of the three helix-turn-helix domains D1, D2 and D3 (residues 44–142) constituting the CA-binding surface of Ank^{GAG}1D4. The amino acids are presented in the same colour code as in the ribbon model. **B** Surface model of Ank^{GAG}1D4. The ankyrin constant domain and variable surface are presented as *grey and coloured areas*, respectively. Amino acids in the variable binding surface are presented in the same colour code as in (**A**)

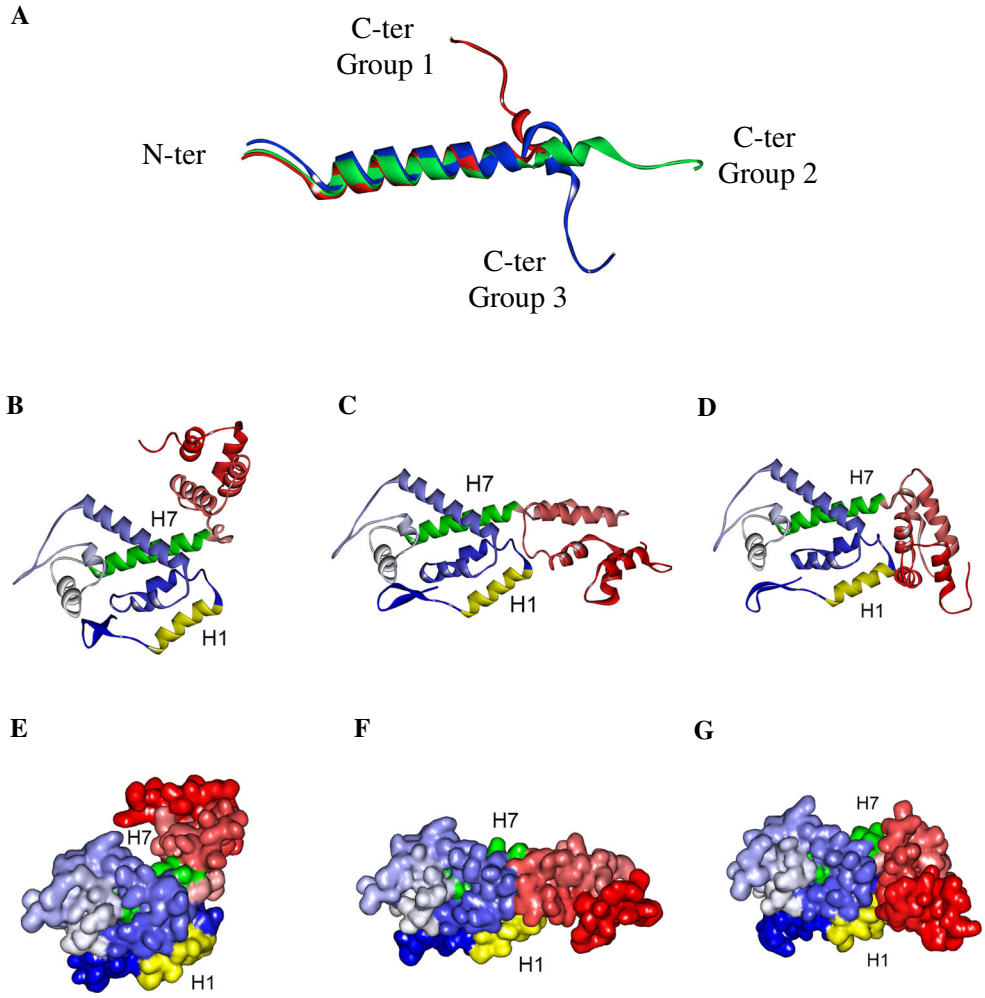

Fig. 3 Structural models of the HIV-1 CA protein. **A** Possible spatial orientations of the YSPTS linker forming the hinge between the N- and C-terminal domains of the CA protein. Each colour represents the candidate selected for each group of models: *red* group 1, *green* group 2, *blue* group 3. The model shown corresponds to the sequence (^{121}NPPIPVGEIYKRWIILGLNKIVRM-*YSPTS*-ILDIRQGPKEP^{160}), overlapping the mini beta-hairpin (NPPIP) and helix 7 (*underlined*) of the NTD^{CA}, the YSPTS linker (*italics*), and the first eleven residues of the CTD^{CA} [46]. N-ter and C-ter refer to the N- and C-termini, respectively. **B–G** Structural characteristics of the monomeric form of the CA protein used as the template for homology modeling (CA_W184/M185A; PDB code: 2LF4). *Top row* (**B–D**): ribbon representation of different models of the CA structure. **B** m1; **C** m5; **D** m20. *Bottom row* (**E–G**): surface representation of the three models. **E** m1; **F** m5; **G** m20. The *blue and red gradient colours* correspond to the N- to C-terminus direction. The NTD^{CA} helices 1 (H1) and 7 (H7) are shown in *yellow* and *green*, respectively. Note that the level of exposure and accessibility of H7 is minimal in model m1, and maximal in m20

complexes, which were grouped in the same cluster, showed that the complexes also exhibited high frequencies, with 38 and 36 poses for m1 and m5, respectively. In the case of H7, the highest frequency of poses was obtained with m20H7 (127 poses; Table 2). Out of the 20 possible structural models of CA_W184/M185A [38] distributed among three groups, 14 models were in favour of the interactions of Ank^{GAG}1D4 with H1. Although the chance of H1 being exposed to the solvent is statistically higher than that of H7, this analysis suggested that the formation of complexes between Ank^{GAG}1D4 and CA could be mediated by H1 as well as H7.

Prediction of amino-acid residues involved in Ank^{GAG}1D4-NTD^{CA} interaction

Frequency of amino-acid residue interacting pairs

The prediction of amino-acid candidates for binding to Ank^{GAG}1D4 was performed by analyzing the frequencies of potential atomic interaction pairs occurring at a 5 Å-distance at each amino-acid position in the NTD^{CA} helices 1 and 7. The first-key residue implicated in the binding of NTD^{CA} to Ank^{GAG}1D4 was defined as the one exhibiting the highest frequency of interaction pairs, whereas lower

Table 2 Frequency of the different types of Ank^{GAG}1D4-CA complexes generated by ZDOCK, using six models implicating the helix 1 (m1H1, m5H1 and m20H1) or helix 7 (m1H7, m5H7 and m20H7)

	Number of complexes			
	H1		H7	
	All	Cluster	All	Cluster
m1	323	38	13	2
m5	314	36	61	12
m20	118	31	127	16

Table 3 Nature and position of the first-key and second-key amino-acid residues of NTD^{CA} involved in its interaction with Ank^{GAG}1D4, in 38 candidate poses of the m1H1 model and 16 candidate poses of m20H7

m1H1			m20H7		
Residue and position in CA	Key residue		Residue and position in CA	Key residue	
	First	Second		First	Second
Q13	–	1	R82	–	1
P17	–	1	N121	1	3
R18	23	4	K131	–	1
N21	12	3	R132	10	4
K25	–	14	N139	–	1
E28	1	3	R143	5	6
E29	2	5			
E35	–	6			
T58	–	1			

frequency values would identify second-key residues. The candidates of each cluster of m1H1 and m20H7 were evaluated to identify the key residues in helices 1 and 7, respectively.

Among the first-key residues in 38 ZDOCK poses of m1H1, arginine at position 18 (R18^{CA}) was the most represented amino acid with 23 poses, and asparagine at position 21 (N21^{CA}) ranked second with 12 poses (Table 3). Considering the second-key residue, lysine at position 25 (K25^{CA}) was the most represented amino acid with 14 poses. In the Ank^{GAG}1D4 binding site assigned to helix 1 of the NTD^{CA}, the side chain of R18^{CA} was therefore predicted to be a hot-spot residue, whereas the N21^{CA} and K25^{CA} showed a lesser probability of contribution to the protein interaction. Notably, when R18^{CA} was replaced by A18^{CA} in the p269 model (m1H1), to mimic the R18A^{CA} mutation (see below), and compared to the wild type form, the frequency of atomic interaction pairs with a distance threshold of 5.0 Å at the position 18 was significantly decreased: 20 interaction pairs were calculated for the R18A^{CA} mutant versus 174 pairs for the wild type form R18^{CA} (data not shown). At position 25, the K25A^{CA} mutant showed only 4 interaction pairs, whereas the wild type form K25^{CA} showed 46 pairs.

Table 4 Nature and position of the first-key and second-key amino-acid residues of Ank^{GAG}1D4 involved in its interaction with NTD^{CA} in 38 candidate poses of the m1H1 model and 16 candidate poses of m20H7

m1H1			m20H7		
Residue and position in Ank1D4	Key residue		Residue and position in Ank1D4	Key residue	
	First	Second		First	Second
R23	4	5	K16	–	2
S45	1	2	D44	1	–
I46	1	–	S45	2	3
Y56	11	–	Y56	2	–
Y57	2	2	Y57	2	–
S78	1	2	S78	1	–
T79	1	1	T79	–	2
R89	4	10	R89	3	1
L90	2	–	K122	2	2
A111	–	2	K123	1	2
W114	–	1	H125	–	1
K122	7	7	H144	–	1
K123	4	3	F145	2	–
F145	–	1	N156	–	2

For the binding of helix 7 to Ank^{GAG}1D4, arginine residue at position 132 (R132^{CA}) had the highest probability to be the first-key residue (10 poses; Table 3), followed by arginine-143 (R143^{CA}; 5 poses; Table 3). Therefore, R132^{CA} was considered to be the main hot-spot. In the case of mutations in the p29 model (m20H7), the number of interaction pairs calculated for the wild type R132^{CA} was 184 versus 4 for the R132A^{CA} mutant, and 81 pairs for the wild type R143^{CA} versus 7 for the R143A^{CA} mutant (data not shown).

Amino-acid residue(s) of Ank^{GAG}1D4 critical for NTD^{CA} binding

To identify the hot-spot residue(s) of Ank^{GAG}1D4 involved in the binding to NTD^{CA}, the same strategy of frequency analysis as described above was performed on the 38 candidate poses of m1H1 and 16 candidate poses of m20H7. For the m1H1 complex, tyrosine at position 56 (Y56^{ANK}) was ranked first among the hot-spot residues, since it was the most represented amino acid as first-key residue, with 11 poses (Table 4). Lysine at position 122 (K122^{ANK}) and arginine at position 89 (R89^{ANK}) ranked second as hot-spot residues, with 7 poses for K122^{ANK}, and 10 poses for R89^{ANK} (Table 4). Although the differences in

the frequency values were less pronounced for the m20H7 complex, residue R89 and the two pairs Y56/Y57 and K122/K123 emerged from the list, as well as S45 (Table 4).

Experimental analysis of amino-acids critical for Ank^{GAG}1D4-NTD^{CA} interaction

Hot-spot residues in the HIV-1 CA protein

In order to experimentally support our model prediction, the hot-spot arginine residues in helices 1 and 7 were substituted by alanine in recombinant H_6-tagged CA protein, to generate the R18A^{CA}(H1), R132A^{CA}(H7) and R143A^{CA}(H7) mutants (Online Resource 3). The effect of these mutations was evaluated by an ELISA-modified method (AMELIA), consisting of ankyrin-dependent protein capture followed by a conventional immunoassay (Online Resource 5). The three R-to-A mutants showed a significantly reduced binding activity with immobilized Ank^{GAG}1D4, compared to wild-type H_6-CA (Fig. 4). This confirmed that R18, R132 and R143 in the NTD^{CA} domain were key players of the CA-Ank^{GAG}1D4 interaction. However, R18A^{CA}(H1) showed only a 50 % decrease in Ank^{GAG}1D4 binding, whereas R132A^{CA}(H7) and R143A^{CA}(H7) almost totally abolished the binding. This suggested that the binding of the CA to Ank1D4 via H7 could substitute, at least partially, for the lack of binding via H1. On the opposite, R132A^{CA}(H7) and R143A^{CA}(H7) mutations could not be compensated for by the binding of CA to Ank^{GAG}1D4 via H1.

Hot-spot residues in Ank^{GAG}1D4

The amino-acid residues of Ank^{GAG}1D4 critical for the interaction with CA protein were investigated by sequencing naturally occurring ankyrins which were found to be negative for CA binding in our ankyrin library screening process [26]. The prerequisite was to compare the amino acid sequence of ankyrins with the same molecular mass as Ank^{GAG}1D4 (163 residues, 17.9 kDa) and the same number of alpha-helical modules, viz. three. Ank-2E3 and Ank-2D3, two ankyrins of ca. 18 kDa characterized as non-binders [26], were thus selected for sequencing. Ank^{GAG}1D4, Ank-2E3 and Ank-2D3 showed highly conserved sequences, but positions 45, 56, 89 and 122 were not occupied by conserved residues in the three ankyrins. In Ank^{GAG}1D4, S^{45}, R^{89} and K^{122} were replaced by N^{45}, V^{89} and A^{122} in Ank-2E3, and by E^{45}, F^{89} and A^{122} in Ank-2D3 (Online Resource 8). For position 56, Y^{56} in Ank^{GAG}1D4 was replaced by L^{56} in Ank-2D3. More striking differences between Ank^{GAG}1D4 and the two non-binders were found when sequence comparison was performed at the dipeptide level. Dipeptides ^{45}SI^{46}, ^{56}YY^{57}, ^{89}RL^{90} and ^{122}KK^{123} in Ank^{GAG}1D4 were replaced by the nonconservative changes ^{45}NT^{46}, ^{56}YS^{57}, ^{89}VY^{90} and ^{122}AS^{123} in Ank-2E3, and by ^{45}EA^{46}, ^{56}LS^{57}, ^{89}FG^{90} and ^{122}AH^{123} in Ank-2D3 (Online Resource 8). This confirmed the status of hot-spots for positions 45/46, 56/57, 89/90 and 122/123 in the Ank^{GAG}1D4 sequence.

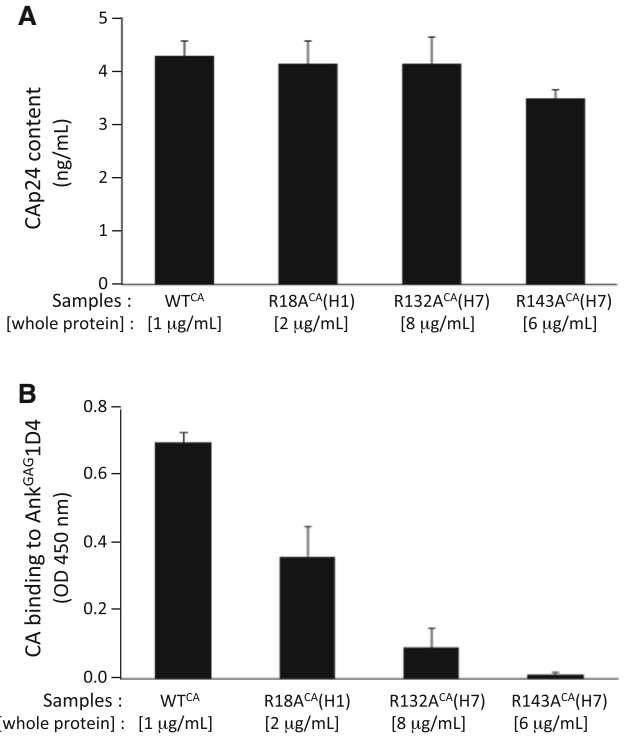

Fig. 4 Influence of R-to-A mutations in the helices 1 and 7 of NTD^{CA} on the interaction of the CA protein with Ank^{GAG}1D4. **A** *Normalization of the CA protein*. The normalization of the CA protein amount in each sample immobilized on ELISA plate was performed using a sandwich CAp24 detection kit (refer to "Materials and methods"). **B** *Binding assays* The binding of Ank^{GAG}1D4 to the wild-type CA protein (WT^{CA}) and CA mutants was performed using AMELIA, an ankyrin-mediated protein capture and enzyme-linked immunoassay (refer to Online Resource 5). The CA mutants consisted of arginine-to-alanine substitutions at positions 18, 132 and 143, corresponding to hot-spot residues for binding to Ank^{GAG}1D4, identified by prediction analysis

Molecular modeling of the Ank^{GAG}1D4—NTD^{CA} interaction

The candidate complexes of m1H1, which involved R18^{CA} and Y56^{ANK} as both first-key residues, were selected for further analysis and 3D modeling. These complexes corresponded to the complex pose numbers 123 (p123), p132, p265, p269 and p273. Out of five selected complexes, Y56^{ANK} in p269 and p273 was the first-key residue for the interaction with R18^{CA}. The p269 complex had a higher ZDock score, 3.64, compared to 3.1 for p273 (data not

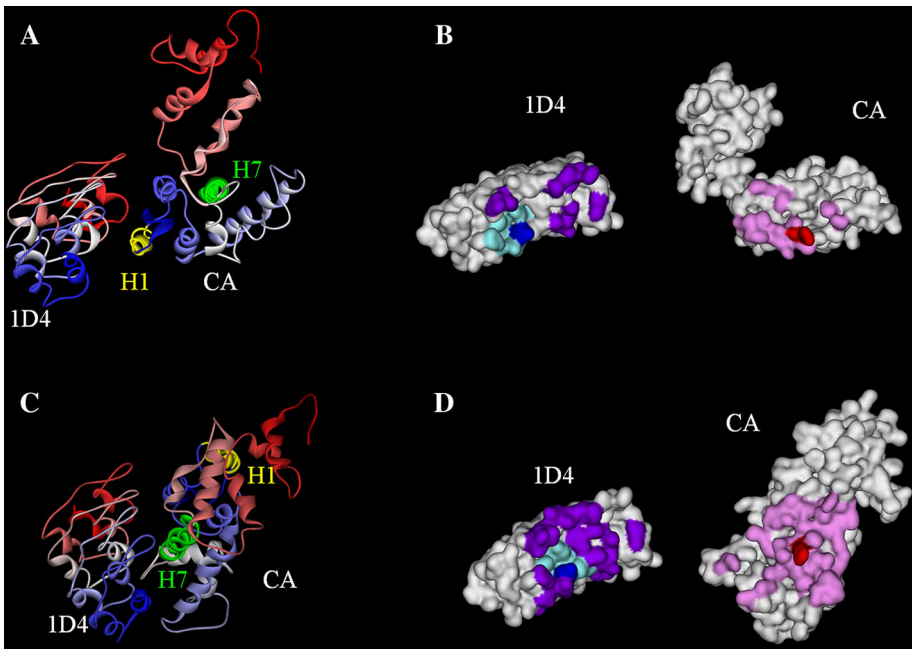

Fig. 5 Structural models of Ank^{GAG}1D4-CA complexes. **A**, **B** Structure of the p269 complex (m1H1). **C**, **D** Structure of the p29 complex (m20H7). The proteins are depicted as ribbon style (**A**, **C**) and surface style (**B**, **D**). In **A**, **C**, the *blue and red gradient colours* correspond to the N- to C-terminus direction, and the NTD^{CA} helices 1 (H1) and 7 (H7) are shown in *yellow* and *green*, respectively. In **B**, **D**, the colours in the surface representation showed the interacting surfaces of both partners. At the NTD^{CA} surface (**B**, **D**), the hot-spot residues for the binding to Ank^{GAG}1D4, R18^{CA} (model p269) and R132^{CA} (model p29) are in *red*, and the minor binding residues are in *pink*. At the Ank^{GAG}1D4 surface, the major binding residue Y56^{ANK} is in *dark blue* (**B**, **D**); the minor binding residues, interacting with R18^{CA} in helix 1 (**B**) and R132^{CA} in helix 7 (**D**), are in *aqua blue*, whereas other accessory binding residues are in *purple*

shown). In the p269 complex, R18^{CA}, located in a highly accessible region of the NTD^{CA}, could interact with a hydrophobic groove of Ank^{GAG}1D4 delineated by residues L53^{ANK} and Y86^{ANK} (Figs. 5A, B, 6A–C). This model implied the occurrence of three hydrogen bonds, involving respectively (1) the guanido group of R18^{CA} and the hydroxyl group of Y56^{ANK} (R18:O—Y56:HH); (2) the amino group of Q13^{CA} in the beta-hairpin of NTD^{CA} and the hydroxyl group of S45^{ANK} (Q13:HE22—S45:OG); (3) the carboxyl group of E29^{CA} and the amino group of K122^{ANK} (E29:OE1—K122:HZ3; Fig. 6A). The lengths of the three H-bonds were 2.2, 2.2 and 1.9 Å, respectively. Interestingly, K122^{ANK} showed a relatively high frequency of ZDOCK poses (refer to Table 4), suggesting that it could also play the role of hot-spot residue, besides Y56^{ANK}.

In the case of the m20H7 complex, which involved R132^{CA} and Y56^{ANK} as first-key residues, the p29 docking model was selected to analyze the interaction of Ank^{GAG}1D4 with helix 7 (Figs. 5C, D, 6D–F). Five hydrogen bonds would occur in the p29 model. (1) One would take place between the guanido group of R132^{CA} and the hydroxyl group of Y56^{ANK} (R132:HH11—Y56:OH); and four additional stabilizing hydrogen bonds would take place (2) between R82^{CA} and A24^{ANK} (R82:HH21—A24:O); (3) between R143^{CA} and S45^{ANK} (R143:HH11—S45:O); (4, 5) and two extra hydrogen bonds between N139^{CA} and S45^{ANK} (N139:HD22—S45:O) and (N139:O—S45:HG; Fig. 6D). The lengths of these four H-bonds were 2.4, 2.2, 2.4 and 2.2 Å, respectively. In terms of interaction energy, the H7-bound CA-Ank^{GAG}1D4 complex (as in the p29 docking model) had a lower energy than H1-bound CA-Ank^{GAG}1D4 (p269 model), −30.6 kcal/mol for the former, compared to −7.9 kcal/mol for the latter. The difference of 22.7 kcal/mol between the two potential interactions implied a higher stability of the H7-mediated CA-Ank^{GAG}1D4 complex.

In summary, the putative residues of Ank^{GAG}1D4 interacting with the NTD^{CA} helix H1 would be, in the order of pre-eminence, Y56, K122, K123, R89, R23, Y57, L90, S45, S78, T79, and I46. Those interacting with helix H7 would be, in the same type of ranking, R89, K122, S45, Y56, Y57, F145, D44, S78 and K123. Interestingly, S45, Y56, R89, K122 and K123 would be common CA interactors in both helices.

Discussion

Among the strategies envisaged to confer cellular resistance to HIV-1, different types of protein-based, peptide-based or peptidomimetic inhibitors targeting viral

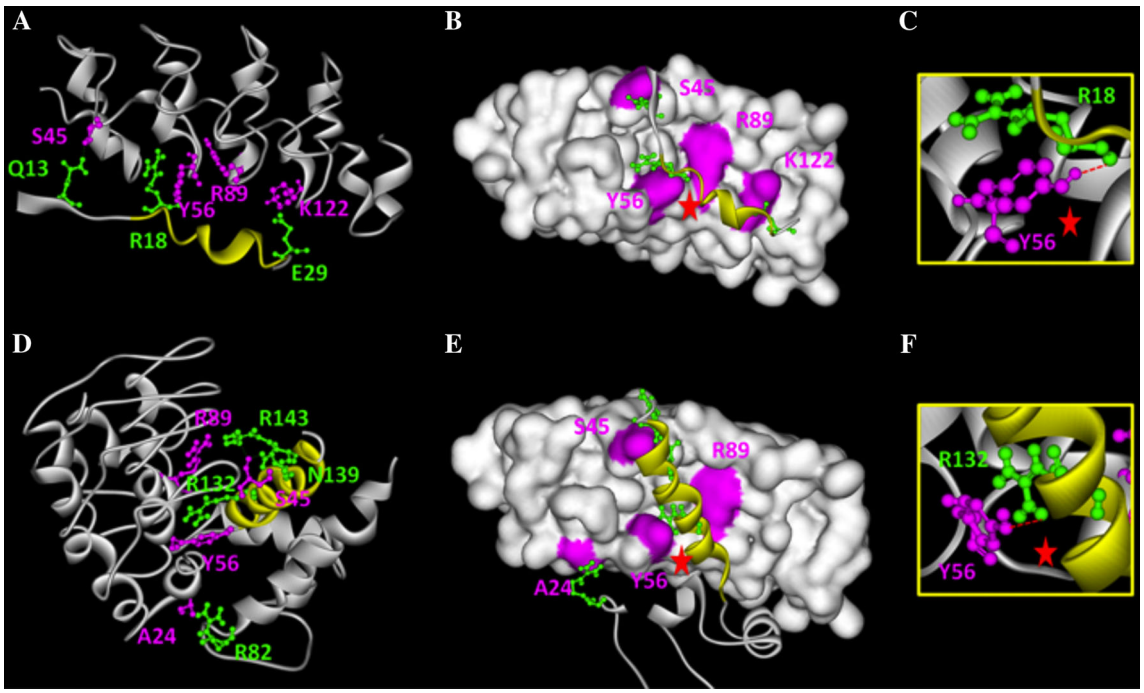

Fig. 6 Amino-acid interactions in the p269 and p29 models of Ank^{GAG}1D4-CA complexes. The Ank^{GAG}1D4 molecules complexed with the helix 1 (m1H1; panels **A**, **B**, and **C**) or helix 7 of NTD^{CA} (m20h7; panels **D**, **E**, and **F**) are shown in different presentations. In the ribbon style the presentation (**A**) and (**D**), the hot-spot residues and residues involved in intermolecular hydrogen bonds are shown in *green* for those belonging to NTD^{CA} and in magenta for those of Ank^{GAG}1D4. The *yellow ribbon* represents the helix 1 in (**A**), and helix 7 in (**D**). The hot-spot residues in NTD^{CA} include R18^{CA} in the complex model p269 and R132^{CA} in the complex model p29. The surface presentation (**B**, **E**), shows the hot-spot residues R18^{CA} and R132^{CA} located in a groove of the Ank^{GAG}1D4 protein. Panels **C** and **F** show the hydrogen bonds between R18^{CA} and Y56^{ANK} (**C**), and R132^{CA} and Y56^{ANK} (**F**), in enlarged views of the areas marked by *red stars* in (**B**) and (**E**), respectively

components have been proposed and evaluated. These include trans-dominant negative mutants of viral proteins [47–51], and innate human cell inhibitors engineered to act as specific antivirals, such as the TRIM5–cyclophilin fusion protein [52]. Small molecules targeting the HIV-1 CA protein and the viral assembly pathway have also been characterized. CAP-1 has been designed to specifically interact with NTD^{CA}, based on computational docking experiments [53]. Likewise, dodecapeptide CA-I has been identified by phage display as the ligand of CTD^{CA} [54, 55]. It has been suggested that both CAP-1 and CA-I inhibit the viral capsid assembly by disruption of the NTD^{CA}–CTD^{CA} interface and disorganization of the Gag assembly process [2–5]. More recently PF-74, a small molecule partner of NTD^{CA} [56], has been found to negatively interfere with the capsid assembly and destabilize the viral cores in vivo [57].

By screening a phage-display library of ankyrin repeat proteins [58] on the HIV-1 MA–CA polyprotein as the target, we have isolated and characterized Ank^{GAG}1D4, an artificial ankyrin exerting its antiretroviral activity at the intracellular level. We have shown that Ank^{GAG}1D4 binds to the N-terminal domain of the CA protein, and blocks the assembly and release of the viral progeny [26]. The aim of the present study was to better understand the molecular mechanism of the negative effect of Ank^{GAG}1D4 on HIV-1 assembly and virus release, and shed some light on the relationship between ankyrin structure and function.

The first step was to identify the Ank^{GAG}1D4 binding site(s) on the CA protein, by using a method referred to as ankyrinotope mapping and which was inspired from that used for epitope mapping [59]. As in conventional epitope mapping by competition ELISA [60–62], the NTD^{CA} domain of the HIV-1 CA protein was dissected into overlapping synthetic oligopeptides used as competitors of the Ank^{GAG}1D4-CA interaction. Unlike linear epitopes, which are efficiently mapped by peptide scanning methods, the protein interface in binary complexes of ankyrins with their cognate targets involves a surface of the target protein which is in its native conformation [63]. It was therefore unclear if our mapping strategy based on a peptide scan could provide unambiguous information on binding determinants. Our results clearly showed that peptides corresponding to local stretches of the NTD^{CA} sequence competed more efficiently than others for binding to Ank^{GAG}1D4. The competing peptides designated two

discrete regions of NTD^{CA}, corresponding to helices 1 and 7, as possible contributors to the binding interface.

In the CA protein, helices H1 and H2 have been identified as an interface essential for the CA assembly into the viral core, as they make the primary contacts between six NTD^{CA} which stabilize the CA hexamer [6]. Thus, the interaction of Ank^{GAG}1D4 with helix H1 would block the virus assembly by preventing the formation of (or destabilizing) the CA–CA interface. However, the mechanism of the Ank^{GAG}1D4-mediated antiviral effect might be more complex. A novel protein–protein interface has been recently identified in the HIV-1 CA protein. This interface differs from the well-known cyclophilin A-binding loop [64], and involves the cellular polyadenylation specific factor 6 (CPSF6) and the helices H4, H3 and H5, and the helix 5/6 turn in the NTD^{CA} [65]. The binding of Ank^{GAG}1D4 to H7 would partially overlap with the binding site for CPSF-6, and would also mask a small molecule inhibitor binding pocket recently identified (pocket 2; [66]), and delineated by helices H4, H5, and H7 on the surface of the NTD^{CA} [67–69].

The next step consisted of the crystallization of the Ank^{GAG}1D4 protein, which allowed us to construct a 3D structural model of Ank^{GAG}1D4, and a model of its binding surface deduced from the crystal structure. The characterization of this binding site was used to critically predict the formation of possible Ank^{GAG}1D4-CA complexes by molecular docking analysis. Our results suggested that Ank^{GAG}1D4 could bind to helices H1 and H7 of NTD^{CA}. Structure based-method showed that $R18^{CA}$ and $Y56^{ANK}$ were both first-key residues of the CA-Ank^{GAG}1D4 interaction mediated by helix 1. $Y56^{ANK}$ was found to be situated in a hydrophobic groove, and our model predicted that three hydrogen bonds of 1.9–2.5 Å in length would contribute to the formation and stabilization of the complex These H-bonds would take place between $R18^{CA}$ and $Y56^{ANK}$, $E29^{CA}$ and $K122^{ANK}$, and $Q13^{CA}$ and $S45^{ANK}$, respectively.

$R18^{CA}$, considered as the hot-spot residue of the H1-mediated binding of NTD^{CA} to Ank^{GAG}1D4, is a highly conserved residue in the CA sequence [66]. Out of 150 unique sequences reported for HIV-1 CA which we examined, 147 showed R at position 18, and 3 showed K (data not shown). The important structural and functional role of $R18^{CA}$(H1) has been demonstrated by extensive mutagenesis of HIV-1 *gag*: a double mutation involving $R18^{CA}$ (R18A/N21A) reduced the efficiency of capsid formation, provoked the formation of abnormally shaped and multiple capsids, and also reduced infectivity more than 20-fold [6]. Interestingly, $N21^{CA}$ in H1 was identified by ZDOCK poses as a second order residue in terms of interaction with Ank^{GAG}1D4. The status of H1 as a privileged target for assembly inhibitors has been demonstrated by other studies: antiviral CAP-1 and other benzodiazepine- and benzimidazole-derivatives have been found to bind to the hydrophobic pocket of NTD^{CA} involving H1 [69, 70].

The integrity of helix H7 is also critical for CA assembly: the L136D mutant in H7 showed an assembly-defective phenotype with very few conical cores, and 100-fold less infectivity compared to wild type virus [6]. Furthermore, a double mutant involving $R132^{CA}$ (E128A/R132A) and the single mutant R143A also showed an assembly-defective phenotype with few conical, and six-fold less infectivity compared to wild type virus [6]. $R132^{CA}$ is also a highly conserved residue [66]: out of the 150 sequences examined, R was found 134 times, K 14 times, and Q one time (data not shown). From our molecular docking analysis however, the formation of Ank^{GAG}1D4-CA complexes involving H7 was apparently less statistically favorable than those involving H1. Based on the CA crystal structure, H7 would only be partially accessible, in contrast to H1 [3]. This was in apparent contradiction with our results. (1) Experimentally, a higher efficiency of competition was observed with H7-derived peptides, compared to H1-peptides; (2) our mutagenesis of arginine residues in H7 showed more detrimental effect on the binding that the R18A mutation in H1; (3) in molecular docking analysis and complex modeling, the Ank^{GAG}1D4-CA complex mediated by H7 would involve five hydrogen bonds, versus three for H1; (4) Ank^{GAG}1D4-CA complex mediated by H7 would form with a lower interaction energy, and hence a higher stability, than the one involving H1.

Several mechanisms could be proposed to reconcile this apparent contradiction. In the first mechanism, there would be a simultaneous contribution of both H1 and H7 to the interaction of the CA protein with Ank^{GAG}1D4. This would require that the orientation of H1 and H7 relative to each other would change to permit their accommodation to the plane of the Ank^{GAG}1D4 binding surface. Such a mode of interaction would be unlikely for two reasons: (1) it would necessitate a strong distortion of the CA structure; (2) the probability of this type of event was statistically very low (1 out of 135 complexes).

In the second hypothetic scenario, the two binding sites in H1 and H7 could be used alternatively by Ank^{GAG}1D4, and the equilibrium between free Ank^{GAG}1D4, H1-bound CA-Ank^{GAG}1D4 and H7-bound CA-Ank^{GAG}1D4 complexes, would progressively evolve towards a majority of stable H7-bound CA-Ank^{GAG}1D4 complex, versus a minority of less stable H1-bound CA-Ank^{GAG}1D4 complex. This scenario is reminiscent of the avidity effect commonly observed with immunoglobulins, or more generally when one of the binding partners is an oligomer. The multiplicity of binding possibilities due to neighboring

binding sites in an oligomeric protein results in an enhancement of the apparent affinity between the two partners. Thus, the dual interaction of Ank^{GAG}1D4 with the CA via H1 or H7 could possibly induce a similar effect.

In the third mechanism, the initial binding event between Ank^{GAG}1D4 and NTD^{CA}, mediated by the highly accessible helix H1, would result in a conformational transition of the NTD^{CA} which would increase the accessibility of helix H7. Following this discrete allosteric change in NTD^{CA}, the Ank^{GAG}1D4 would then shift to a more stable Ank^{GAG}1D4-CA complex mediated by helix H7. The last two mechanisms are not mutually exclusive, and the last one would imply a significant degree of malleability and flexibility of the NTD domain of the CA protein, and not only between NTD^{CA} and CTD^{CA}, as previously described [3].

Several lines of experimental evidence indicate that the Pr55Gag polyprotein is flexible. Free in solution or bound to RNA, the Pr55Gag polyprotein adopts a folded conformation, although the formation of immature virus particles that bud at the plasma membrane requires Pr55Gag to be in an extended conformation (reviewed in [71]). It has been suggested that the simultaneous interaction of Pr55Gag with the plasma membrane via its N-terminal domain, and with the viral genomic RNA via its C-terminal domain, promotes the conformational changes leading to the Pr55Gag extension [72]. This model of alternate conformation of Pr55Gag as free within the cytoplasm versus bound to the plasma membrane is consistent with our previous analysis of the antiviral activity of both non-N-myristoylated (Myr(−), cytoplasmic), and N-myristoylated (Myr(+), membranal) Ank^{GAG}1D4 [26].

The results of the present study confirmed that artificial ankyrins or DARPins represent a class of macromolecules with multiple interests and potential applications. (1) In terms of fundamental virology, artificial ankyrins could serve as molecular probes to identify domains or motifs of viral components, which are essential for the virus, such as the conserved residues $R18^{CA}$ in helix H1, and $R132^{CA}$ in helix H7. As a corollary, $R18^{CA}$(H1) and $R132^{CA}$(H7) were designated by Ank^{GAG}1D4 as two highly specific targets for inhibitors of virus assembly. (2) Artificial ankyrins could also contribute to disclose transient states or/and discrete viral steps(s) of the viral machinery, which would permit the identification of new specific viral targets. (3) In terms of development of novel drugs and therapeutic strategies [73], artificial ankyrins can serve as antivirals which would act extracellularly [22] or intracellularly [21, 26], and interfere negatively with the virus life cycle. Like *bona fide* HIV restriction factors (reviewed in [74]), artificial ankyrins would bind and titrate viral components which would then become limiting factors of intracellular reactions, or compete with cellular partners of HIV-1 indispensable for achieving certain steps of the virus life cycle.

Acknowledgments The work in Thailand was supported by the National Research University project under the Thailand's Office of the Commission on Higher Education, the Research Chair Grant of National Sciences and the Technology Development Agency at the Faculty of Associated Medical Sciences, Chiang Mai University, and the Thailand Research Fund. The work in Lyon was supported by the French National Agency for Research on AIDS and Viral Hepatitis (Inserm-ANRS contract No. 11344/2011-2013). We are grateful to thank Prof. Dr. Kiat Ruxrungtham for his kind gift of synthetic pentadecapeptides, and the NIH AIDS Research and Reference Reagent Program for providing us with the pNL4-3 plasmid. The authors would like to thank the National Electronics and Computer Technology (NECTEC) for Discovery Studio software and also thank the University of Malaya for UMRG (Grant No. RP002-2012D) and FRGS (FP008-2012A), and for its support with computational work. We are also indebted to the supporting staffs at beamline BL13B1 at NSRRC and beamline BL12B2 at SPring-8 for the technical assistance. The crystallographic work was supported in part by National Science Council Grants 101-2628-B-213-001-MY4 and NSRRC Grant 1013RSB02 to C.-J. Chen.

References

1. Cimarelli A, Darlix J-L (2002) Cell Mol Life Sci 59:1166
2. Ganser BK, Li S, Klishko VY, Finch JT, Sundquist WI (1999) Science 283:80
3. Ganser-Pornillos BK, Cheng A, Yeager M (2007) Cell 131:70
4. Ganser-Pornillos BK, von Schwedler UK, Stray KM, Aiken C, Sundquist WI (2004) J Virol 78:2545
5. Ganser-Pornillos BK, Yeager M, Sundquist WI (2008) Curr Opin Struct Biol 18:203
6. von Schwedler UK, Stray KM, Garrus JE, Sundquist WI (2003) J Virol 77:5439
7. Adamson CS, Freed EO (2007) Adv Pharmacol 55:347–387
8. Freed EO (1998) Virology 251:1
9. Freed EO (2002) J Virol 76:4679
10. Göttlinger HG (2001) AIDS 15(Suppl. 5):S13
11. Cardone G, Purdy JG, Cheng N, Craven RC, Steven AC (2009) Nature 457:694
12. Darlix JL, Cristofari G, Rau M, Péchoux C, Berthoux L, Roques B (2000) Adv Pharmacol 48:345
13. Fassati A (2012) Virus Res 170:15
14. Pornillos O, Ganser-Pornillos BK, Kelly BN, Hua Y, Whitby FG, Stout CD, Sundquist WI, Hill CP, Yeager M (2009) Cell 137:1282
15. Purdy JG, Flanagan JM, Ropson IJ, Craven RC (2009) J Mol Biol 389:438
16. Amstutz P, Binz HK, Parizek P, Stumpp MT, Kohl A, Grütter MG, Forrer P, Plückthun A (2005) J Biol Chem 280:24715
17. Binz HK, Stumpp MT, Forrer P, Amstutz P, Pluckthun A (2003) J Mol Biol 332:489
18. Kummer L, Parizek P, Rube P, Millgramm B, Prinz A, Mittl PR, Kaufholz M, Zimmermann B, Herberg FW, Plückthun A (2012) Proc Natl Acad Sci USA 109:E2248
19. Li J, Mahajan A, Tsai MD (2006) Biochemistry 45:15168
20. Mosavi LK, Cammett TJ, Desrosiers DC, Peng ZY (2004) Protein Sci 13:1435
21. Schweizer A, Roschitzki-Voser H, Amstutz P, Briand C, Gulotti-Georgieva M, Prenosil E, Binz HK, Capitani G, Baici A, Pluckthun A, Grutter MG (2007) Structure 15:625
22. Schweizer A, Rusert P, Berlinger L, Ruprecht CR, Mann A, Corthésy S, Turville SG, Aravantinou M, Fischer M, Robbiani M, Amstutz P, Trkola A (2008) PLoS Pathog 4:e1000109

23. Sedgwick SG, Smerdon SJ (1999) Trends Biochem Sci 24:311
24. Zahnd C, Wyler E, Schwenk JM, Steiner D, Lawrence MC, McKern NM, Pecorari F, Ward CW, Joos TO, Plückthun A (2007) J Mol Biol 369:1015
25. Tamaskovic R, Simon M, Stefan N, Schwill M, Pluckthun A (2012) Methods Enzymol 503:101–134
26. Nangola S, Urvoas A, Valerio-Lepiniec M, Khamaikawin W, Sakkhachornphop S, Hong SS, Boulanger P, Minard P, Tayapiwatana C (2012) Retrovirology 9:17
27. Buranapraditkun S, Hempel U, Pitakpolrat P, Allgaier RL, Thantivorasit P, Lorenzen SI, Sirivichayakul S, Hildebrand WH, Altfeld M, Brander C, Walker BD, Phanuphak P, Hansasuta P, Rowland-Jones SL, Allen TM, Ruxrungtham K (2011) PLoS ONE 6:e23603
28. DaFonseca S, Blommaert A, Coric P, Hong SS, Bouaziz S, Boulanger P (2007) Antiviral Ther 12:1185
29. Gay B, Tournier J, Chazal N, Carrière C, Boulanger P (1998) Virology 247:160
30. DaFonseca S, Coric P, Gay B, Hong SS, Bouaziz S, Boulanger P (2008) Virol J 5:162
31. Kitidee K, Nangola S, Gonzalez G, Boulanger P, Tayapiwatana C, Hong SS (2010) BMC Biotechnol 10:80
32. Otwinowski Z, Minor W (1997) Methods Enzymol 276:307
33. Monroe N, Sennhauser G, Seeger MA, Briand C, Grütter MG (2011) J Struct Biol 174:269
34. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, Mccoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Acta Cryst D 67:235
35. Emsley P, Cowtan KD (2004) Acta Crystallogr D Biol Crystallogr 60:2126
36. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) J Appl Crystallogr 26:283
37. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Nucleic Acids Res 25:3389
38. Shin R, Tzou YM, Krishna NR (2011) Biochemistry 50:9457
39. Schwede T, Kopp J, Guex N, Peitsch MC (2003) Nucl Acids Res 31:3381
40. Chen R, Li L, Weng Z (2003) Proteins 52:80
41. Chen R, Weng Z (2003) Proteins 51:397
42. Hwang H, Vreven T, Pierce BG, Hung JH, Weng Z (2010) Proteins 78:3104
43. Li L, Chen R, Weng Z (2003) Proteins 53:693
44. Wisitponchai T, Nimmanpipug P, Lee VS, Tayapiwatana C (2013) Chiang Mai J Sci 40:232
45. Cressey R, Pimpa S, Chewaskulyong B, Lertprasertsuke N, Saeteng S, Tayapiwatana C, Kasinrerk W (2008) BMC Biotechnol 8:16
46. Gitti RK, Lee BM, Walker J, Summers MF, Yoo S, Sundquist WI (1996) Science 273:231
47. Apolloni A, Lin M-H, Sivakumaran H, Li D, Kershaw MHR, Harrich D (2013) Hum Gene Ther 24:270
48. Checkley MA, Luttge BG, Soheilian F, Nagashima K, Freed EO (2010) Virology 400:137
49. Fackler OT, D'Aloja P, Baur AS, Federico M, Peterlin BM (2001) J Virol 75:6601
50. Trono D, Feinberg MB, Baltimore D (1989) Cell 59:113
51. Walker RC Jr, Khan MA, Kao S, Goila-Gaur R, Miyagi E, Strebel K (2010) J Virol 84:5201
52. Neagu MR, Ziegler P, Pertel T, Strambio-De-Castillia C, Grütter C, Martinetti G, Mazzucchelli L, Grütter MG, Manz MG, Luban J (2009) J Clin Invest 119:3035
53. Tang C, Loeliger E, Kinde I, Kyere S, Mayo K, Barklis E, Sun Y, Huang M, Summers MF (2003) J Mol Biol 327:1013
54. Bartonova V, Igonet S, Sticht J, Glass B, Habermann A, Vaney MC, Sehr P, Lewis J, Rey FA, Kräusslich HG (2008) J Biol Chem 283:32024
55. Sticht J, Humbert M, Findlow S, Bodem J, Muller B, Dietrich U, Werner J, Kräusslich H-G (2005) Nat Struct Mol Biol 12:671
56. Blair WS, Pickford C, Irving SL, Brown DG, Anderson M, Bazin R, Cao J, Ciaramella G, Isaacson J, Jackson L, Hunt R, Kjerrstrom A, Nieman JA, Patick AK, Perros M, Scott AD, Whitby K, Wu H, Butler SL (2010) PLoS Pathog 6:e1001220
57. Shi J, Zhou J, Shah VB, Aiken C, Whitby K (2011) J Virol 85:542
58. Nangola S, Minard P, Tayapiwatana C (2010) Protein Expr Purif 74:156
59. Saint-Remy JM (1997) Toxicology 119:77
60. Lenstra JA, Kusters JG, van der Zeijst BA (1990) Arch Virol 110:1
61. Stanley KK, Herz J (1987) EMBO J 6:1951–1957
62. van Regenmortel MH (1989) Immunol Today 10:266
63. Sennhauser G, Grütter MG (2008) Structure 16:1443
64. Zhao Y, Chen Y, Schutkowski M, Fischer G, Ke H (1997) Structure 5:139
65. Price AJ, Fletcher AJ, Schaller T, Elliott T, Lee K, KewalRamani VN, Chin JW, Towers GJ, James LC (2012) PLoS Pathog 8:e1002896
66. Li G, Verheyen J, Rhee S-Y, Voet A, Vandamme A-M, Theys K (2013) Retrovirology 10:126
67. Goudreau N, Coulombe R, Faucher AM, Grand-Maitre C, Lacoste JE, Lemke CT, Malenfant E, Bousquet Y, Fader L, Simoneau B, Mercier JF, Titolo S, Mason SW (2013) Chem Med Chem 8:405
68. Lamorte L, Titolo S, Lemke CT, Goudreau N, Mercier JF, Wardrop E, Shah VB, von Schwedler UK, Langelier C, Banik SS, Aiken C, Sundquist WI, Mason SW (2013) Antimicrob Agents Chemother 57:4622
69. Lemke CT, Titolo S, von Schwedler UK, Goudreau N, Mercier JF, Wardrop E, Faucher AM, Coulombe R, Banik SS, Fader L, Gagnon A, Kawai SH, Rancourt J, Tremblay M, Yoakim C, Simoneau B, Archambault J, Sundquist WI, Mason SW (2012) J Virol 86:6643
70. Kelly BN, Kyere S, Kinde I, Tang C, Howard BR, Robinson H, Sundquist WI, Summers MF, Hill C (2007) J Mol Biol 373:355
71. Rein A, Datta SA, Jones CP, Musier-Forsyth K (2011) Trends Biochem Sci 36:373
72. Datta SA, Heinrich F, Raghunandan S, Krueger S, Curtis JE, Rein A, Nanda H (2010) J Mol Biol 406:205
73. Engelman A, Cherepanov P (2012) Nat Rev Microbiol 10:279
74. Li L, Oliveira NMM, Cheney KM, Pade C, Dreja H, Bergin A-MH, Borgdorff V, Beach DH, Bishop CL, Dittmar MT, McKnight A (2011) Retrovirology 8:94